

John J. Heldrich Center for Workforce Development

research brief

Evaluating Workforce Programs: A Guide to What Policymakers Need to Know to Structure Effective, User-Friendly Evaluations

by Kathy Krepcio, William Mabe, and Charyl Staci Yarbrough

This brief discusses the value and purpose of program evaluations, highlights different evaluation tools and techniques, and illustrates how policymakers and program managers can structure and implement evaluations of workforce development programs.

Introduction

With nearly 14 million Americans unemployed and growing competition from low-cost, high-skill workforces abroad, improving education, training, and employment outcomes for job seekers should be a top priority for policymakers, funders, and training providers. In an era of declining budgets, lawmakers and funding organizations are, more than ever, looking to allocate funds to workforce programs and practices that can provide evidence of effectiveness at a reasonable cost. Program evaluation is the means for assessing program effectiveness and it can benefit policymakers and funders in a number of important ways:

- Learning what works and what does not work for diverse groups of people;
- Understanding what the program has accomplished, why, and at what cost;

- Documenting effective practices for replication internally and elsewhere;
- Identifying barriers to success and program weaknesses;
- Getting evidence needed to take early corrective action; and
- Making multidimensional evidence-based information just as available as anecdotes and stories.

In an effort to foster the more widespread use of program evaluations, the John J. Heldrich Center for Workforce Development at Rutgers University has prepared this guide for a wide variety of public and nonprofit organizations that implement publicly and privately funded workforce development programs serving a diverse array of job seekers. The guide seeks to explain evaluation tools and techniques in a straightforward manner, debunk some evaluation myths, and demonstrate how policymakers and program managers can structure effective, user-friendly evaluations of workforce development programs that best fit their unique program needs.

What is Evaluation?

Evaluation is the systematic and objective process by which a researcher assesses the quality, effectiveness, or value of an “evaluand,” defined as the subject of evaluation, such as an organization, program, policy, or activity (rather than a person). It is important to consider each of the pieces of this definition:

- The process is systematic because it follows established rules of scientific inquiry.
- The process is objective in the sense that any neutral observer would arrive at the same conclusions about the program if she used the same methods as the evaluator. Moreover, because the evaluator, unlike program implementers, lacks any stake in the program, he or she can be thought of as the voice of program participants, attempting to provide a neutral assessment of how well the program meets the needs of those affected by it.
- A researcher may assess the quality of an evaluand by studying **how** the organization, program, policy, or activity operates.
- Effectiveness relates to whether the evaluand achieves the goal it seeks to achieve.
- A value assessment places the effectiveness of the evaluand in the context of its costs and refers to the extent to which the evaluand is cost effective.

What Evaluation is *Not*

Although the following activities are important and may be necessary for an evaluand to be successful, they are analytically distinct from evaluation.

- **Evaluation is not auditing.** Audits are related to evaluation because they are both implemented to support program implementation and organizational process. However, the core purpose and strategies for each are quite different. Audits are primarily intended to verify the accuracy and truthfulness of information. Evaluations provide insight on best practices for utilizing organizational capacity and determine if an evaluand’s efforts are yielding the intended results.
- **Evaluation is not a needs assessment.** A needs assessment is often necessary for any program or policy to be effective because it enables the designers to identify the goals that the policy or program should target. Needs assessment is, therefore, prior to evaluation.
- **Evaluation is not customer satisfaction.** Participants in programs or individuals affected by a policy may have opinions of the quality or effectiveness of the program or policy, and the evaluator should take this information into account when conducting an evaluation. However, a thorough program evaluation must consider far more than just the degree to which participants are satisfied with a program or policy, including whether the program or policy is effective at achieving its goals, the extent to which it is well run, etc.
- **Evaluation is not technical assistance.** The individuals who are implementing a program must have the knowledge and competence to implement the program successfully. The role of the evaluator is to assess how well these individuals implement the program. Although the recommendation sections of evaluation reports will necessarily convey information for technical assistance, the evaluator’s principal role is to convey this information, not to guide the implementation of the recommendations.

Evaluation Myths 101

There are many misconceptions about evaluation and these misconceptions often deter people who run programs or implement policies that might benefit from evaluation from engaging in it. There are four principal evaluation myths.

