PUBLIC TRAINING PROGRAMS IN CANADA: A META-EVALUATION

Derek Hum
Wayne Simpson
University of Manitoba
Winnipeg, Manitoba

Abstract: Canada has a history of training individuals for the labour market, and substantial research has accumulated concerning the effectiveness of training programs. There have been many evaluations of public training programs in Canada, both summative and formative, in the last two decades. What have we learned from these evaluations of training programs? What should we continue to do? What should we try to avoid? This article presents an assessment of the summative evaluations conducted to date in Canada, focusing on three questions: What is an appropriate and feasible methodology for summative evaluation of training programs? Has that methodology been consistently and effectively implemented in Canada during the past two decades? What are the prospects for future evaluations in Canada?

Résumé: Depuis longtemps le Canada entraîne des individus pour le marché du travail, et un grand nombre d'études ont été accumulé à l'égard de l'efficacité de ces programmes de formation. Dans les deux dernières décennies, il y a eu plusieurs évaluations des programmes de formation publique au Canada, des études formatives aussi bien que des études sommatives. Qu'est-ce qu'on a appris de ces évaluations des programmes de formation publique? Que devons-nous continuer à faire? Que devons-nous essayer d'éviter? Cet article présente une évaluation des études sommatives entreprises jusqu'à date au Canada, se concentrant sur trois questions: (1) Quelle est une méthodologie appropriée et faisable pour l'évaluation sommative des programmes de formation? (2) Est-ce que cette méthodologie a été exécutée d'une manière conséquente et efficace au Canada pendant les deux dernières décennies? et (3) Quels sont les espoirs pour de futures évaluations au Canada?

The rationale for training programs is often couched in terms of augmenting individual human capital, which in turn will
lead to better economic performance, such as lower unemployment rates or higher personal earnings. Canada has a long history of assisting individuals to train for the labour market, and a substantial body of evaluation research has accumulated in the last two decades concerning the effectiveness of training programs. Much of this research has centred on programs in the federal sector. In future, however, more training programs will likely be delivered by provinces, as the Labour Market Development Agreements (LMDAs) now being developed represent a new direction for the public provision of training in Canada. Whatever the future auspices of delivery, it seems appropriate to ask: What have we learned from “evaluations” of public training programs? What should we continue to do? What should we try to avoid? There have been many evaluations of training programs in Canada, both formative and summative. This article presents an assessment of the summative evaluations of training programs conducted to date in Canada. We do not consider formative evaluations, as many of these are not in the public domain.

Many objectives are associated with training programs, some political, some social, and some economic. For economic policy, one useful summary measure is the “rate of return” to training. Rate of return measures are useful for individuals contemplating whether or not it is worthwhile to pursue training. Such measures also serve to gauge allocative efficiency for government spending; that is, whether public monies spent on training programs are earning acceptable returns, or perhaps should be used elsewhere. There are, of course, other measures besides the economic one of a rate of return to training, such as regional distribution of public funds. Nonetheless, the rate of return criterion is common to virtually all evaluations of training programs, either explicitly or implicitly. Accordingly, we adopt the rate of return to training as a particular focus in reviewing the evaluation reports.

We structure our discussion to ask the following three questions: What is an appropriate and feasible methodology for summative evaluation of training programs? Has that methodology been consistently and effectively implemented? And what are the prospects for future evaluations in Canada? The next two sections address the first two of these questions; the last section considers the question of future evaluation directions in Canada by offering comments for guidance rather than prediction.
METHODS FOR EVALUATING RETURNS TO PUBLICLY PROVIDED TRAINING

Assessment of past evaluations requires some preliminary remarks on modern evaluation methods and their associated problems and controversies. We begin with a brief discussion of experimental and nonexperimental approaches, and then use insights from this examination to assess the Canadian research on public training programs.

Evaluation of training programs involves a comparison of program costs and benefits. Direct costs are typically available from program administrative files. Indirect costs, such as the forgone earnings of those undertaking training, can be estimated from the earnings of program nonparticipants with similar characteristics. Program benefits can then be estimated by comparing the outcomes of program participants with comparable nonparticipants acting as a control group. In short, program benefits net of costs (or simply returns) are compared for participants and nonparticipants. However, estimation of program benefits is complicated by the fact that factors other than the program(s) in question may affect labour market outcomes. Therefore, methods of analysis that ignore these factors or incorrectly specify their influence will produce misleading estimates of outcome effects.

