Should the government sponsor training for the disadvantaged?

William P. Warburton[1]

and

Rebecca N. Warburton[2]

[1] Director, Economic Analysis, BC Ministry of Human Resources, Victoria, BC

[2] Assistant Professor, School of Public Administration, University of Victoria, Victoria, BC

Note to reader:
The authors gratefully acknowledge the use of British Columbia Ministry of Human Resources data. This paper is not an official document of the British Columbia Ministry of Human Resources or any other organization, and the authors are responsible for any errors or omissions.

Should the government sponsor training for the disadvantaged?

William P. Warburton and Rebecca N. Warburton

Introduction

Should the government sponsor training for the disadvantaged? The instinctive response is a resounding "yes." This sentiment was expressed by George Eliot's Mr. Tulliver when he said, "Tom's eddication 'ull be so much capital to him." And instinct seems to have been confirmed by economic theory from Adam Smith to the present. For example, an introductory economics text expresses the same idea, albeit less eloquently, when it says, "Individuals can thus influence their future stream of wages by paying for training in the current period.[1]" Yet some individuals, on their own, cannot afford training. If training does increase the wages of participants by a large enough amount, however, then the government should support training for the disadvantaged. Government support can enable the disadvantaged to obtain training they could not otherwise afford, thereby increasing their future wages, reducing poverty, reducing government expenditure on income support programs, and increasing tax revenue; outcomes that are appealing to voters on both the right and the left.

But lately training has received some bad press that conflicts with this view. For example, a *Globe and Mail* headline announced "Axworthy loses faith in jobs training,"[2] and *The Economist* reported, "Nobody seems to be saying that government-supported training is often a waste of money—nobody, that is, except researchers who have examined existing schemes."[3]

Unfortunately, it is easy to understand the disenchantment expressed in the *Globe and Mail* and in *The Economist*. Methodologically sound studies, mostly from the US, have shown that training for disadvantaged workers is often not effective. With the link between training and higher subsequent wages broken, training programs can leave the participants disappointed and cynical, while increasing costs to government; outcomes that are distasteful to voters on both the right and the left.

But training is neither always good nor always bad. The same studies that show that funding training is not always right also show that it is not always wrong. Estimates of the impacts of California's GAIN program made by the Manpower Demonstration Research Corporation (MDRC) are broadly regarded as reliable. On the positive side, MDRC found that in the Riverside county site, GAIN increased participants' incomes by 49% while returning to government $2.84 in increased tax revenue and decreased support payments for every $1.00 invested. On the negative side, the same study found very small impacts on earnings in other sites (e.g. 6% in Los Angeles county and 5% in Tulare County) and overall net costs to government (76¢ return for every $1.00 invested.)[4]

These results mean that the answer to our question, "Should the government sponsor training for the disadvantaged?" is easy for politicians. Government should sponsor training programs like the GAIN program as it was administered in Riverside California. It should not sponsor programs like the GAIN program as it was administered in Los Angeles or Tulare counties.

---

[1] John T. Addison and W. Stanley Siebert, *The Market for Labour* Goodyear, Santa Monica, 1979, page 110.

[2] July 28, 1995.

[3] The Economist, April 6, 1996, page 19.

[4] Variability in results across sites in a common finding. When the US Department of Labor implemented its Job Training Partnership Act programs it estimated its impact using random assignment in 16 sites. Abt Associates reported estimates of impact of training on the incomes of male youths varied from a $4,424 increase to a $6,581 decrease across the sites.

But if government takes this advice and directs a senior administrator to implement training programs, but only effective ones, it might leave that administrator wishing for a career change. Three problems could cause her to wake up screaming at night:

1. The impact of apparently similar training programs vary dramatically from site to site. Therefore, the administrator cannot simply find a program that worked well in another jurisdiction and import it. An administrator implementing a program like California's GAIN program might get results like those in Riverside county or like those in Los Angeles county. She could only find out which results she was getting by estimating the impacts in her own jurisdiction. At that point she would face a second problem:

2. Most of the estimates of program impact that are offered to the administrator will be unreliable. If she overcomes that difficulty and gets reliable estimates, she will run into a third problem:

3. Politicians and the public may have unreasonable expectations, either so high that even reasonable results may be interpreted as failure, or so low that any results showing that a program is effective are not believed.

This paper is intended as a toolbox specifically for dealing with these three problems. Since the administrator will have to commission estimates of program impacts, the paper contains a discussion of program impacts including a definition and a description of why impacts are difficult to estimate. To protect administrators from indefensible estimates of program impact, this essay gives five questions that an administrator can ask of studies to determine whether the results are likely to be reliable. For the public-spirited administrator, this section outlines a strategy for improving the overall quality of estimates of program impact[5]. And for dealing with false expectations, the paper reports estimates of the range of training programs for welfare recipients in British Columbia in the late 80's.

Estimating Impacts

What does "estimating impacts" mean?

It is important for administrators to appreciate that simply measuring what happens to people after they participate in programs is only half the task of estimating the impacts of programs. This point was driven home to a Canadian politician in the early 1980's. She was told that only 20% of welfare recipients going through a program remained dependent after 3 months, and bragged about the program's success. But her bragging backfired because the opposition quickly divided the number of welfare cases by the number of people going through the program and asked if she predicted an end to welfare in 30 months. Her staff had gotten her into trouble by ignoring half of the estimating-impacts task. She also needed to know how many of the participants in that program would have become independent of welfare even without the program.

The impact of a program on subsequent welfare dependence is the difference between:

1. the number who become independent with the program and

2. the number who would have become independent in the absence of the program.

This is illustrated by Figure 1. The subsequent welfare dependence of program participants is illustrated by the solid line. The subsequent welfare dependence that we expect that these individuals would have experienced in the absence of the program is illustrated by the dotted line. (Please suspend disbelief about the possibility of determining what the individuals would have experienced in the absence of the program until the next section.)

Putting this into concrete terms, if 100 welfare recipients enroll in the program then three months later we observe that about 20 of them remain dependent (point A). If they had not enrolled in the program, we

[5] Unless the administrator has an unusually long tenure, the improvement will come too late for her peace of mind, but it might prevent her successors from suffering from the same nightmares.

would expect about 50 to remain dependent (point B). The difference, 30 people becoming independent, is the impact of the program. This means that the program has reduced the caseload by 30, three months later.

Figure 1 also shows that the impact of the program diminishes over time. Ten months after the start of the program, its impact is only 7, that is, the caseload is only 7 lower than it would have been in the absence of the program.

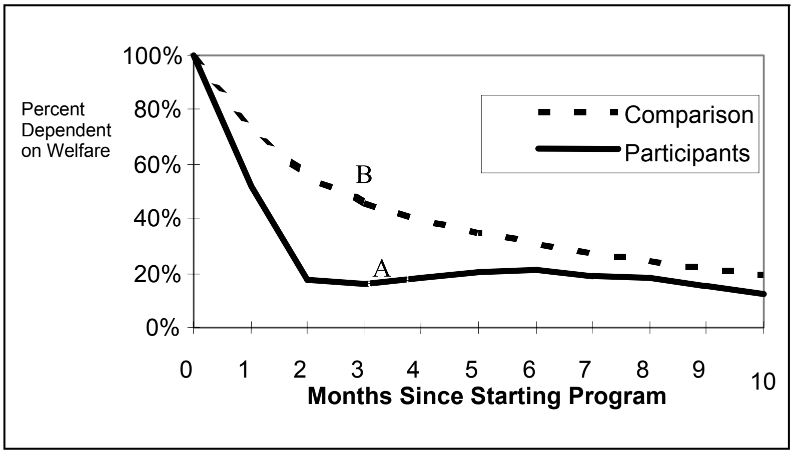

Figure 1: Percentage dependent on welfare, Participants and Comparison group, hypothetical program

Now that we have an estimate of the program's impact we can make statements about the benefits of the program in terms of, for example, reduced welfare caseloads and expenditure. The savings in month three equals the extra cost that the welfare agency would have borne if those 30 additional people had remained dependent. The total savings due to the program equals the sum of the savings in each of the months.

In a similar way, we could make calculations of the increased incomes of the participants, the increased success of their children in school, increased tax revenue and so on. It is this type of information that we need in order to make funding decisions.

Why is estimating impacts difficult?

We need to know the impacts of programs in order to make funding decisions. To estimate impacts, we need to estimate what would have happened in the absence of the program. This is difficult because it is impossible to know what would have happened to any *individual* if he/she hadn't entered a program. However, statistical methods allow us to estimate the impacts on *groups* of individuals. The next four paragraphs explain the basic statistics behind estimating impacts.

If we select a group of individuals (a sample) from a population[6] at random, and observe some characteristic (e.g. height, age or income) then the average of this characteristic for the sample will be about the same as the average for the population, and the larger the sample, the closer its average will be to the average of the population. However, the more the characteristic varies between individuals, the further the average of the sample may be from the average of the population. Although these relationships are intuitively obvious, it is only within the past century that they have been proven mathematically and the relationships between the average of the group and the average of the population determined. Notice that that the statements were made, just as the mathematicians made their proof, without regard to what we are taking the average of. The average height, weight, income, percentage with blue eyes, percentage wearing blue jeans, percentage married or the percentage dependent on welfare will all be approximately the same

[6] We call the larger group, from which the sample or subgroup is selected a "population" to avoid confusion in terms.

for the group selected as for the population from which they were selected, *so long as the selection is random*. That is, every member of the population must have an equal chance of being included in the sample.