Myth #1. Evaluation is a Gotcha. Evaluation is not a search for what is wrong with a policy or program. Unfortunately, this myth has arisen because some evaluators have adopted a “gotcha” approach. Properly conducted, an evaluation is a partnership between program and evaluator in which the evaluator engages in a process of inquiry that helps the program identify what about it works well and what about it needs to be improved. Although the evaluator must necessarily look for aspects of the program that detract from its value, the spirit of this search is not rooted in an intent to “show up” the program, but rather to identify how it can best be implemented to the benefit of the program’s constituency.

Myth #2. Evaluation cannot establish the effectiveness of all programs. Some program implementers contend that their programs are highly contextual and for this reason their effects cannot be measured. Even if all of the disparate effects of an initiative cannot be measured, the core intended effects of any initiative should be clear and measurable. Vague or ill-conceived program goals will always be difficult to measure and evaluate. Poorly defined goals reflect on poor program development rather than on the utility of evaluation as a tool for assessing program effectiveness. In fact, one of the benefits of conducting an evaluation is that it can help a program clarify the outcomes it aims to achieve. Moreover, an evaluation that assesses goals and incorporates context can be put in place for any program that has a goal or set of goals that are clear and measurable.

Myth #3. Evaluation is just about numbers. While many evaluations do collect numerical data (for example, number of people placed in employment), they are not always just about the numbers. For example, a process evaluation can help uncover the context in which the program is operating and can help to shed light on how the program is working. At the other extreme, however, some evaluators believe that they do not need to know any details about the program, and that all they need to evaluate a program is the numeric program data. This approach to evaluation is misguided. An evaluation can only be effective and of benefit to program managers if the evaluator understands the context of the program, the population that the program is serving, and the goals that it aims to achieve. The evaluator cannot evaluate the quality of program implementation and overall program effectiveness without this qualitative data.

Myth #4. Evaluations must be complex to be successful. Although some evaluation methodologies are highly complex, involving experimental or quasi-experimental designs and cutting-edge statistical analysis techniques, many others are more straightforward and employ simpler techniques. The sophistication of the methods used depends on the nature of the program being evaluated and the goal of the evaluation. Depending on the program being evaluated, qualitative techniques may, for example, generate more useful information than the most sophisticated statistical models.

Why Should I Evaluate?

There are many reasons to conduct program evaluations of workforce programs. Although at the most basic level, a program may engage in evaluation because the funder — whether a foundation or government agency — requires it, the true value of evaluation lies in allowing the workforce program to establish how effective it has been at serving job seekers. By applying a systematic analytical process, evaluations can generate credible evidence of the effectiveness of a workforce program. By establishing a program's net impact and overall effectiveness, evaluation can be vital to building a program's sustainability.

As important, evaluation can play a critical role in helping to make workforce programs as effective as they can be. Evaluation can benefit workforce programs by:

- Learning what works and what does not work for diverse groups of program participants;
- Understanding what the program has accomplished, why, and at what cost;
- Documenting effective practices for replication internally and elsewhere;
- Identifying barriers to success and program weaknesses;
- Getting evidence needed to take early corrective action; and
- Making multidimensional evidence-based information just as available as anecdotes and stories.

In other words, systematic evaluations can help program staff to identify the components of their programs that are effective and working well, point out aspects of programs that detract from program success, and help to pinpoint barriers to program success. As

important, an evaluation can uncover the barriers — programmatic, environmental, or participant-related — that may limit a program's effectiveness. By generating all of this knowledge, evaluations give rise to recommendations for how policymakers and managers can improve their programs.

Types of Evaluations

Systematically conducted evaluations can generally be divided into two different categories: **process evaluations** designed to improve the implementation of programs and identify the factors that are contributing to a program's success or failure, and **outcome evaluations** that seek to measure how effective the program is at achieving its objectives. These evaluations can make use of quantitative or qualitative data and are usually conducted on a one-time basis (for a designated period of time). Other types of evaluations that will not be discussed in this brief include performance monitoring (such as completing quarterly progress reports, reviewing program metrics such as number of participants served compared to targeted numerical goals, etc.) and cost-benefit studies (such as analyzing program financial data to address how much the program and/or its components cost, especially in relation to the benefits being produced by the program).

Process Evaluations

Process evaluations are generally established to understand what is happening in a program and how it is producing the results or outcomes it has been set up to achieve (for example, employment at wages above the minimum wage). As demonstrated in Table 1, a process evaluation can involve the collection of data through various modes, including focus groups with stakeholders or program participants, surveys of participants, interviews with key stakeholders and project staff, as well as participant observation.