Evaluation of training in Canada has concentrated on nonexperimental data because such evidence can be readily obtained at low cost. However, the use of nonexperimental data raises a number of concerns. Some experts claim that the benefits estimated from nonexperimental data are likely to be less reliable than estimates obtained from a properly designed experiment in which eligible participants are randomly allocated to the program and control groups (Ashenfelter, 1987; Burtless & Orr, 1986). Two major problems with nonexperimental research evaluation designs are (a) the choice of an appropriate control group and (b) the specification of the model and the nonprogram factors that affect labour market outcomes. Reliable estimates of program benefits require a control group of comparable individuals who are nonparticipants, as it is only the difference between the program and control groups that can be properly classified as a program effect. The choice of a control group is complex in practice, however, and the properties of the chosen control group are critical to the success of the estimation of program impacts (Heckman & Robb, 1985). Even if a suitable control group is found, the estimated
program effects will still be biased if the models, typically multivariate regression models, are misspecified either in terms of the nonprogram factors that affect earnings (such as personal and family background variables) or the process by which the program and control groups were selected (Burtless, 1995, pp. 68–73).

Experiments that randomly allocate eligible participants to program and control groups can potentially solve both of these problems and improve the reliability of the estimated program benefits. A randomized experiment relieves the analyst of the onerous task of constructing a comparable control group from those who did not participate in the program, as the control group is derived through random allocation. It also eliminates the need for a complete specification of the underlying nonprogram factors that generate labour market outcomes, because the program benefits from a properly designed experiment can be estimated by direct comparison of the change in outcomes between the program and control groups, and should not be sensitive to nonprogram factors included in the analysis (Keeley, 1981). Analyses in Canada of experimental data of this type include the negative income experiment (Hum & Simpson, 1993b) and, more recently, analysis of the impact of a temporary earnings supplement for income assistance recipients in the Canadian Self-Sufficiency Project (Card, Michalopoulos, & Robins, 1999). In each case, the analysis used variations of the “differences estimator,” which subtracts the change in labour market outcome of the program group from the change in outcome of the control group to obtain an estimate of the program effect. Neither experiment, however, focused on training per se.

Given the paucity of true experiments in Canadian policy research, and none that were designed specifically to assess training outcomes, the essential question is whether nonexperimental studies can still give trustworthy results. Some researchers in the U.S. have examined this question in detail. Lalonde (1986) uses the National Supported Work Demonstration experiment to assess the reliability of nonexperimental estimates of training benefits. After obtaining the estimates of the training benefits using the differences estimator, he discards the experimental control group and draws a control group from sources typically used in nonexperimental evaluations of training programs in the U.S. He then uses conventional econometric modelling approaches with nonexperimental data to assess their reliability in terms of the experimental results. He concludes pessimistically that
the econometric estimates often differ significantly from the experimental results ... this evidence suggests that policymakers should be aware that the available nonexperimental evaluations of employment and training programs may contain large and unknown biases resulting from specification errors. (p. 617)

If nothing else, these findings warn us that evaluations from nonexperimental data must be done with considerable care.

As experimental evaluation results are expensive and not easily generalizable (Bjorklund, 1987; Heckman & Smith, 1995, pp. 94–107), policy makers must often rely on evaluations from nonexperimental data. No one has persuasively argued that such evaluations are worse than no evaluation at all. Moreover, Heckman and Robb (1985) and Heckman and Hotz (1989) emphasize that program evaluations from nonexperimental data can, in principle, be accurate. Heckman and Hotz (1989) specify six generic models involving a levels estimator, a differences estimator, and a difference of differences estimator, each with variants that do, and do not, allow training impacts to vary across individuals. They show how a series of tests on the preprogram data, the postprogram data and the model restrictions can be used to identify nonexperimental models that produce estimates of training impact comparable to those obtained from experimental data, albeit with larger standard errors. It is interesting to note, however, that different nonexperimental estimators are required for the two groups they analyze — high-school dropouts and women on welfare. Thus, there is no evidence that there exists a single appropriate estimator, such as the differences estimator, that would consistently produce accurate estimates. Moreover, Friedlander and Robins (1995) show that specification testing alone will not produce nonexperimental estimates that correspond to experimental results under commonly used practices for obtaining a comparison group of nonparticipants, particularly the conventional practice of choosing a comparison group from another region without the training program.