How big do the selected groups or samples have to be? The answer depends on the variance (variability between individuals) of the characteristics in the population, and on the degree of confidence required. As an example, the percentage of males in a selected group of 400 will be within five percentage points of the percentage of males in the population, 19 times out of 20 (95% confidence).

This relationship is essential for estimating program impacts. If we randomly select two sufficiently large groups of individuals from the same population, then the average value for any characteristic (e.g. employment, income, welfare dependence, etc.) will be similar in the two groups. If one of these groups participates in a program, and the other does not, then the average earnings, employment, and welfare dependence of the non-participants will be a reliable estimate of what the average earnings, employment, and welfare dependence of the participants would have been in the absence of the program. So the difference between the participants (often called the treatment group) and the non-participants (often called the control group) gives an estimate of the impact of the program.

This method of estimating impacts is usually referred to as random assignment. It is the only method that can be proven mathematically to give unbiased estimates of the impact of programs[7]. Random assignment has the additional advantage of being easy to understand. If the participants in a program have truly been selected randomly, then estimating the program impact is as easy as calculating the average outcomes for the treatment and control groups and finding the difference between them.

Random assignment is not a panacea

Although random assignment has sound theoretical justification and appealing simplicity, there can be many practical problems in using it to produce useful estimates. I list five of these problems below. Those interested in a more thorough discussion should see the articles by Heckman and Smith (1995) and by Burtless (1995).

1) Where there is good reason to believe that a program is effective, and individuals have a legal or moral right to participate, then random assignment may not be appropriate, since in a random assignment study some individuals who are otherwise eligible, will not participate. This is not normally an issue in training programs, since usually there are many more individuals eligible for training than spaces available. The American National Research Council's Committee on Youth Employment Programs concluded that, "in situations in which program resources are scarce and program effectiveness unproven, it [random assignment] is ethical." (Betsey et. al. 1985, page 30.)

2) People who have been assigned to the treatment group may not participate in the program, and people who have been assigned to the control group may participate in either the program or one similar to it. This will tend to make observed outcomes more similar for treatment and control groups, so that the estimate of the impact of the program produced by comparing the average for the treatment group with the average for the control group will understate true impacts of programs.

3) The impacts of random assignment programs may not be the same as impacts from similar programs that do not involve random assignment. This can occur for two main reasons. First, administrators will be acutely aware that the impacts of their programs are being estimated if random assignment is used. Consequently administrators who feel that their programs are working well may be more likely to volunteer to be part of a random assignment study, and once the program is running, they may exert additional effort knowing that they are being monitored. Second, the individuals who participate in random assignment studies may in fact be different from those who participate in programs in which random assignment is not used. Generally, program administrators tend to select individuals who they feel will

[7] The randomization does not have to be caused by the program. Two stage techniques can take advantage of a random process that is unrelated to the estimating-impact exercise. Nonetheless, estimates produced using two stage techniques should be viewed with extreme caution. See Section 3.

benefit most from the program, while random assignment tends to select participants typical of the average member of the population. If program administrators can accurately identify and select as participants those individuals best able to benefit from the program, then random assignment studies, which measure impacts for average members of the population, will show lower program impacts than studies that do not use random assignment.

4) Random assignment is often unpopular with program administrators because it explicitly interferes with or replaces their usual selection and referral process which is based on staff and client judgment. This type of judgment-based referral may be seen by staff as an important part of a program.

5) Because random assignment studies involve referral to the program as well as tracking the outcomes for the treatment and control groups, they can take longer and cost more than studies that do not involve random assignment.

If not random assignment, what?

For these reasons, random assignment is often not the first choice for estimating the impacts of training programs. But if random assignment is not used we need another reliable method. An obvious alternative is to select from the remaining population (the non-participants) a sample of individuals who appear to be similar to program participants, and use this sample as a comparison group. The difference between the average for the participants and the average for the comparison group would then provide an estimate of program impact. This method sounds simple, but its accuracy suffers whenever the comparison group differs from the participants in some way that has not been noticed or measured. Unfortunately, it is very common for there to be unmeasured differences between participants and non-participants. For instance, among welfare participants, individuals who have been incapacitated in some way (for example those with a chronic physical illness) are less likely to take training and also less likely to move into employment on their own. Differences in subsequent employment between program participants and a comparison group of non-participants might result from higher rates of illness in the comparison group rather than from the program itself. Also, individuals who are more highly motivated are generally felt to be more likely to enter training programs and more likely to become employed. If after receiving training, participants do better than non-participants, we will be left wondering if they did better as a result of the training or because they were more highly motivated to begin with. Estimates that are wrong because they falsely attribute the impacts of characteristics of participants to programs are said to suffer from selection bias.

Clearly the danger of selection bias depends partly on the quality of data that the researcher has to work with. If the researcher can record incapacitation or motivation, then he/she can draw a comparison group that is truly comparable to the participants and the estimates will not be biased. The danger of bias also depends on the nature of selection into the program. If selection criteria are stringent or the program is particularly demanding, then participants may be particularly motivated (and hence different from non-participants) so that estimates of impacts would tend to overestimate program effects.

Selection bias has received a great deal of attention in estimates of the impacts of training programs, largely as a result of the experience of the US in the early 1980's. There, different researchers, using the same data, and estimating the impacts of the same programs, arrived at qualitatively different estimates of the impacts of the Comprehensive Employment Training Act (CETA) programs. A blue ribbon panel concluded that the problem was due to selection bias and recommended that the impacts of subsequent training programs be estimated using random assignment.

Although the problem of selection bias has received the most attention, it is by no means the only potential problem in studies of program impacts. For this reason, program administrators and other users of estimates of impacts need a checklist with which to assess the reliability of estimates. The next section presents such a checklist.

Five questions that administrators should ask

One source of nightmares for senior administrators is the general lack of credibility of estimates of impacts. To avoid disappointment, administrators need to become more sophisticated critics of study methods. Five critical questions, and criteria for evaluating answers to them are given below.

1. Were *impacts* studied? Does the study compare outcomes for the treatment group (participants) and a comparison or control group (non-participants)?
2. Is the comparison group valid? If the study is not based on random assignment, were the pre-program characteristics of the participants and the comparison group carefully compared? (This means comparing welfare dependence, Employment Insurance dependence, earnings and employment using *monthly* (not annual) data.)
3. Was the survey response rate adequate? That is if the study was based on a survey, was the response rate *honestly measured* at 80% or more?
4. Was regression analysis used thoughtfully? That is, was the functional form tested?
5. Was the validity of two stage methods tested? That is, if a two stage technique was used, was there separate sensitivity analysis performed?

Were impacts studied?

A surprising number of studies still only report outcomes. ("Eighty percent of the people who went through our program got jobs!") As discussed earlier, measuring what happens to people after they participate in programs is only half the task of estimating the impacts of programs. If we want to know whether the program did any good or not, whether it affected the caseload or not, whether it increased incomes or not, or any other interesting question, impacts rather than outcomes must be studied. That is to say, the estimate must be a comparison between what happened to participants and some estimate of what would have happened to them in the absence of the program.

Is the comparison group valid?

Selection bias is the main source of uncertainty in estimates of the impact of training programs. It occurs when characteristics of the participants rather than (or perhaps in addition to) the program itself cause differences in the outcomes of interest. If we had information on all characteristics of participants and non participants, then we could compare the outcomes for participants with the outcomes for non-participants with the same characteristics, and we could be confident that the difference in outcomes would be due to the program and not to characteristics of the participants.

Of course, it is impossible to measure all the characteristics of an individual, particularly with administrative data. However, it may not be necessary for two reasons. First, it may not be necessary because some characteristics will not affect the outcomes of interest. Second, it may not be necessary because some observed characteristics may be good enough proxies for others that are unobserved, so including both is unnecessary. For example, if motivation affects employment then it will have affected employment in the past. Comparing participants to non-participants with similar employment histories may make it unnecessary to measure motivation directly. In assessing the impacts of real programs, the question of which characteristics must be included becomes an empirical one.

In order to be valid and reliable, estimates must use a comparison group of non-participants with similar histories of employment, welfare, and UI dependence as participants. These history comparisons must be based on monthly data. That is, participants who are unemployed in the month before the program must be compared with non-participants who are also unemployed in that month. Annual data are not precise enough, and tend to introduce bias. With annual data, participants who were employed in the year before they entered the program are compared with non-participants who were also employed in the previous year. But people most commonly enter training programs when they are unemployed. So, of the people who were employed in the previous year, those who lose their jobs in the current year are more likely to take training. But those who have lost their jobs are more likely to be unemployed in the future. As a result, when annual data is used program participants are less likely to be employed than the comparison group, and so studies using comparison groups based on annual data are unlikely to be reliable.