Table 1. Overview of the Components of Process Evaluation

Other Names	<ul style="list-style-type: none"> ▪ Process Evaluation, Implementation Evaluation, Formative Evaluation
Answers these Questions	<ul style="list-style-type: none"> ▪ What is occurring in the program? ▪ How is the program being implemented and operated? ▪ How is the program producing the results that it does?
Methods Used to Collect Data or Information	<ul style="list-style-type: none"> ▪ Conducting site visits ▪ Collecting and analyzing program administrative data ▪ Conducting focus groups with program participants ▪ Interviewing key informants such as program participants, local delivery staff, program managers, key program partners ▪ Conducting surveys (e.g., web, telephone, in-person) to gather information from program participants
Can Illuminate	<ul style="list-style-type: none"> ▪ How the program is operating ▪ Why program performance goals are/are not being met ▪ What is required to make the program or practice work successfully ▪ What is required of program managers and/or staff to successfully deliver services that results in realizing intended program outcomes ▪ The cultural context in which the program works or does not work ▪ The strengths of the original program model and/or program operations ▪ The weaknesses of the original program model and/or program operations ▪ The need for more, less, or the collection of different program data
Can Provide	<ul style="list-style-type: none"> ▪ Meaningful and practical recommendations to improve and/or change the program model and/or operations and information to help correct program model shortfalls, especially if an intent is to replicate the model

Process Evaluation of New Jersey's Literacy Labs

In 2001, the New Jersey legislature allocated approximately \$6 to \$8 million a year from the state's Unemployment Insurance trust fund to support the New Jersey Workplace Literacy Program, which established "literacy labs" at One-Stop Career Centers and affiliates throughout the state. The labs allow participants to access computers and a variety of multimedia technology tools designed to improve reading, math, communication, computer, and general workplace readiness skills. The labs' ultimate goal is to assist participants to obtain, maintain, or advance within a job. However, shorter-term objectives include helping participants to increase their basic academic and workplace-related skills, achieve a recognized educational credential such as a GED, and/or enter an approved occupational training program following completion of literacy lab services.

In 2003, the State Employment and Training Commission contracted with the Heldrich Center to perform a process evaluation of the state's literacy labs. The purpose of the evaluation was to identify both effective program practices as well as practices that needed to be improved to better serve literacy lab clients.

The Heldrich Center conducted site visits to literacy labs, administered structured interviews with program staff, and held focus groups with program participants. In addition, the Heldrich Center worked with nationally recognized experts in adult literacy and education to review the software and video tools used in the labs for self-directed study.

The evaluation found that there were a number of ways that the implementation of the literacy labs could be improved. A key issue was the difficulty of serving individuals with vastly different literacy needs. The New Jersey Labor Department's policy was to allow everyone, at any literacy level, to receive instruction through a literacy lab, but the labs did not have sufficient staffing to support the diverse instructional needs of low-level English learners, on the one hand, and more advanced students, on the other. Adult literacy professionals who served on the expert panel convened for the study, and the scholarly literature in the field of adult literacy, agreed that individuals with different literacy needs require different types of services. The labs were under-staffed and so lab staff faced a trade-off between providing the intensive one-on-one instruction that lower-level English learners require and being available to support the self-directed study of high-level learners and GED students.

The Heldrich Center also learned that although the labs used assessment results to select software programs and other technology tools for participants to use, the labs did not use detailed assessment results to create varied and highly customized learning activities that are closely connected to participants' job-specific goals. Best practice calls for a transparent and purposeful approach to developing curricula for individuals, which means that the rationale behind all learning activities should be clear to the learner and relate directly to the learner's needs and goals. Through interviews, the Heldrich Center learned that one impediment that made it difficult for literacy lab staff to create highly customized plans of study for literacy lab participants was that many of the individual employment plans that had been developed by the clients' One-Stop counselors were unclear, because they included neither specific, obtainable job goals nor action steps for achieving job goals.

Based on these and other findings, the Heldrich Center offered a series of recommendations to increase staffing at the literacy labs, improve the assessment process, and enhance coordination between One-Stop and literacy lab staff.