Moffitt (1991) offers a sensible primer on the use of nonexperimental data to estimate the benefits of training programs. He argues that the crucial evidence to identify program impacts involves the search for a valid comparison group and the collection of longitudinal data involving multiple observations before and after the training intervention. This will permit nonexperimental model specifications to be
tested in the fashion described by Heckman and Hotz (1989). Moffitt views the process of testing as crucial to the accumulation of knowledge about what information to collect in future nonexperimental evaluations, and where experimental trials might play a useful role in understanding which training programs are effective:

One can imagine a sequential search strategy in which a particular data collection protocol is adopted, the results analyzed, and tests for over identifying restrictions conducted ... and the key sources of remaining sensitivity to different just identifying, untestable assumptions are determined. A subsequent new data collection protocol based on these results could then be devised to test those assumptions and the process repeated. Thus the benefits of a single investigation extend to other studies as well ... [E]xperimental trials could be reserved only for situations in which a sequence of nonexperimental investigations of the type just mentioned is unable to resolve differences ... [or] a randomized trial could be conducted simultaneously with a nonexperimental evaluation to identify ... which nonexperimental data yield the same estimate of program impact as that obtained from the randomized trial. (p. 311)

The important lesson from this debate is that the evaluation of training programs is a complex exercise that requires a careful research strategy. Simple estimates from carelessly designed data collection are unlikely to produce valuable evidence on the returns to training for individuals or society. Evaluation exercises need to be taken seriously both in the design of programs and data collection and in the statistical analysis of the data. Knowledge of what works in evaluation design and analysis must be carefully accumulated and applied to subsequent programs. The potential payoff is that evaluation research will then provide a better guide to what works — for prospective trainees as well as program deliverers. The research may also improve our understanding of how labour markets operate (Heckman & Smith, 1995).

EVALUATION OF PUBLIC TRAINING PROGRAMS IN CANADA: PAST AND PRESENT

How far has evaluation strategy advanced in Canada? What have we learned about the outcomes and returns to Canadian training
programs? Evaluation of training programs in the federal government was mandated by Treasury Board in 1977 and became a formal part of the planning process with the introduction of the policy and expenditure management system in 1980. Since then, Human Resources Development Canada (HRDC) and its forerunner, Employment and Immigration Canada, have produced a variety of evaluations using nonexperimental, primarily administrative, databases. These studies have generally been conducted by private consultants in close consultation with the evaluation branch of EIC/HRDC. Some studies have also been produced in conjunction with provincial governments, particularly in recent years as Ottawa embarked on a series of labour market development agreements that transferred responsibility for the design and evaluation of training programs to the provinces. However, these evaluation efforts have largely remained an internal government process with limited involvement from the academic social science community, unlike evaluation research in the U.S. (Riddell, 1991, pp. 52–53) This situation probably accounts for the absence of an experimental approach to training evaluation in Canada.

In order to review and assess the publicly available evaluation studies, it is useful to group the evaluation studies into three time periods: the “first generation” or early evaluation research (1981–85) prior to the implementation of the Canada Jobs Strategy; the “second generation” or mature evaluation research (1986–94) of the Canada Jobs Strategy; and recent evaluation research (1995 to date).

The initial response of Employment and Immigration Canada (EIC) to the introduction of the policy and expenditure management system was a series of evaluations of the Canada Manpower Industrial Training Program (CMITP) and the Critical Trade Skills Training Program published between 1981 and 1985. As might be expected, these early studies exhibit serious flaws consistent with the state of evaluation research at that time. In particular, benefits were determined only by comparing the preprogram and postprogram earnings of training participants; there was no comparison group and no attempt to control for observable factors that might have affected program participation. As a result, the estimated benefits, or program impacts, from these studies are not credible today. Although the analyses are primitive by current standards, the focus was in our view the right one: benefits are estimated, program costs to government and individuals are measured, and benefit-cost ratios are calculated. Hum and Simpson (1996a) have recently shown that the
information provided in these reports, if credible, could be used to estimate rates of return to government-sponsored training; for example, the evidence in the evaluation of the CMITP indicated a favourable rate of return of 27% to training the unemployed. As a result, the early evaluations were a promising beginning to evaluating the returns to public training programs in Canada.