An example of this was provided by Warburton (1996a). There, a program that had a modest positive impact when correctly evaluated using monthly data, appeared to have a large negative impact when the comparison group was drawn based on annual data. The annual-data comparison group was invalid because it contained people with less history of unemployment than the participant group.

Was the survey response rate adequate?

If the estimates are based only on survey data, then at least 80% of the sample should have responded to the survey. If the response rate is less than this, then we will encounter a special form of selection bias usually called non-response bias. Those who respond in a survey tend to be different from those who don't respond, and typically we get different response rates in the treatment and comparison groups. Again, if we find that participants had higher incomes on average than non participants, we will be left wondering whether the difference was due to the program or to pre-existing differences between the respondents and non-respondents.

In a test of the seriousness of non response bias, Warburton (1996b) found that a survey with a 75% response rate generated positive and statistically significant estimates of program impact, when full information (based on monthly administrative data) indicated that the programs actually had no impact. Now that computers have become so inexpensive and computer data bases so widespread problems with non-response bias can often be avoided by the use of administrative data. Where survey data is used, it is essential to achieve a response rate of at least 80%, and to test for differences in relevant characteristics between respondents and non-respondents.

Was regression analysis used thoughtfully?

Regression analysis is often used to produce estimates of the impacts of training programs. In regression analysis, the functional form (an algebraic relationship between the outcome of interest and various characteristics of the individual) is specified and its parameters estimated. For example, if income is thought to be affected by age, schooling and program participation, then the relationship might be:

Income = $a_0 + a_1 \cdot \text{age} + a_2 \cdot \text{schooling} + a_3 \cdot \text{program participation}$.

The parameters to be estimated are a_0, a_1, a_2, and a_3. Regression analysis is useful in the estimation of program impacts because, if we don't have selection bias or non response bias, and the functional form is specified correctly, then the estimate of program impact will equal a_3. That is, if the average age and schooling is the same for participants and non participants, then the regression analysis estimate[8] of a_3 will equal the difference in average income between the participants and the comparison group.

Regression analysis is an old and well established technique. (Gauss is generally credited with its development almost 200 years ago.) In addition, because of the wide availability of computers and appropriate computer programs, it is very easy to produce regression estimates. Unfortunately, we have also been aware for quite a long time that estimates can go seriously awry when the functional relationship between the variables is mis-specified. (See e.g. Bryant and Rupp, 1987 for an example in employment and training programs.) For this reason, where regression analysis is used, tests of the functional form should be included[9].

[8] Using ordinary least squares

[9] e.g. by using an 'F' test for changes in the coefficients when the coefficients are estimated separately on high-income and low-income subsamples.

Was the validity of two stage techniques tested?

Within regression analysis, there is a class of techniques for eliminating selection bias called two stage techniques. One of these techniques, two-stage least squares, is well established, having been developed more than 50 years ago. It works as follows. When assignment to a program is random, differences between treatment and control group gives an estimate of the impact of a program. When assignment to a program is biased, but includes random[10] components, a technique known as two stage least squares can generate unbiased estimates of program impact. For example, if referral to a particular case officer is random, and some workers assign a higher percentage of their clients to programs, then two stage least squares can generate unbiased estimates of program impact. Unfortunately, it is very difficult to calculate the variances of estimates produced by two stage least squares, and computer programs report an approximation called the *asymptotic* variance, which is only true if the sample size is infinite. In most cases these understate the true variance by a wide margin, so researchers imply that their estimates are much more reliable than they really are. Consequently, policy makers should steer clear of estimates based on this technique unless an accurate estimate of the variance is included. (For a discussion on estimating variances in two stage techniques see Cragg, 1967.)

Also, within the class of two stage techniques there are a number of methods that purport not to need any random component. These are of theoretical interest and worthy of further development by academic researchers, but at this stage are not sufficiently reliable for empirical work on which program decisions will be based.[11].

How can we improve our estimates?

Most administrators will be discouraged by the number of papers the above checklist will lead them to discard, and may wonder what they can do to improve the general quality of the estimates. Three things come to mind.

First, we can provide reliable estimates of program impact to serve as benchmarks against which to compare other estimates. To produce these benchmarks, we must (as the Americans have) use random assignment in the design of programs from time to time.

Second, we must give researchers access to data, and lots of it. This could be accomplished by linking administrative data on employment, incomes, UI and welfare dependence. It would be very important to protect the confidentiality of the data, and to specify the uses to which linked data would be put, but reasonable limits would still permit useful new research.

Finally, if researchers are to make complete estimates of the impacts of programs as they actually operate, they need time. It simply isn't realistic to expect conclusive estimates of impacts in fewer than five years,[12] although indications of program success can often be produced within the first few years.

Estimates of the impact of BC's programs in the late 80's

This section reports estimates of the impact of four types of training program:

1. Classroom training

---

[10] Technically, the factors that affect assignment into the program do not have to be purely random but only uncorrelated with the outcome of interest.

[11] Not enough time has passed since these techniques were developed for us to have the necessary level of confidence in the techniques. For example, a technique developed in the early 70's was found to give unreliable results in a paper published in 1983. If we use the latest techniques we may find in five to ten years that the results weren't valid.

[12] Our favourite example of unrealistic timeframes is the apocryphal story of the group that was required to set up a prenatal nutrition program and report on its effectiveness within six months.

2. Job clubs
3. On-the-job training in the private sector
4. On-the-job training in the public sector

With the exception of the estimates of the impact of classroom training, these estimates get a "yes" for all relevant questions (not all studies used survey data) and so are quite reliable. The classroom training estimates get a "yes" to all of the questions except the first. The estimates are based on monthly data, but it was not possible to control for histories of employment or UI dependence when these estimates were made. Unfortunately, the new results were not available in time for inclusion in this essay.

A more important caveat comes from California's experience with GAIN. Even if programs in other jurisdictions appear to be the same as these, they may have substantially different impacts.

Classroom Training

In British Columbia, about 4,000 welfare recipients participate in classroom training at public institutions at the post secondary level. They account for less than 5% of the 115,000 FTE students in the public post secondary system. Many more welfare recipients participate in upgrading and training courses and private institutions. Participation in classroom training is significant from the point of view of the welfare system, but it accounts for only a small portion of the post secondary education system.

The conventional wisdom in Canada is simply that classroom training is not an effective tool for helping disadvantage workers. The Canadian federal National Institutional Training Program Evaluation (Abt, 1985 page 7) concluded that classroom training participants "show no significant benefit from participating in the training." This conclusion has been reinforced by American studies. For example, LaLonde in his review of the American literature, (1992, page 2) concludes that "when training raises the earnings of the economically disadvantaged, the gains are modest in size." And Barnow (1987) concludes "public service employment and on-the-job training were generally found to be more effective than classroom training."

Until about five years ago I shared this view. A 1986 BC study found that classroom training had no impact on subsequent welfare dependence (Jamieson 1987). That study tracked the welfare dependence of welfare recipients who enrolled in a course of classroom training beginning in 1982. But that study, like most studies of classroom training, was limited by data which lumped all types of training together. Information on the type of training, and even whether the individuals had shown up for and participated in the training, was not available.

However, some more recent work, using data that identifies the type of classroom training received, gives us more reason to be optimistic. First, the Urban Institute's evaluation of Massachusetts' ET Choices (Nightengale, 1991) provided separate estimates of the impact of three types of classroom training, English as a second language (ESL), upgrading or adult basic education (ABE), and vocational training. It found no impact on subsequent welfare dependence from either ESL or ABE, but significant impacts from vocational training.

This finding encouraged us to approach Camosun College in 1989 to look for similar results using disaggregated data. They agreed and we undertook a joint study. The study followed 1,388 individuals who were dependent on welfare at the time they enrolled in training.

- 760 enrolled in adult basic education (ABE). These courses, from basic literacy and numeracy to high school equivalence, were all at the secondary, rather than post-secondary, level;
- 169 enrolled in career-technical training. Courses were directly job related and normally lasted 24 months. Examples include criminal justice, visual arts and electronic technology.
- 339 took vocational training. Courses were directly career-related and normally lasted nine months. Examples include plumbing, welding, secretarial and dental hygiene.
- 120 enrolled in academic courses. These are university level courses.

Results of our study for each type of classroom training are presented below.

Adult Basic Education

The study tracked the welfare dependence of the participants over the five years following enrollment. This is shown by the solid line in the top half of Figure 2 below. After five years, about half were still dependent on welfare. In order to estimate the impact we had to make an estimate of the number who would have been dependent on welfare if they hadn't entered the program. This was done by tracking a group of non-participants who were judged comparable, over the same period. This is shown by the dotted line. The impact is the difference between the two lines. For ease of reading the difference is graphed on the lower part of the page.