Outcomes-based Evaluations

Outcomes-based evaluations look at the impact and/or changes to program participants that the program's services or other interventions are designed to effect. Outcomes-based evaluations critically depend on the program staff to do an effective job of collecting participant data, both on participant attributes, such as demographic characteristics, and dosage, including the quantity and intensity of the services that participants received. For an outcomes-based evaluation to be effective, evaluators must have access to administrative data and/or the diligent collection of follow-up data on participants after they complete the program.

There are four primary types of outcome evaluations. First are outcome evaluations that look at performance outcomes without a control or comparison group. So an evaluator, for example, might estimate the employment rate or average earnings, but would not compare these outcomes to those of similar individuals. Such outcome information can provide valuable information on program performance, but without something to compare the outcomes to, it is difficult to say whether the program being evaluated is effective.

To inject something comparative into the analysis, an evaluator might make comparisons over time. Did the average individual's situation improve over some time period from before the program to after it? Over time comparisons provide incredibly valuable information. At the same time, however, a comparison over time alone is not sufficient to establish that a program has been effective. The reason is because many things besides the program are always going on in the world and some of these things affect the same outcomes that any training program seeks to influence. The economy, as it has done over the past three years, may have declined. As a result, the employment

rate for program participants may be lower after program completion than it was before the program started, and this reduced employment rate may have nothing at all to do with the quality of the program but instead with the larger economic forces that are beyond the program's control. It is easily possible that the employment rate declines would have been greater had the person not participated in training. If a researcher were to restrict over time comparisons to only those individuals who participated in the program, she may wrongly conclude that the program is ineffective when it may have actually been effective.

In order to avoid drawing incorrect conclusions about program effectiveness, it is therefore necessary to also make a second comparison, across individuals. Did the program participants fare better over time than similar people who did not participate in the program? There are two ways to establish this sort of comparison. The most well-known such strategy is an experimental design, whereby applicants to a program are randomly assigned either to participate in the program or to receive a separate set of services. Randomization in assignment is assumed to create two groups that are all but identical except for the fact that one group participates in the program, while the other does not. Experimental designs have been thought to yield the most accurate estimates of program effects for the participants studied. There may, however, be circumstances — as when it may be difficult to ensure that control group members do not receive treatment services, the program is a universal service program in which participants cannot be turned away, or when the program does not receive enough applicants to populate both a treatment and a comparison group — that may limit the ability to implement an experimental design. Experimental designs are also the most expensive evaluation designs to implement.

Similar to an experimental design, a quasi-experimental design compares program participants to a comparison

group of individuals who are similar to the participants. These designs use administrative data sources on individuals who have received services that are similar to the ones being offered by a particular workforce program as a comparison pool. Researchers then engage in probabilistic matching in order to select a subset of individuals from the comparison pool who are as similar as possible to program participants based on key characteristics, such as age, sex, race, income, work history, etc. Quasi-experimental designs can yield reliable estimates of program effects at a fraction

of the cost of experimental designs. The potential weakness of a quasi-experimental design is that there may be unobservable attributes that make the members of the treatment group systematically different from members of the comparison group that cannot be controlled for using statistical methods.

As illustrated in Table 2, outcomes-based evaluation is fundamentally used to tell an organization whether its programs are having a positive effect on the people they are serving.

Outcome Evaluation of an Occupational Training Program

A nonprofit organization enlisted the Heldrich Center to conduct a process and outcome evaluation of a program that it manages to prepare low-income residents of an urban area for careers in the construction industry. The program aims to achieve this goal by preparing its graduates for apprenticeships with a construction and building trades union. During the 10-week program, students receive intensive and highly targeted academic preparation in math, reading, and critical thinking; are introduced to the different building trades through hands-on work and site visits; and receive instruction in life skills. After completing the program, graduates apply for apprenticeships with the building trades. The program, which trains about 100 individuals a year, has developed strong ties to the building trades unions to ensure that its candidates are considered for employment.

To conduct the outcome evaluation, Heldrich Center researchers used a combination of program data on the services that participants received and the demographics of participants, state Unemployment Insurance wage record data, and Employment Services data. Heldrich Center researchers used the Employment Services data on individuals who completed different types of training programs in the same geographic area as the program to create a comparison group of individuals who were as similar as possible to program participants on key variables, including age, sex, race, and prior employment history. The researchers used probabilistic matching software to select the most similar individuals from the Employment Services data for the comparison group. By selecting only individuals who completed training programs and by matching on prior employment history, the evaluators were able to control for the motivation of participants. (Individuals who complete training are typically more highly motivated and thus more likely to do well in the labor market than individuals who do not complete training. By including only training completers in the comparison group, the researchers limited the confounding effects of differences in motivation.)