The systemic flaws in the assessment of training benefits were largely eliminated in a series of evaluation reports issued between 1985 and 1994. Beginning with the project report of the evaluation of the National Institutional Training Program (NITP) (EIC, 1985a), the outcomes of training program participants were compared to a “control” group of individuals who did not participate in the program but who were, in other relevant respects, deemed comparable to those who did participate. The impacts of the program were estimated using a “differences” estimator that compared the change in the earnings or other outcomes (principally “employability,” generally defined as the proportion of time spent in employment) of the participant and comparison groups. A variety of control variables, including variables to correct for sample selection bias arising from the nonrandom selection of the participant and comparison groups (Heckman, 1979), were also included to test the sensitivity of the impact estimates. As a result, the estimated benefits from these “second-generation” reports are much more credible than earlier attempts.

Nonetheless, the comparison group for the evaluation of the NITP was disappointingly small (only 200 compared to 4,000 program participants). This may have affected the statistical reliability of the results. The document Evaluation of the Job Development Program: Final Report (EIC, 1989), for example, used program and comparison samples of 3,983 and 1,500, respectively, and subsequent evaluations (EIC, 1992; HRDC, 1991, 1994a, 1994b, 1994c, 1994d) all provide a respectable balance between the size of the comparison and participant samples. These studies also relied on the differences estimator of program impacts, and many corrected for selection bias and performed other specification tests that enhanced the credibility of the estimated training benefits.

An important limitation of these studies is that they follow training participants and the comparison group for a very short period of time, never exceeding two years. As a result, little is known about the permanence of the benefits that are estimated. In some cases,
the estimated positive benefits of training may decline over time, as is suggested in the SARS evaluation report (HRDC, 1994d). In other cases, the estimated negative or zero benefits of training may be attributable to the short time period between graduation from the training program and the follow-up survey, as is suggested in the reports of the Canada/Newfoundland and Canada/New Brunswick Youth Strategies and the CEEMI Sectoral Training Fund (HRDC, 1994a, 1994b, 1994c). No study has been conducted to answer these questions. We simply do not know whether the effects reported are long lasting rather than temporary, or delayed in detection, or both.

These reports provide some indication of which training programs “work.” Still, several programs reported statistically insignificant, sometimes even negative, estimates of program impacts. With respect to earnings, these include the NITP (EIC, 1985a) (which may be attributable to the small comparison group used), the Job Development Program except for men in the Individually Subsidized Jobs option, and the Canada/New Brunswick and Canada/Newfoundland Youth Strategies (1994a, 1994b). On the other hand, the Skill Shortages Program (HRDC, 1991), the Skill Investment Program (EIC, 1992), and the Employability Initiatives for SARs in CJS (HRDC, 1994d) provide reasonably compelling evidence of significant positive impacts that should inform policy makers that these programs have something to offer.

Unfortunately, the potentially most useful and easily communicated information about these apparently successful programs — namely, their rates of return — is not calculated. The Skill Shortages Program and the Skill Investment Program evaluation reports provide direct program costs per trainee, but do not provide an estimate of the forgone earnings of trainees necessary to estimate benefit-cost ratios or, indeed, to examine “cost-effectiveness” as they claim to do. The evaluation report of Employability Initiatives for SARs in CJS, on the other hand, correctly estimates program costs as direct administrative/course costs plus the opportunity cost of training time, which is estimated from the earnings of the comparison group over the average duration of the training program. Indeed, the report provides a very clear, model methodology for the evaluation of the costs and benefits of training programs that should serve as the basis for future evaluation reports in this area.

The estimated social costs and benefits of training in the evaluation report of Employability Initiatives for SARs in CJS are used to pro-
vide benefit-cost ratios that are difficult to interpret and therefore provide poor guidance for decisions to allocate training resources. The benefits are evaluated over five years, discounted at 7.5% per year, under two scenarios: first, the estimated training impacts decline uniformly to zero over five years, and second, the estimated training impacts persist over five years.¹ Using this approach, the report calculates social benefit-cost ratios for the two scenarios of 2.0 and 3.2 for survey data and 1.2 and 1.9 for administrative data. As the appropriate cost and benefit data have been estimated, however, we can calculate the social rates of return to the Employability Initiatives for SARs in CJS. The returns are 41% and 74% over five years for the survey data under the two scenarios, and 15% and 37% for the administrative data. Because the report effectively assumes that training benefits do not exceed five years, these very attractive rates of return to the program are potentially understated for the second scenario. In addition, it would have been useful to have estimates of the after-tax earnings impact of training and the after-tax forgone earnings of trainees to calculate private rates of return, to inform and attract prospective program participants.