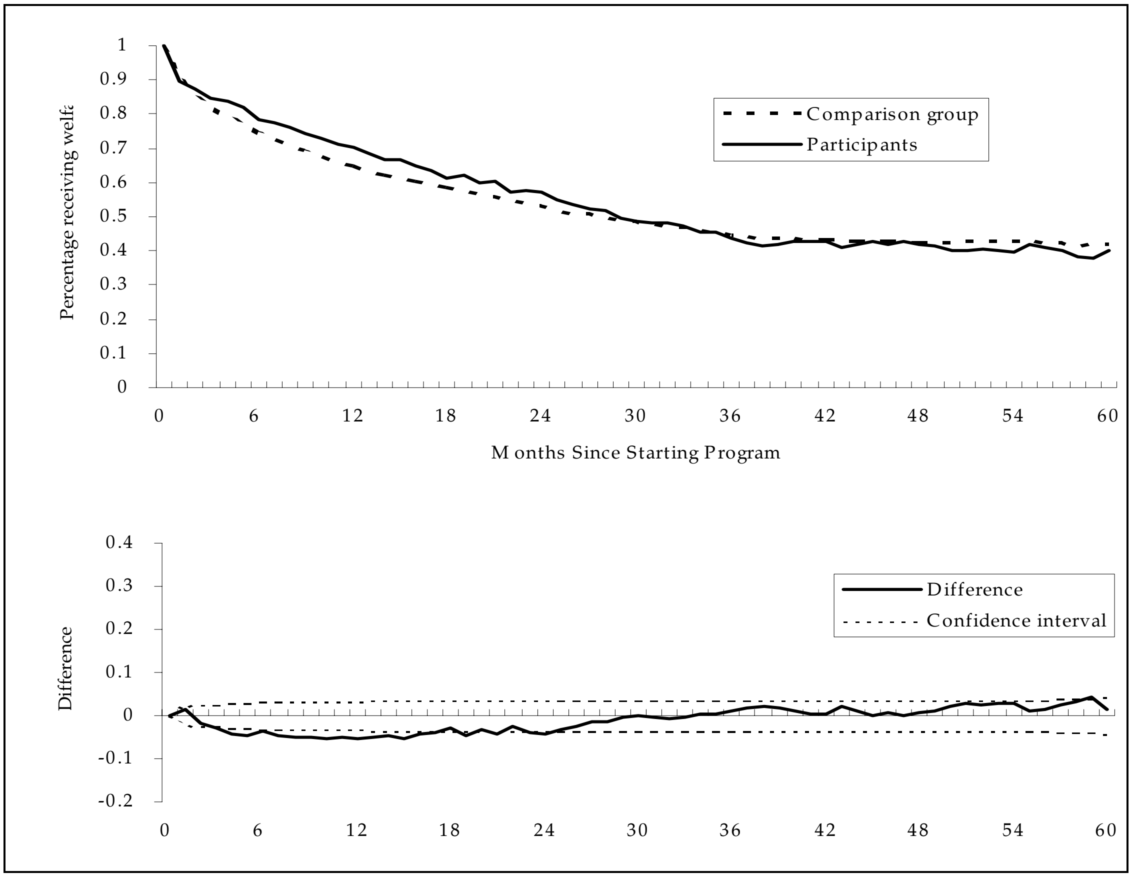

Figure 2: Percent dependent on welfare: ABE participants and comparison group

The results for ABE were similar to the results found in the Urban Institute's ET Choices study. ABE had no apparent impact on the subsequent welfare dependence of participants. The increased dependence in the first 18 months is understandable. Participation in training would reduce the intensity of job search. However, the absence of an impact even after tracking the individuals for five years is clearly disappointing.

Career-Technical

In direct contrast to the results for adult basic education, the study found the impact of career-technical training was large and sustained. Figure 3 shows the onset of an impact at around month 24, the time at which the participants would complete the course. The impact is undiminished three years later. It appears that career-technical training has a permanent impact on the welfare dependence of participants.

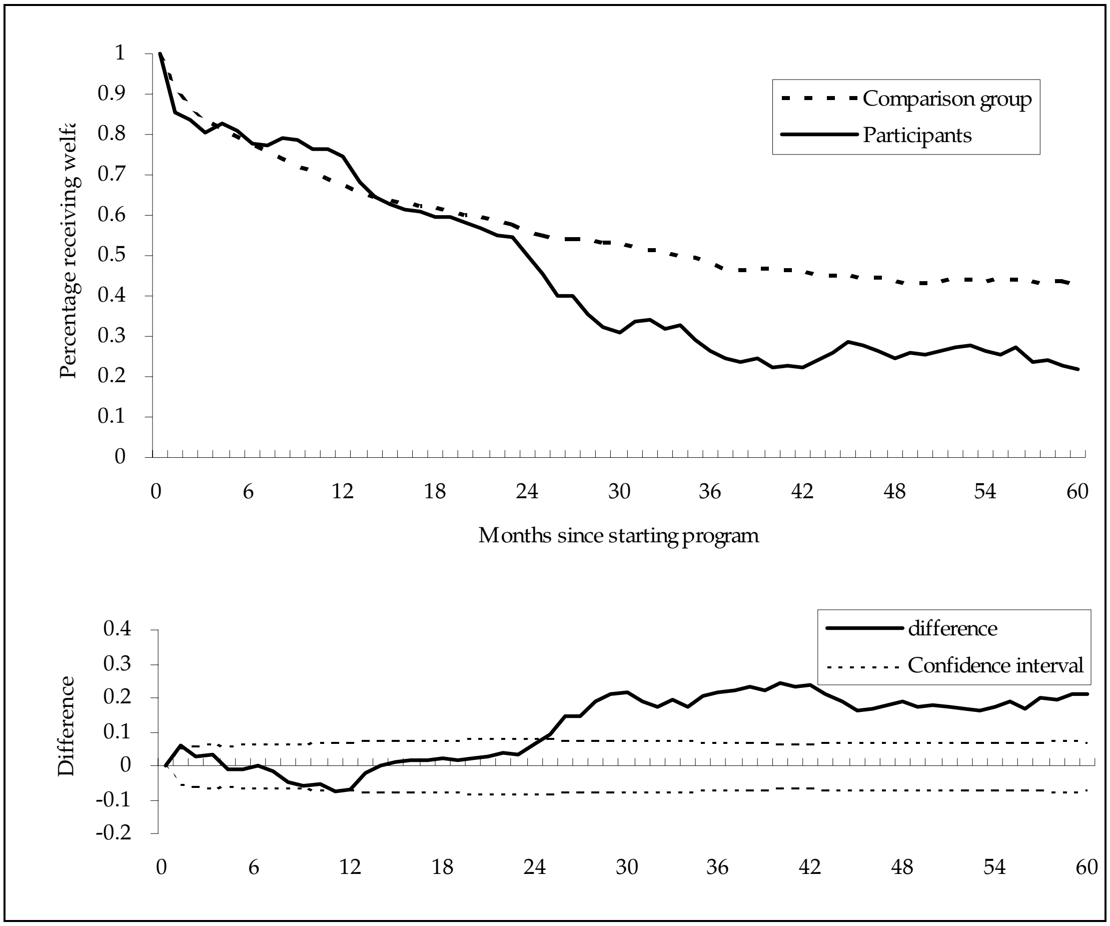

Figure 3 Percent dependent on welfare, career-technical training participants and comparison group

Vocational Training

The study found that vocational training, like career technical training, showed a large and sustained impact. Positive effects became apparent about nine months after the beginning of training, coincident with the normal completion time, and were sustained over the next three years.

Although no attempt to quantify the financial costs and benefits of these programs has been made here, it should be noted that the brevity of vocational training will improve its attractiveness to both participants and funders alike for two reasons: the cost is lower than for career-technical or academic, and benefits are realized more quickly. So, although the study found that career technical training had a bigger impact than vocational, it also found that vocational training might provide more help for a given amount of money.

Academic

Figure 5 shows the comparable results for academic training. Again strong positive effects are found, lasting throughout the five-year period studied (although the impact and the sample size were smaller, putting the estimates of impact at the edges of statistical significance). I speculate that the impacts in months 1 to 8 and 12 to 20 result from movement to Student Financial Assistance during the academic year, and back onto welfare during the summers.