After the comparison group was created, the researchers compared the earnings growth from before training to after exit from training for the program participants and the comparison group members. The results showed that program participants witnessed significantly higher earnings growth than similar individuals who completed other types of training programs. The researchers attributed these results to the strong relationships that the program had built with the labor unions, which provide access to relatively high-wage jobs, and partly to the skill improvements that participants realized over the course of the program.

Table 2. Overview of the Components of Outcomes-based Evaluation

Other Names	<ul style="list-style-type: none"> ▪ Impact Evaluation, Summative Evaluation
Answers this Question	<ul style="list-style-type: none"> ▪ Did the program have its intended effect(s) on the program participants?
Methods Used to Collect Data or Information	<ul style="list-style-type: none"> ▪ Collecting and analyzing individual and summary administrative records ▪ Administering and analyzing time-interval participant surveys (in-person, telephone, web)
Can Illuminate	<ul style="list-style-type: none"> ▪ The extent to which a particular service or strategy is reaching its objective (that is, changing a condition, changing a behavior) ▪ Whether there are changes in outcomes among program participants
Can Provide	<ul style="list-style-type: none"> ▪ Information about whether the program's efforts have improved people's general condition (for example, employment situation)

Key Outcomes for Workforce Program Evaluations

The ultimate goal of any workforce program is to help an individual find, keep, or get promoted in a job. In order to position the person to achieve this goal, there are a number of intermediate outcomes — sometimes referred to as outputs — that a program might hope to achieve, such as improving the person's skill level in an occupation or improving the individual's English language skills.

It is critical at the outset of any project to clearly identify both the ultimate and the intermediate outcomes that the program hopes to achieve. By engaging in this process, program staff can take their goals and translate them into measurable outcomes that can be tracked in order to monitor how well the program is achieving its objectives.

For job seekers, key employment outcomes can include whether the individual found employment, how long she retained the job, the amount that she earned on the job,

whether the occupation in which she was hired was related to her training, and/or whether the job offers benefits. Key outcomes for incumbent workers include earnings, retention in the job, and whether the individual receives a promotion.

Intermediate outcomes can include skill gains, which can be different for various types of programs. Whereas an occupational training program might seek to measure gains in job-specific skills, a basic literacy program would look to measure whether the participants gained one or more literacy levels. As intermediate outcomes, programs can also measure whether participants earned a degree or an industry-recognized credential.

The exact outcomes that a program chooses to measure should reflect the goals that the program is trying to achieve but be balanced with the ease or difficulty of measuring that specific outcome.

Data Collection for Evaluation

In order for a workforce program to be evaluated, it is necessary for the program to systematically collect and electronically store data on each participant. For every program participant, program staff need to collect data on participant characteristics, the services that each participant receives, and all outcomes that the participant realized. For example:

- Participant characteristics include basic demographic information (age, race, sex, veteran status, disability status); education, literacy, and/or English as a Second Language level; employment status, earnings, and occupation at the time of enrollment in the program; and any attributes of program participants that might bear on how well they perform in the program, how they should be served, and how they are likely to fare after program completion.
- Services received include detailed information on the type of service (e.g., case management, occupational training, job search assistance, etc.) that an individual received; the frequency of the service received (number of hours per day, number of days per week, number of weeks); attendance; and other factors that might affect the client's ability to achieve the outcomes the program hopes he will achieve.
- Outcomes for workforce programs can roughly be divided into three categories: skills gains; the attainment of a degree, industry-recognized credential, or other certificate; and employment outcomes, including employment, earnings, and retention in employment. Often, the evaluator will play a key role in collecting outcomes information.

The evaluator may identify additional intermediate outcomes that the program should measure to track its progress.

It is also important for any workforce program to coordinate with its funder at the start of the program to identify the specific data elements that it must collect to meet the funder's reporting requirements. Unless a workforce program engages in systematic data collection and uses a reliable mechanism for tracking and storing individual-level data on participant characteristics, the services each participant received, and the outcomes of each participant, the program may not be able to be evaluated.

When Should I Evaluate?