Compared alongside the evaluation research published between 1989 and 1994, recent evaluation reports of public training programs must be viewed as disappointing. Several evaluations are basically formative and contain no outcome assessment, including Integrated Training Centres for Youth (ITCY) (1996), the Community Skills Centres (CSC) (1997), and the Employment Benefits and Support Measures (EBSM) (1998). In the case of the ITCY and EBSM programs, summative evaluations of program objectives were “planned” but have not been made public. In the case of ITCY, the summative evaluation was scheduled for 1997; the EBSM summative evaluations will be conducted under the labour market development agreements with the provinces. In the case of the CSC program, a telephone survey of past program participants was conducted, but no summative evaluation appears to have been planned.

Other evaluations, which were not explicitly formative, made little attempt to estimate training outcomes. The Evaluation of the Investing in People Initiative — Year Two Final Report (1996) examined only reliance on social assistance under the rationale that the goal of the program was self-sufficiency. However, the report admitted that the postprogram follow-up period was too short to permit graduates of the program to find employment. Similarly, the Final Evaluation Report of the Sectoral Partnership Initiative (1997)
examined only program take-up and effectively abandoned the planned summative evaluation after an initial round of data collection.

The best recent evaluations of training programs are *Nova Scotia Compass (NSC): A Summative Evaluation Final Report* (HRDC, 1997) and the *Employability Improvement Program (EIP) Evaluation* (HRDC, 1995). The NSC evaluation was designed and analyzed by outside consultants who also assisted HRDC in the development of a primer on quasi-experimental evaluation (HRDC, 1998). The primer reviews the issues associated with the estimation of training outcomes from nonexperimental programs that we discussed above. Indeed, many of these issues were addressed in the evaluation of training programs between 1989 and 1994. Therefore, from that perspective, the NSC evaluation itself is disappointing. Although the evaluation provides two years of preprogram data on program participants and a matched sample of nonparticipants, the postprogram follow-up is limited to one year and includes only about one quarter of the initial sample. As a result, although they find higher employment and reduced social assistance reliance attributable to the training and job experience programs in NSC, the authors can say little about the durability of these effects. The disappointing follow-up evidence, in terms of both quality and quantity, likely also accounts for the lack of statistically significant evidence of program effects on earnings. There is an interesting lesson here, however. Although a simple comparison of the earnings of participants and nonparticipants one year before and after the program suggests that training is beneficial — a difference estimator without controls yields a very impressive estimate of $1,765, or about 58% of the preprogram earnings of participants — the result does not stand up to further econometric testing. The estimated impact of training is generally insignificant and varies from positive to negative when various model specifications, involving linear regression and difference estimators, are used. As with previous evaluation reports, no attempt is made beyond this point to determine which estimators are acceptable in the fashion advocated by Heckman and Hotz (1989) and Moffitt (1991), likely because the follow-up data proved to be disappointing.

The follow-up survey for the EIP evaluation was considerably better both in coverage and in length, averaging 66 weeks from the end of the program. A control group was constructed from a random sample of the administrative files of individuals who were potential cli-
ents for EIP, based on contact with a Canada Employment Centre or a UI claim during the preprogram period. In comparison with this control group, the estimated change in annual earnings for graduates of the training programs in the first year was 54% for job opportunities, 42% for project-based training, and 39% for purchase of training. These estimates are based on regression models that include sociodemographic characteristics, labour force history, and contextual variables (involving the timing of programs and data collection). The models are also tested for selection bias arising from nonrandom program participation. Although this is clearly the most useful recent evidence on the effect of training programs, more testing of the models along the lines suggested by Heckman and Hotz (1989) would have been useful to increase the credibility of the results. Further, no data on program costs are presented and no estimates of the returns to training are carried out. For example, if the report had emulated the approach of the Employability Initiative for SARs (HRDC, 1994d) and provided information on the earnings of nonparticipants during the program period, it could have been used to estimate training costs properly and thereby to estimate the social rate of return to training