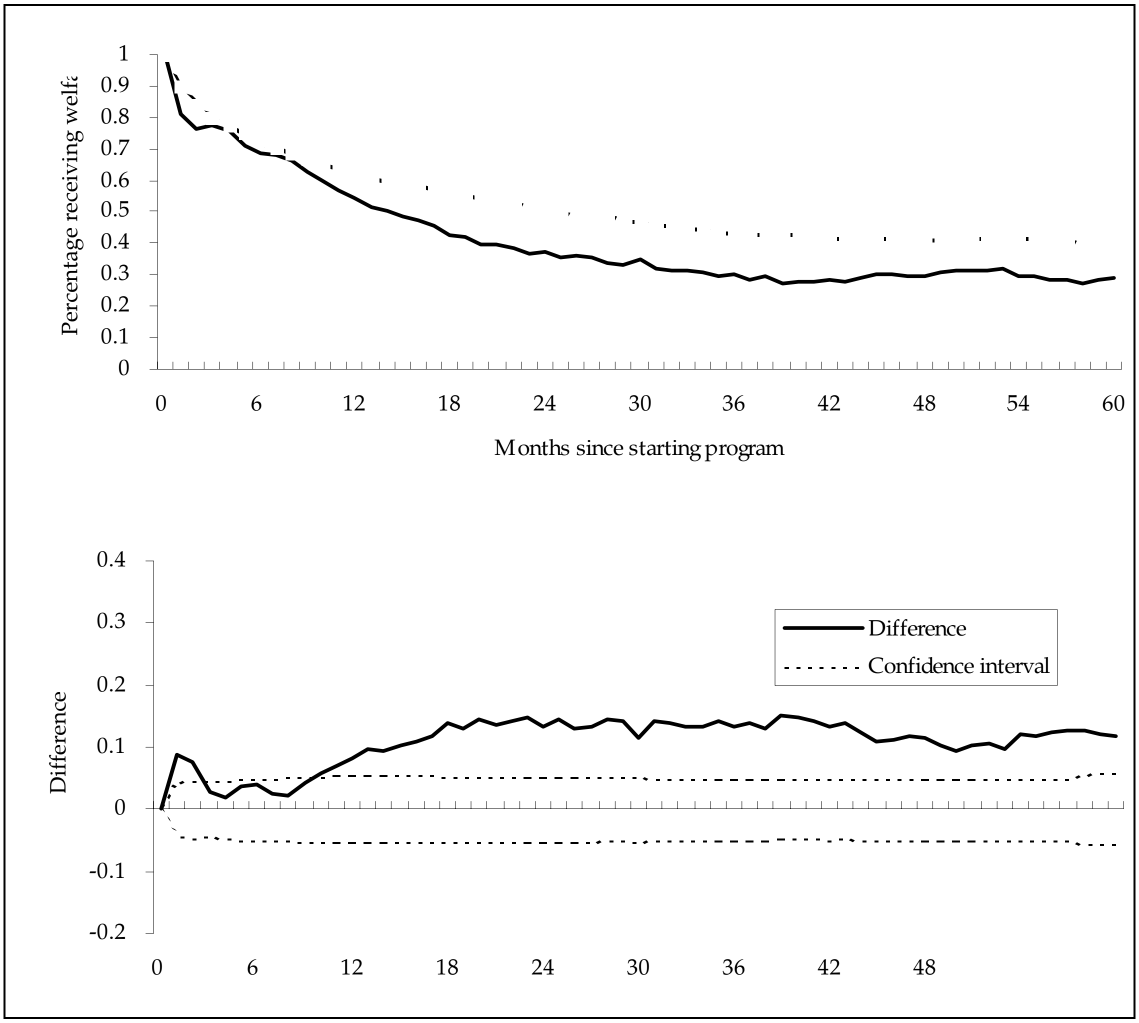

Figure 4 Percent dependent on welfare, vocational training participants and comparison group

Overall conclusion for classroom training

Contrary to the conventional wisdom, this research shows that some forms of classroom training can be effective in assisting welfare recipients to become independent. As noted above, this does not mean that it will be effective for all groups and in all jurisdictions.

Job Clubs

Job clubs are probably the most common type of training offered to the unemployed across North America. Although they can vary in duration and content, they have three essential ingredients. First, they use a group setting, usually involving 10 to 15 individuals. Second, they provide some job search preparation, such as information on the job market, techniques for finding jobs, career counseling, interview and job search skills and resume preparation. Third, participants are given targets for the intensity of their job search, for example, to contact 10 employers per day. Success in achieving these targets is monitored for each individual and reported to the group as a whole. Job clubs typically last for from three to six weeks.

Job clubs have repeatedly been shown to be effective at speeding participants' return to employment. In their study for Human Resources Development Canada., Crémieux et al. (1995) found that individuals' job search intensity decreased after they had been unemployed for nine months. They further concluded that after 18 months, for all intents and purposes it had stopped. Job clubs may have their greatest impact by increasing the intensity of job search.

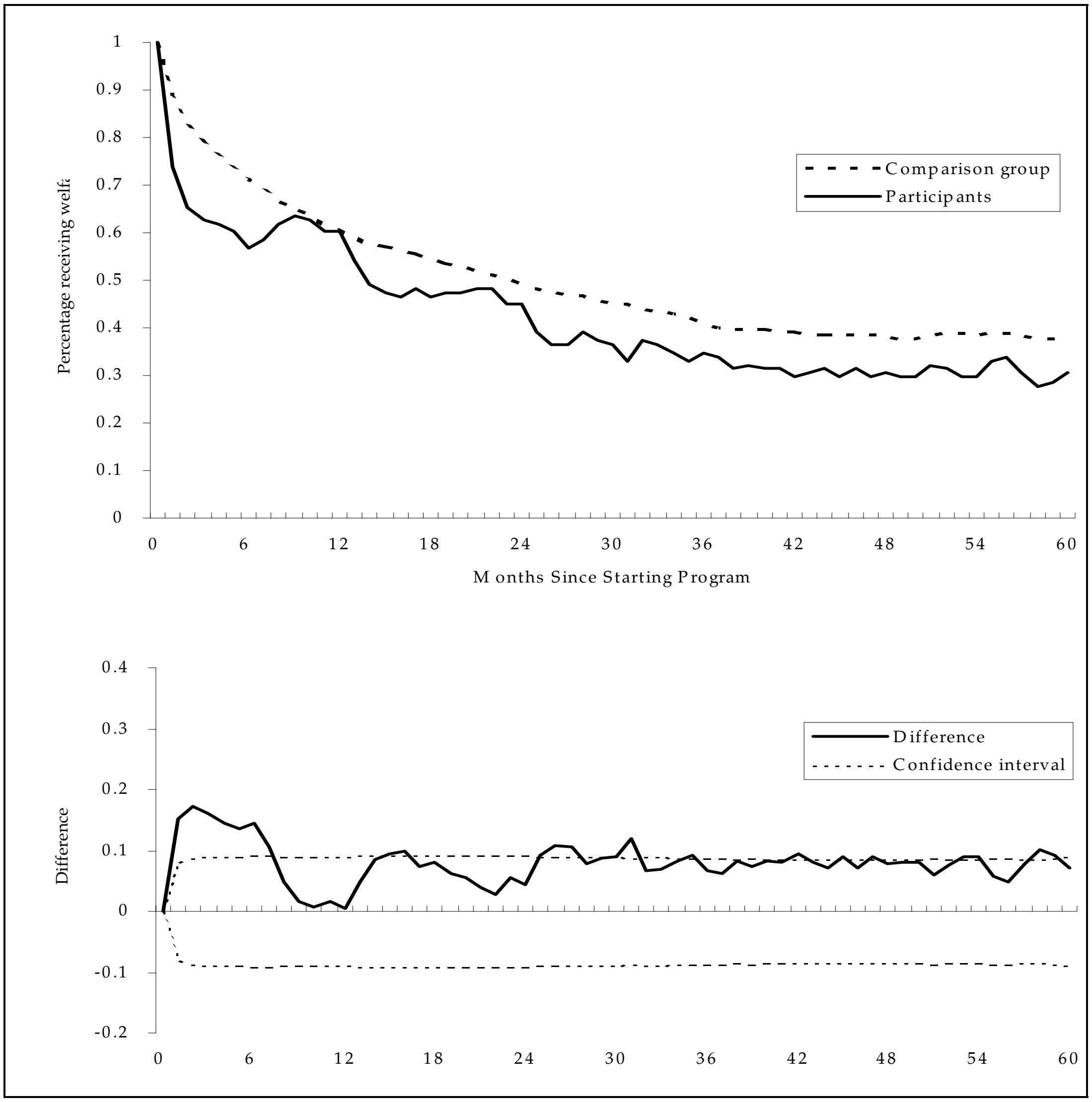

Figure 5 Percent dependent on welfare, academic participants and comparison group

In 1989, BC's Ministry of Social Services (MSS) conducted a job club pilot project. Participants were selected randomly from the pool of eligibles, generating a true control group with which to compare the success of the participants. The subsequent welfare dependence of the participants and the controls is shown in Figure 6. Dependence was reduced by a statistically significant amount in the first few months, but the difference petered out within eight months.

Nonetheless, the reduction in welfare payments saved the Ministry $1.38 for each $1 spent on the job club. (Assuming that the job clubs were full and cost $650 per participant—$500 for the contract plus $150 in participant expenses.)

Despite the general acceptance of the effectiveness of job clubs, they have their detractors who say "participants get jobs faster, but they also lose them faster" and "job clubs force individuals into low paying jobs." This study found that this was not in fact the case. Of the people who left welfare, a smaller percentage of participants than controls returned after one month of independence.