Every workforce program should be established with an eye on evaluation from the outset. Programs should be set up so that it will be possible for an evaluator to assess the extent to which the programs are achieving their goals, the effectiveness of various components, and the aspects of the programs that may be impeding success without imposing on routine processes and overburdening personnel. Including evaluation as a component of program implementation from inception can help keep program managers focused on how to make their programs as effective as possible as well as substantially help programs achieve their goals.

Although the importance of evaluation throughout the life of a program is constant, the focus and best methods of evaluating each program change over time. Like products, programs too have life cycles beginning with conceptualization, then piloting, then widespread implementation, maturity, and possibly phase out or replication elsewhere. One of the most significant facts about program evaluation is that the evaluation requirements of programs change over their life cycles. The program life cycle is a critically important concept for anyone who wants to conduct a program

evaluation because where a program is in its life significantly affects both the goals that the program manager can hope to accomplish through an evaluation and the type of evaluation that the program manager may want to conduct.

In the early stages of a program's implementation, the needs are different than when the program has been in operation for a while. For example, at inception, a program cannot be evaluated but needs to be set up

in such a way that the information that will be needed to perform an evaluation can be collected. Therefore, programs require different types of evaluations when they are at different stages in their life cycles. As explained earlier, evaluations can take many forms — from process evaluations designed to improve the implementation of programs and identify the factors that are contributing to their success or failure to outcomes-based evaluations that seek to establish how well the program is working.

Program Evaluation Versus Performance Measurement

Many public, as well as foundation-funded, workforce development programs are required to demonstrate the effectiveness of their programs in terms of certain performance measures. A performance measure is a numeric summary or description of how a program has functioned. The hope is that requiring programs to meet performance measures will make them more effective at delivering services. Performance measures for workforce programs can include output measures, such as the number of individuals served, or outcome measures, such as the percentage of program exiters who find employment.

Although performance measures can provide valuable information on how a program is functioning, they are different from and less informative than program evaluation. A workforce program can be said to be effective if it makes the people it serves better off than they would have been had they not participated. Program evaluation can answer this question by carefully comparing the outcomes of program participants with similar groups of non-participants. Performance measures typically capture short-run quantities, typically up to a year after program exit. The implicit theory behind performance measures is that they are an accurate proxy for how well program participants will do over the long run.

Unfortunately, research shows that short-run performance measures do not accurately predict how successful a program participant will be in the labor market over the long run. Although the performance measures in the Job Training Partnership Act (JTPA) — the forerunner to the Workforce Investment Act (WIA) — incorporated a dose of comparison by granting states the flexibility to adjust performance standards depending on the population served and economic conditions, performance measures are a blunt instrument when it comes to comparison. (WIA does not include these adjustments in its performance measures.) What performance measures cannot answer is, for example, whether an X percent employment rate is good for the types of people the program served and in the economy in which they had to find work. Whereas a carefully designed evaluation can answer this question, performance measures really cannot. Heckman, Heinrich, and Smith¹ assessed the extent to which a variety of performance measures used under JTPA were related to participants' long-term labor market success, as measured in experimental studies. They found that there was no relationship between how program participants fared on the short-term JTPA performance measures and how they were doing in the labor market 18 and 30 months after starting to receive JTPA services.

The bottom line is that to learn whether a program is effective, performance measures alone are inadequate and rigorous program evaluation is needed.

Lessons Learned from Past Evaluations of Workforce Programs

From evaluating dozens of workforce programs, the Heldrich Center has gleaned a number of lessons that program managers may find useful in implementing their programs. Although there are many different specific lessons that could benefit a variety of different programs depending on their focus, the Heldrich Center has identified the following four lessons as broadly applicable to a wide range of workforce programs.

Lesson #1. Participant Recruitment. Getting enough people to participate in the program often presents a significant challenge to new workforce programs. In an economy with a high rate of unemployment, one might assume that job seekers would flock through the doors of any program aiming to help them find a job. New programs, however, often face the challenge of getting the word out about the program and also in persuading potential participants that the program has a sound approach for helping them obtain employment. Developing partnerships and strong relationships with public workforce and nonprofit organizations that serve populations similar to the one that a workforce program serves **before** the program begins can help the program meet its recruitment targets.