PROSPECTS FOR FUTURE EVALUATION RESEARCH IN CANADA

The focal point for government intervention in training revolves around the existence of gaps between the benefits and costs of training that accrue to individuals and society. An activity may be costly to provide to one individual but, once provided, may be virtually costless to extend to others. An example of such a “public good” is information on the benefits and costs of adult education and training to individuals, or simply private rates of return. The government can contribute to efficient resource allocation by collecting and providing this information to society at large as a public-good service. In addition, adult education and training may confer benefits on society that exceed benefits accruing to individuals. As a result, individuals will tend to underinvest in these activities without government support, such as subsidies to increase investment levels, until the social rate of return equals the opportunity cost of capital.

Both these traditional arguments for government intervention require detailed information on the private and social rates of return to adult education and training that are currently not available. No recent evaluation of public training programs provides estimates of private or social returns. In order to do so, an evaluation would re-
quire preprogram and postprogram data on trainees and a comparison group to analyze the benefits of training, in addition to detailed information on program costs. The private benefits of training are represented by the gain in after-tax earnings, and the social benefits are traditionally represented by the gain in before-tax earnings. Hence, preprogram and postprogram earnings data should be gathered on both before- and after-tax bases. The private and social direct costs of program participation should be collected, where the indirect costs of program participation in terms of foregone earnings are estimated from the earnings of the comparison group during the program period. These data requirements are not onerous, but would provide the basis for the assessment of the private and social returns to training.

In the process of gathering this information, policy analysts might also be able to investigate the barriers to training. For example, one barrier to training and re-employment may be the high implicit rates of taxation on social assistance recipients, the so-called welfare wall. This will be apparent if the gains in earnings from training are collected both before and after tax; it may also be the case that the welfare wall is a greater problem for some groups, such as lone parents, than others. In this case, it would be useful to collect information systematically, where sample sizes permit, on the returns to training across different groups: lone parents, youth, unattached adults, the unemployed, the disabled, and so on. The 1981 evaluation of the CMITP, for example, estimated separate returns for employed and unemployed individuals, and those with special needs, and showed that the returns to training differ substantially across these groups. Another possible barrier is access to funds to finance adult education and training. We can learn something about this barrier if we collect data on the private and total costs of training across individuals, particularly if we use the data to construct prospective costs for training and use it to analyze training participation decisions in nonexperimental data.

The collection of this data in a systematic fashion will be a challenge for governments in Canada. Following Moffitt (1991), we suggest three stages of data collection. First, it is possible to use existing longitudinal surveys to analyze the benefits and costs of adult education and training programs. Hum and Simpson (1996a) illustrate how the Labour Market Activity Survey (and hence its more ambitious successor, the Survey of Labour and Income Dynamics) can be used to determine the benefits of public training programs. They
compare the earnings between 1988 and 1990 of the unemployed who received training under the Canada Jobs Strategy in 1988 with those unemployed who did not receive such training during the same period. They find that earnings rise for all workers over time, due partly to inflation, but that earnings for men with training grew by 73% between 1988 and 1990 compared to only 31% more for those without training. For women, the results were just as impressive: earnings grew by 98% for trainees compared to 53% for nontrainees. These gains translate into estimated returns to training of 34% for men and 41% for women if typical training program costs are used. Even if three quarters of all jobs acquired by trainees involve displacement, the returns still would exceed benchmark rates in the range of 10% that are conventionally cited.

More sophisticated analysis of earnings gains from training before and after tax, a determination of forgone earnings of the comparison group, and a determination of the direct costs of training programs from administrative files would permit a determination of private and social rates of return. This analysis could be supplemented by specific questions on the nature of training, its frequency and duration in one of the panels of a future Survey of Labour and Income Dynamics in order to improve estimates of training costs. Similar opportunities in other longitudinal data sets, such as the Youth in Transition Survey and the Workplace and Employee Survey now being undertaken by Statistics Canada, could also be exploited.