Similarly, the assertion that job clubs induce individuals to take lower paying jobs than they would otherwise was not borne out. The average wage of the 38 participants who responded in a survey was

$8.04. For the 28 controls who responded, the average was $7.94. This difference was not statistically

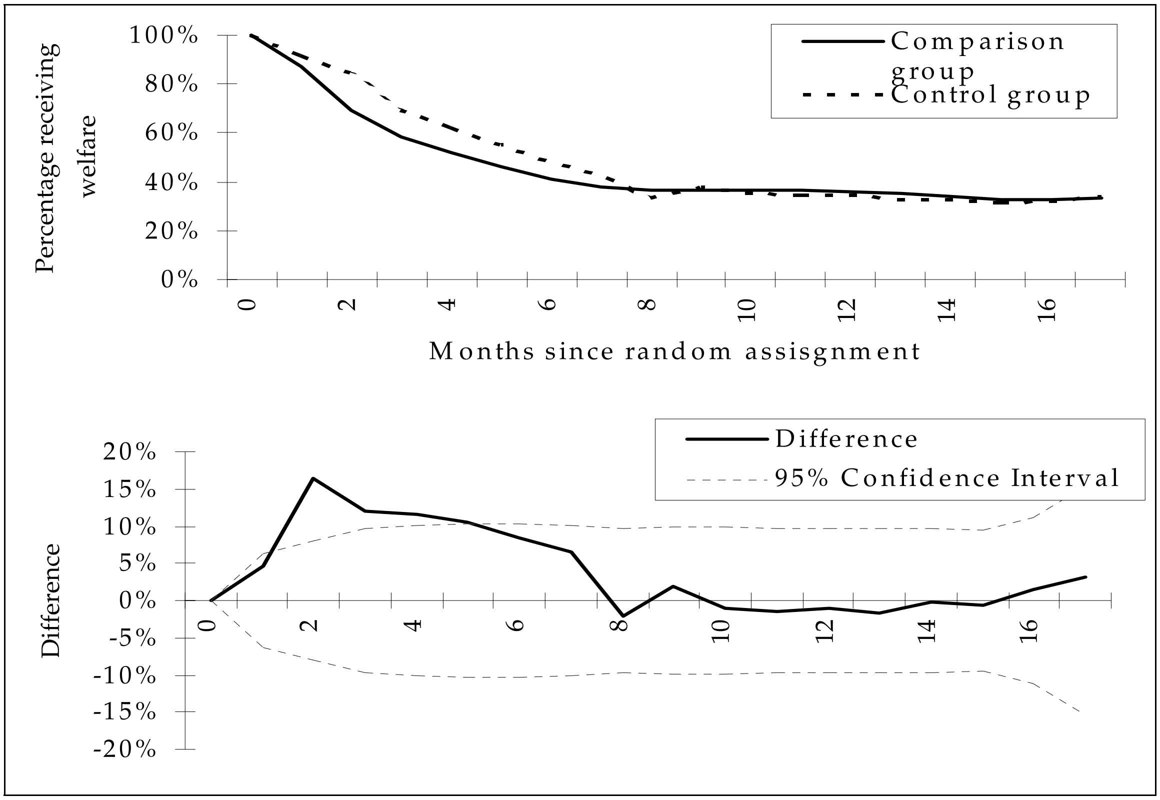

Figure 6 Percent dependent on welfare, Job Action participants and control group

significant.

On-the-job training in the private sector

In the late 1980's the flagship employment and training program for welfare recipients was on-the-job training supported by a wage subsidy program for private sector businesses. The Ministry subsidized of half the wage up to a maximum of $3.50 per hour. The Ministry referred a number of potential candidates to each employer, who hired one.

Figure 7 shows the subsequent welfare dependence of participants and a comparison group. Like participants in vocational training, the subsidized workers left welfare more quickly and were generally less dependent on welfare than the comparison group. Overall, the program saved the Ministry $2.43 for each $1 spent on subsidies.

The subsidized employment generated eligibility for Unemployment Insurance, and did result in increased UI dependence as is shown in figure 8. But even with this taken into account the program returned $1.21 for every dollar spent on either subsidies or additional UI.

Fortunately, this program was big enough to allow separate estimates for people with different characteristics. Figure 9 shows that savings were greatest for families with dependents, especially single parent families. Figure 10 divides the participants according to the number of months of welfare benefits that they received within the previous 25 months. It shows that savings were larger for individuals who had been on welfare longer, although this impact declines slightly for those who have been on continuously for the past 25 months.

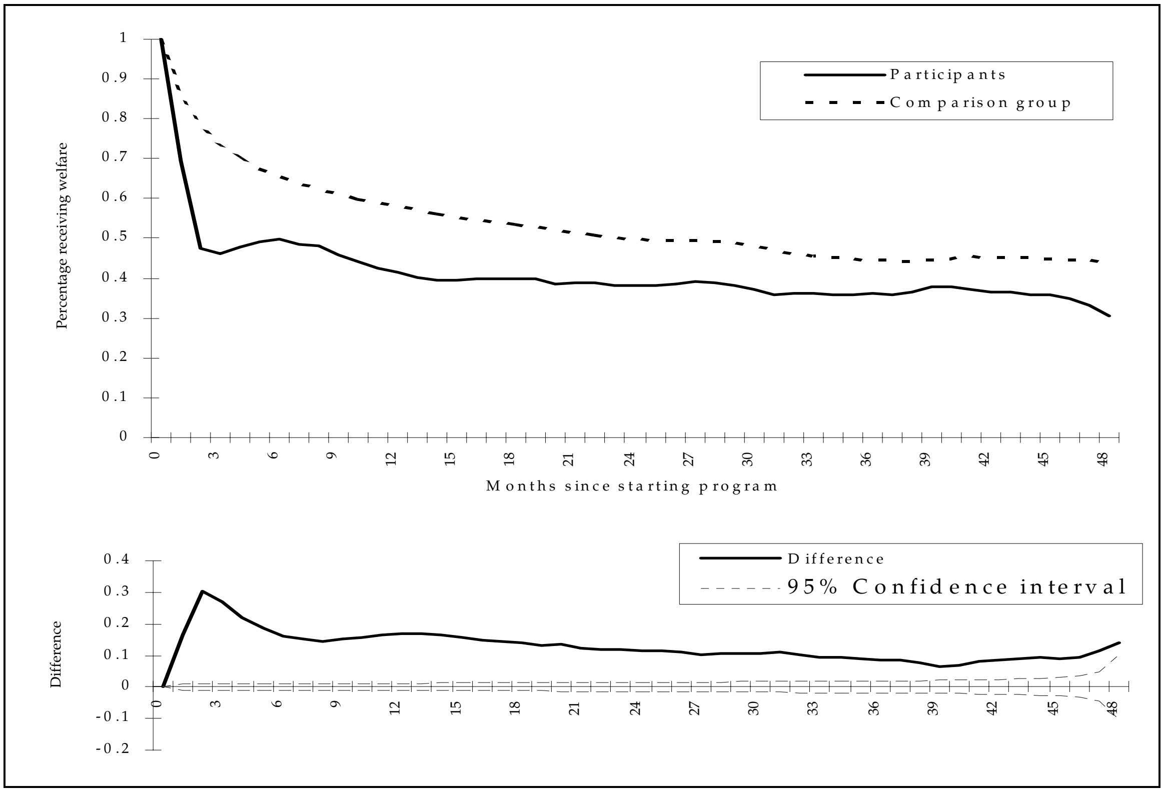

Figure 7: Percent dependent on welfare, EOP participants and comparison group

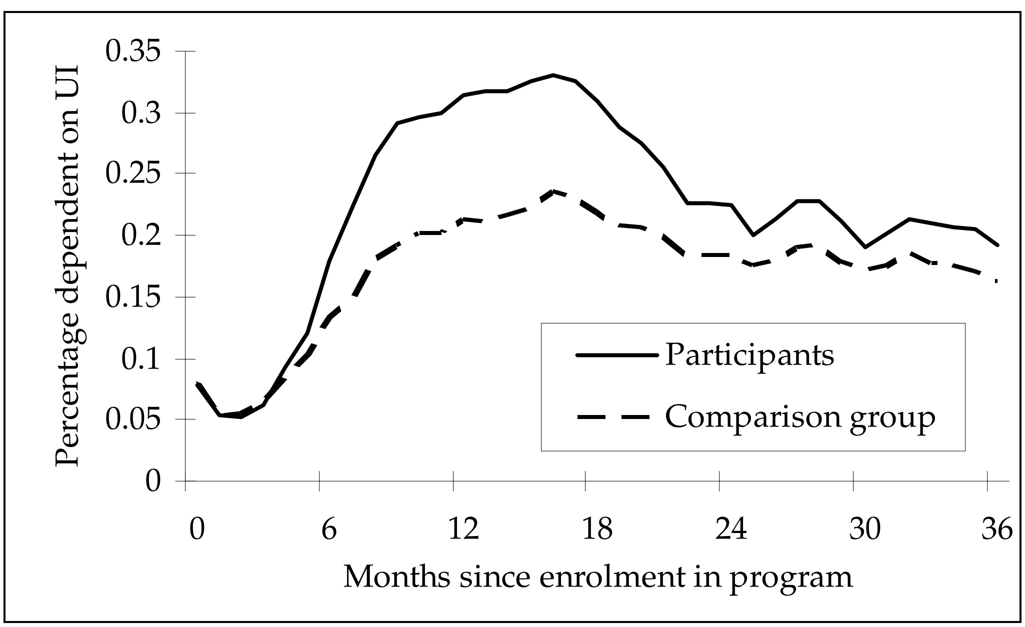

Figure 8: Post program UI dependence of participants and comparison group (EOP)