Lesson #2. Business Engagement. Employers only hire the workers who have the skills they need. Therefore, it is vital that before a program begins accepting clients that it reach out to businesses and identify their specific skill needs. Programs that provide occupational skills training can seek feedback on their curriculum from employers to ensure that the skills they will be teaching participants are the skills that businesses say they need, and, as importantly, that the curriculum is thorough enough to enable students to learn these skills during the course of the program.

Lesson #3. Address Participants' Multiple Barriers to Employment. Many participants face multiple barriers to becoming employed and/or advancing in their jobs. These barriers have been well documented in the workforce development literature and include a lack of basic math and English skills, limited proficiency in English, lack of transportation, mental health issues, a physical or cognitive disability, lack of child care, and a criminal record or outstanding warrants, among others. The workforce programs that are the most successful are the ones that seek to address the multiple barriers that can prevent the participants from getting or keeping a job. Programs that take a case management approach can tailor their services to the specific barriers that each of their clients faces.

Lesson #4. Relationships with Outside Agencies. As noted above, it is important to build relationships with outside agencies before a program begins. This is especially true if the program anticipates receiving funds of some sort, such as Individual Training Account funds from One-Stop Career Centers, from an outside organization. It is important for the program to learn the agency's requirements for reporting on the progress of clients referred by that agency as well as the particulars of how that agency prefers to handle billing.

Purchasing Evaluation Services

Many program managers find locating and purchasing evaluation services to be difficult and challenging. An important, but often overlooked, factor is that an evaluator should be identified and begin working with program managers while they are designing the program and before they begin implementing it. In looking for an evaluator, program staff should identify evaluators who not only have established track records in evaluation but also have experience in evaluating related programs.

Finding the Ideal Evaluator

The ideal evaluator is one who has:

- **Education** in research and/or evaluation methods, through either formal training in evaluation or through graduate studies in the social sciences;
- **Experience** in conducting program evaluations as well as in-depth substantive knowledge related to the program, especially in the specific area of program focus; as a result of this experience, the ideal evaluator should have a good reputation for meeting expectations for quality, timeliness, and rigor;
- Extensive experience in using a **wide variety of evaluation methodologies** and does not try to “sell” the manager on one specific methodology; and
- **Commitment to building and maintaining a collaborative relationship** with the program manager.

As noted in Table 3, there are a number of sources for evaluation services, ranging from knowledgeable individual evaluators to small, medium, and large public and private firms and organizations. For programs that are brand new, the program manager may want to enlist the services of a local evaluator who will be able to meet face-to-face with program staff and engage in participant observation in order to get a better sense of how the program operates. Larger programs or more rigorous outcomes-based evaluations may require the services of larger firms or universities that possess the staffing, analytic, and data collection resources necessary to conduct the evaluation.

Budgeting for Evaluation

Some program managers planning for evaluation find it difficult to budget or cost out evaluation services. While the cost of an evaluation is highly dependent on the type of evaluation design (for example, process versus outcome, experimental versus quasi-experimental outcome evaluation designs), a good place to start is to assume a baseline of 10% for process and outcome evaluations. If the evaluator will conduct more specialized experimental and/or quasi-experimental evaluations, programs should budget between 10% and 20% of their overall program budget for evaluation services.

Working with an Evaluator

Workforce program personnel that are well prepared and ready for the evaluation can significantly ease the process of evaluation. The evaluator will develop an evaluation that lays out the purpose of the evaluation, the research questions, and what data will be collected and how it will be collected. As a first step in preparing to work with an evaluator, program staff should clearly specify program goals and strategies for accomplishing the goals. To inform the evaluation plan, program staff should work with evaluators to develop logic models² that lay out their programs’ resources, inputs, outputs, and short- and long-term outcomes. By engaging in the process of developing a logic model, program staff can build a better blueprint for how the program will achieve its goals and the evaluator will develop a better understanding of how the program will operate.