Administrative files can also be used. Human Resources Development Canada (1994) compared the labour market standings of 64,000 Canadians who took training under the Canada Jobs Strategy in 1988 with an equal number of selected nontrainees, taking advantage of access to taxation data. The earnings of trainees compared to nontrainees improved by only 1% between 1985 and 1991, implying a low pay-off to training. In fact, although the earnings of trainees were equivalent to those of nontrainees in 1985, earnings were almost 8% lower in 1991. The average earnings of trainees were higher because the participation of trainees in the work force increased by 43% compared to 31% for nontrainees. In this case, training improved job prospects but not necessarily earnings prospects. When the results are broken down by age, the older group was found to have received more benefits from training: older trainees recorded a 13% improvement in earnings compared to a 26% decline for the youngest group of trainees. These results suggest that emphasis on
training youth may not be justified, but this is only one study, and more research is required. Other results implied that trainees in public institutions did better than trainees in private institutions and that trainees in science gained the most, relative to comparable nontrainees followed by trainees in trade and transport and in health. This type of information, especially if kept up to date and corroborated by other researchers and other data sets, would be useful in deciding what types of training program are most effective.

Second, each new training program provides an opportunity to design an evaluation that will provide credible estimates of the returns to training. In this regard, recent trends are worrisome. As we noted above, recent evaluation research has not produced such estimates. Currently, the federal government is completing labour market development agreements that will transfer all or part of the responsibility for the design and delivery of training programs to each province. Responsibility for the evaluation of these programs will be shared by the federal and provincial governments, except in Quebec, where the province will have sole responsibility for evaluation. The agreements call for summative evaluations to assess the impact of training programs on employability, earnings, and dependency on government transfers by applying conventional evaluation methodologies.

One can hope that this process will provide appropriate incentives to evaluate programs using best practices and to share the information broadly, but we also note that previous co-operative agreements in other arenas leave much room for improvement. For example, the Final Evaluation Report on the Sectoral Partnership Initiative (SPI) (1997) provided federal government assistance to the private sector to develop sectoral councils to coordinate training and other aspects of human resource management. Whereas both formative and summative evaluations were initially planned, the summative evaluation was abandoned. Moreover, the councils were clearly not adequately prepared for any meaningful summative evaluation exercise:

Generally, the councils perceive themselves to be accountable to HRDC based on the funding agreements signed. These agreements, however, tend to only have statements regarding specific delivery issues and not broader outcomes. As a consequence, the councils do not perceive themselves to be accountable for the intended
impacts and effects of the SPI ... The lack of buy-in or accountability for the impacts and effects of the SPI has created problems for the evaluation. Most sector councils were unprepared for the demands of the SPI evaluation on their time and the data requirements associated with the evaluation, and there were limited requirements in the contribution agreements to provide evaluation information and no funding provisions for evaluation activities. (pp. 15–16)

The labour market development agreements with the provinces must improve substantially upon this arrangement to generate state-of-the-art summative evaluation research that is credible and worth sharing.

Third and finally, experiments can be reserved for strategic situations in which evaluations from nonexperimental data fail to reach any consensus or where a major program initiative is under serious consideration. The major benefit of an experiment is to provide clear and credible answers to important policy questions, such as the private and social rates of return to a specific training program or funding mechanism. In particular, experiments are designed to answer specific questions, and as such have clearly defined research objectives, something that is missing from most adult education and training program measures at present. The gap between the evaluation research produced by an experiment, such as the current Self-Sufficiency Project, and a nonexperimental program need not be as large as it is currently, but it can only be narrowed by careful nonexperimental evaluation research designs.

NOTES

1 Although the average duration of the period after training is not provided, the follow-up surveys vary from one to three years after the baseline (pretraining) survey, providing little opportunity to assess the actual persistence of training benefits.

2 This is the traditional “narrow” concept of social returns and is relatively straightforward to estimate.

3 One might be tempted to conclude that training actually hampered earnings prospects, which seems counterintuitive. It is more likely that those who were eligible for training under the Canada Jobs
Strategy and who accepted training would have received lower earnings than the nontrainee group even in the absence of training. Although the nontrainee group was chosen carefully to correspond with the trainee group according to demographic characteristics, there is no guarantee that the two groups are comparable in other, often unobservable, ways that affect subsequent earning performance (see our discussion earlier in this section). We suspect that the returns to training (had they been calculated) are understated in this study.

REFERENCES


Employment Program Analysis Branch, Strategic Policy and Planning, Employment and Immigration Canada, Ottawa.