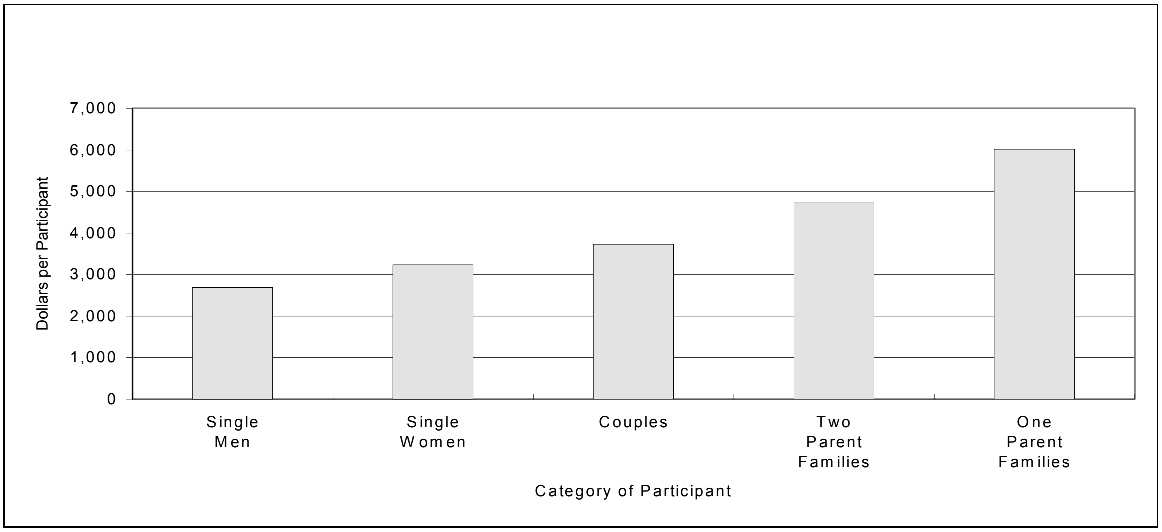

Figure 9: Welfare savings by category of participant

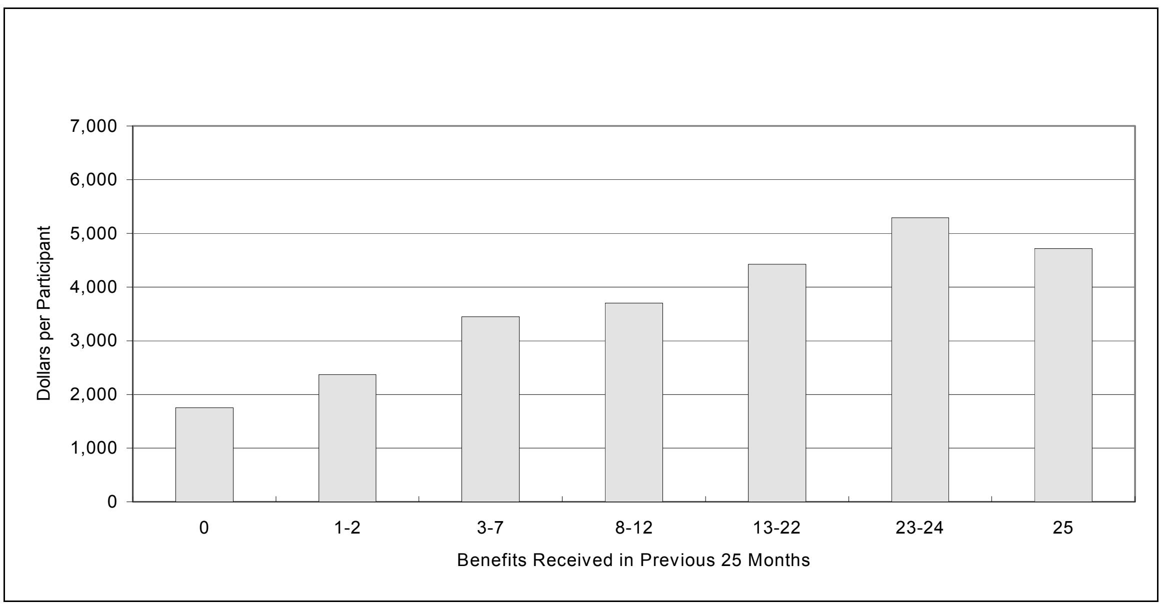

Figure 10: Welfare savings by welfare history

Is displacement a problem?

If the total number of jobs in the economy is fixed, then placing individuals in subsidized employment may simply displace other individuals who may potentially go onto welfare. If this occurs, the impacts listed here overstate the true impacts. Critics of wage subsidy programs point to displacement as a potentially serious problem.

The evidence for displacement is not strong. Over time, the number of jobs in the economy is clearly not fixed. Between 1991 and 1995, about 80,000 in-migrants per year entered BC's labour force, and the unemployment rate declined over that period. Princeton economics professor David Card provides a more detailed analysis of the impact of increasing the labour force on employment and earnings in his study of the effect of the Mariel boat lift of Cubans to the United States in 1980. He concludes, "The Mariel immigrants increased the Miami labor force by 7%, and the percentage increase in labor supply to less-skilled occupations and industries was even greater because most of the immigrants were relatively unskilled. Nevertheless, the Mariel influx appears to have had virtually no effect on the wages or unemployment rates of less-skilled workers."

The wage subsidy program is another way of increasing the supply of labour. As mentioned earlier, a study commissioned by Human Resource Development Canada found that people who have been unemployed for a long time stop looking for work. This means (by definition) that they drop out of the labour force[13]. Subsidized jobs put them back in. If in the process, this displaces another individual from employment, the labour force will still be increased by one if the displaced individual actively seeks employment. And if the economy absorbs increases in the labour force, as the analysis of David Card suggests, then the net effect will be increased employment.

Conclusion

The positive results from the wage subsidy program suggest that there is potential for successful programs for welfare recipients that involve placements with the private sector.

Public Sector Employment Programs

In the late 80's the Ministry also ran a number of programs that provided short-term (up to six months) employment in positions created with non-profit societies. The Community-Tourism Employment Training program was one example. Figure 11 shows that the program reduced welfare dependence for the six months of subsidized employment, and reduced welfare dependence for the subsequent 12 months of UI eligibility, but had no impact on long term welfare dependence.

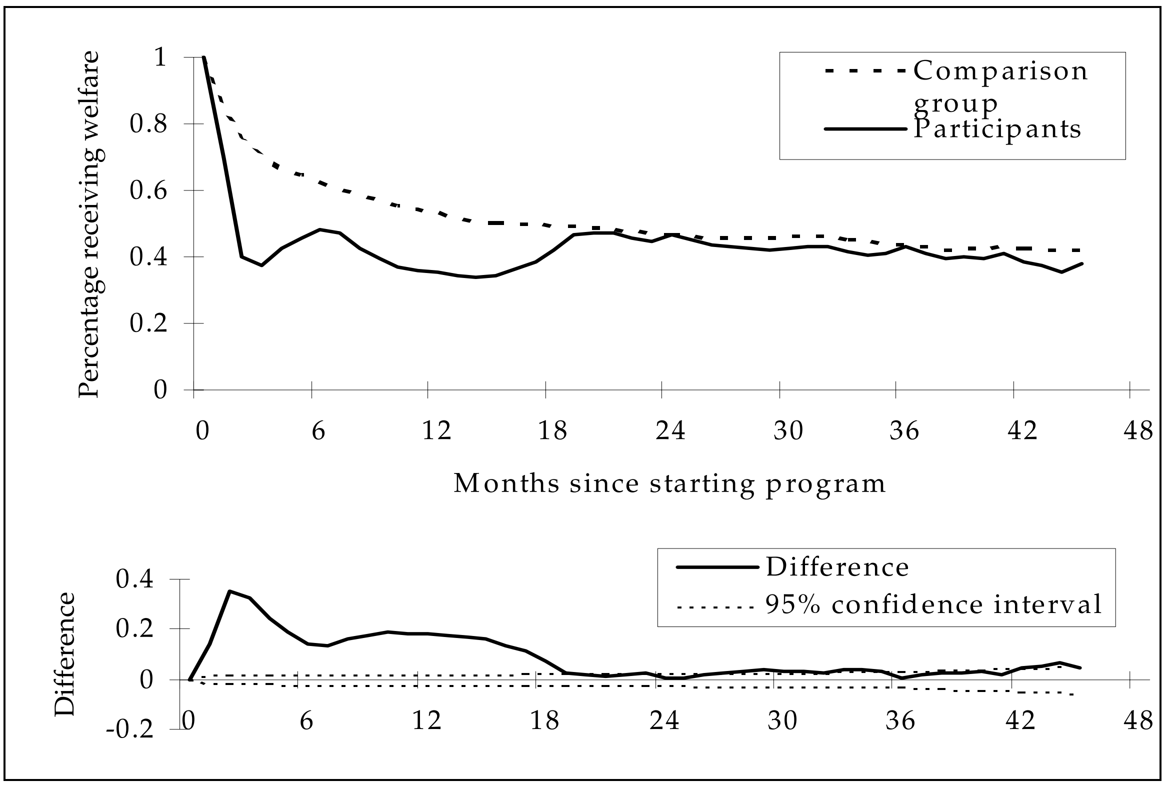

Figure 11: Percent dependent on welfare, CTETP participants and comparison group

[13] Recall that the labour force is defined as those employed plus those not employed but actively seeking employment.

Figure 12 confirms that the decrease in welfare dependence that occurs between months six and 18 is caused by increased UI dependence.

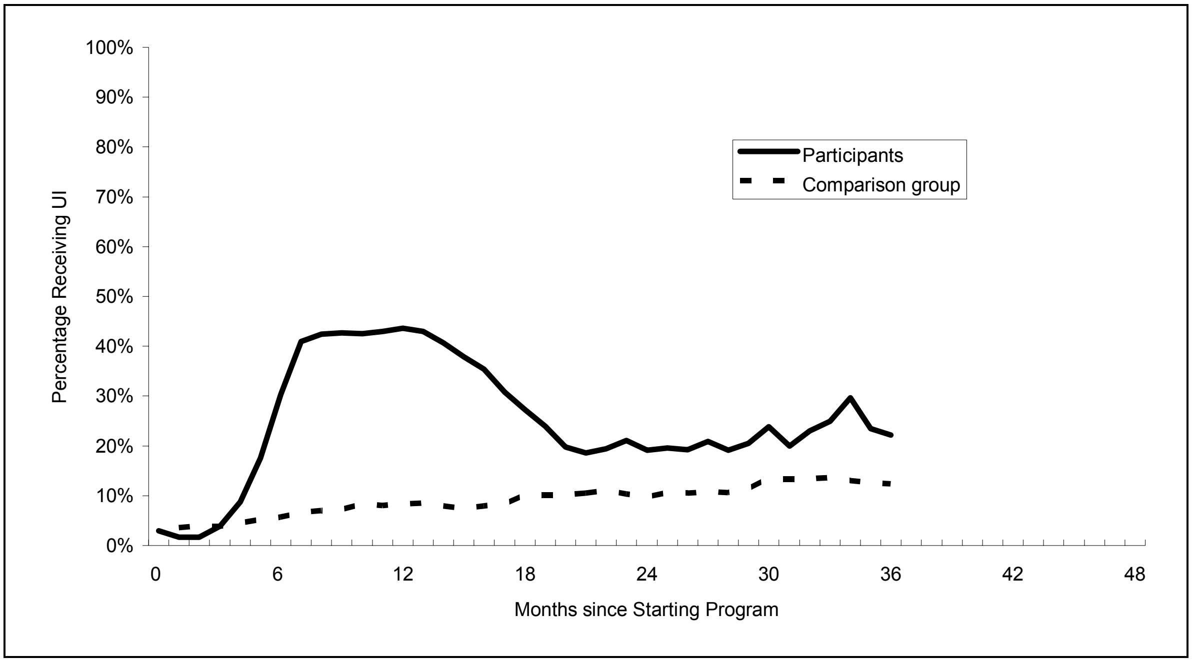

Figure 12: Percentage dependent on UI, CTETP participants and comparison group

Other Benefits

Like Adult Basic Education, public sector employment programs can generate benefits other than reduced welfare dependence. For example, the work that is completed can leave a tangible physical benefit for the community.

Conclusion

I was surprised that public sector employment programs did not have a lasting impact on the dependence of participants. If job search is higher just after the loss of employment as the HRD study indicates, then providing six months of employment should help reattach the participants to the labour force. In addition, I would have expected the employment experience gained on the subsidized job to make the individuals more attractive to potential employers. The fact that it did not have an impact certainly leads to the conclusion that these types of programs must be scrutinized. It may be that the public sector programs simply have the feel of make-work, in contrast to the real jobs provided by the private sector. On the other hand, they may be improved by the addition of elements designed to smooth the transition to work.

Summary and Conclusion

Should the government sponsor training for the disadvantaged? The conclusion of this research is that government should only sponsor training for the disadvantaged if it conducts ongoing, reliable assessment.

The evidence from programs in BC is that some types of training can genuinely help people become independent while reducing expenditure for government. This suggests that government *should* sponsor training for the disadvantaged.

However, experience also shows us how foolish it would be to generalize from any one set of results. Recall that the impact of California's % program and the US National JTPA program varied dramatically from site to site. For this reason, no program should be funded simply because a similar program was shown to be effective in a different setting. Ongoing assessment is essential.

We cannot expect instant success. Thirty years of sponsoring training programs without feedback has left the field open to charismatic, well-meaning, sincere trainers who pitch their programs with religious fervor. In the absence of knowledge of the impacts of programs on the participants, selection of training programs can be affected by the intensity and fervor of the proponents, a rather unreliable guide. To undo this, we must provide clear goals, and feedback to service providers on their success in achieving these goals. We cannot expect that the effects of this long neglect of the need for evaluation evidence will be undone instantly.

Finally, we need to increase our confidence that our estimates of impacts are reliable. This would be facilitated by a set of guidelines for the reliability of estimates. The development of these guidelines would be facilitated by the greater use of random assignment, and by making administrative data more widely available to researchers in and out of government.

Does training make a difference? Properly designed and evaluated, it can. Agencies expending public funds for training have an obligation to ensure that the programs they sponsor live up to their potential.

References

Abt Associates Canada, 1985, Evaluation-National Institutional Training Program Final Report, Toronto

Barnow, Burt S., 1987. "The Impact of CETA Programs on Earnings - A Review of the Literature," *The Journal of Human Resources*, XXII, 2.

Bryant, Edward C., Kalman Rupp, 1987. "Evaluating the Impact of CETA on Participant Earnings," *Evaluation Review,* 11, 4.

Burtless, Gary, 1995. "The Case for Randomized Field Trials in Economic and Policy Research" in *Journal of Economic Perspectives* Vol.9, No. 2.

Cragg, J. 1967 "On the relative small sample propoerties of several structural-equation estimators" *Econometrica,* 35: 89-110

Crémieux, Pierre-Yves, Pierre Fortin, Paul Storer and Marc Van Audenrode, 1995 *Unemployment Insurance and Job Search Productivity,* Human Resources Development Canada

Dickinson, Katherine P., Terry R. Johnson, Richard W. West, 1986. An Analysis of the Impact of CETA Programs on Participants' Earnings, *The Journal of Human Resources,* XXI.

Goldgerger, Arthur S. 1983. "Abnormal Selection Bias," in S. Karlin, T. Amemiya and L. Goodman, eds. *Studies in Econometrics, Time Series and Multivariate Statistics*, Stamford, Academic Press

Griliches, Z., 1986. "Economic Data Issues" in Z Griliches and M Intrilligator, eds. *Handbook of Econometrics,* Vol. 3 Amsterdam: North Holland.

Heckman, James J. and Jeffrey Smith, 1993. "Assessing the Case for Randomized Evaluation of Social Programs" in Karsten Jensen and Per Kongshoj Madsen, eds. Measuring Labour Market Measures, Denmark, Ministry of Labour

Heckman, James J. and Jeffrey A. Smith, 1995. "Assessing the Case for Social Experiments" in *Journal of Economic Perspectives* Vol.9, No. 2.

Jamieson, Joan (1987) *Individual Opportunity Plan Evaluation* Research, Evaluation and Statistics Branch, Ministry of Social Services and Housing, Victoria, BC

LaLonde, Robert J., 1992. *The Earnings Impact of US Employment and Training Programs*. Paper presented at the Canadian Employment Research Forum Workshop, Ottawa, March 1992

Nightingale, Demetra; Douglas Wissoker; Lynn Burbridge; D. Lee Bawden and Neal Jeffries; 1991. *Evaluation of the Massachusetts Employment and Training (ET) Program,* The Urban Institute Press, Washington, DC

Reubens, Beatrice G. (1980) "Review of Foreign Experience", Chapter 5 in Bernard E. Andeerson and Isabel Sawhill eds. Youth Employment and Public Policy. Englewood Cliffs: Prentice hall, Inc.

Stromsdorfer, Ernst W. et al (1985) Recommendations of the Job Training Longitudinal Survey Research Advisory Panel (Washington: US Department of Labor)

US Department of Labor (1992) Self Employment for Unemployed Workers UI Occasional Paper 92-2

Warburton, William P. (1996a) "What went wrong in the CETA Evaluations?" Canadian Journal of Economics XXIX Special Issue Part 1.

Warburton, William P. (1996b) *Estimating the Impact of Selected Programs on Participants' Subsequent Welfare Dependence and Employment in British Columbia* Ph.D. Thesis, University of London.