Table 3. Finding a Qualified Evaluator that is Right for You³

Evaluator Sources	Description	What to Look For	General Cost
Independent researchers	Independent college or university-based faculty and/or graduate student or private independent consultants with expertise in program evaluation and required quantitative and qualitative evaluation methodologies.	Individual(s) with the appropriate credentials, experience, and a reputable track record of conducting evaluation projects.	\$ to \$\$
Educational institutions (college or university)	Postsecondary educational institutions, predominantly found in academic institutes and centers that possess expertise in program evaluation and required quantitative and qualitative evaluation methodologies. Larger universities may possess broad capacities in such areas as survey design and data collection, administrative data and secondary data analysis, and promising practices research.	University or college-based institutes or centers with proven experience in evaluation (including desired required evaluation methodologies) of employment and training, workforce development, and similar social service programs.	\$ to \$\$\$
Private for-profit or not-for-profit firms with expertise in evaluation	Private for-profit or not-for-profit firms that specialize in program evaluation, or have a unit that possesses such expertise. Larger firms may offer broad capacities in such areas as survey design and data collection, administrative data analysis, secondary data analysis, and promising practice research.	Reputable organizations with proven experience in evaluation (including desired required evaluation methodologies) of employment and training, workforce development, and similar social service programs.	\$\$ to \$\$\$

Legend: \$ = low, \$\$ = moderate, \$\$\$ = high

Perhaps the most important single step that program staff can take to facilitate evaluation is to engage in systemic, ongoing collection of all the data that the evaluators need. A good evaluator should either provide technical assistance to program staff to facilitate data collection or partner with the program to gather needed data. Finally, the program can facilitate program evaluation by informing stakeholders about the evaluation and emphasize its importance to making the program as effective as possible.

All in all, it is reasonable for programs to have the following expectations about evaluators and the evaluations they conduct:

- A willingness to learn about the program to be evaluated;
- An interest in developing a collaborative partnership to learn about and evaluate the program;
- Regularly scheduled reports summarizing the findings of the evaluation and recommendations for future action;
- A final report that presents the final findings of the evaluation; and
- Regular communication with program staff about the progress of the evaluation and what has been learned, as well as any information that the program might incorporate to improve its operations.

Conclusion

While recognizing that resources for evaluations are (and will continue to be) limited given current federal and state fiscal constraints, there remains strong pressure from the general public and oversight bodies such as Congress and state legislatures, to demonstrate program effectiveness and impacts. This makes evaluation an important, if not a necessary, part of any program

activity. In the workforce development field, engaging in more systematic evaluation of both public and privately funded employment and training programs provides a chance for program planners, policymakers, and managers to engage in high-level accountability activities that can show evidence as to the effectiveness of services for unemployed and dislocated workers as well as those seeking educational and career opportunities and advancement.

As noted earlier, while there have been difficulties and barriers to undertaking formal evaluation of workforce programs, there are ways to engage in evaluation activities that are relatively low cost, simple to implement, and that yield important, practical information that is relevant to improving program operations. While there are strong arguments for instituting rigorous experimental and quasi-experimental evaluation methods that go further than documenting and describing outcomes, these evaluations require adequate budgets and seasoned evaluators that may sometimes be beyond the fiscal and managerial capacities of many nonprofit agencies or small workforce programs. An appropriate and effective strategy may be to proceed in increments — starting out with process evaluations and more effective performance monitoring using these as platforms for more substantial evaluation efforts at a later date.

The call for more systematic analysis and stronger accountability in workforce program operations and outcomes is most likely going to continue to grow. As noted in this brief, evaluators can design and implement evaluations in a variety of ways and deploy a range of tools to support evaluation activities. A critical first step for program managers is to choose the methods and tools that work best for them, and that offer the greatest utility and usability to program operations.

Resources

Basic Resources About Program Evaluation

American Evaluation Association

<http://www.eval.org/>

The American Evaluation Association (AEA) is an international professional association of evaluators, focusing on the evaluation of programs, policies, products, and organizations. This Web site provides access to a public e-library of evaluators' work, information on AEA conferences and presentations, an online career center providing job and résumé postings, and a search function to find evaluators by expertise or location as well as information on scholarly journals published by AEA.

"Approaching an Evaluation: Ten Issues to Consider," Brad Rose Consulting

http://www.bradroseconsulting.com/Approaching_an_Evaluation.html

This webpage offers 10 key issues to consider when planning an evaluation, offering brief explanations for why certain criteria matter as well as questions that span the scope of an evaluation process from planning to use of findings.

"Basic Guide to Program Evaluation (Including Outcomes Evaluation)," Free Management Library

http://www.managementhelp.org/evaluatn/fnl_eval.htm

This document offers an overview of the main elements in planning and executing an evaluation process in either for-profit or nonprofit organizations, including an overview of basic elements, selecting methods, disseminating results, and pitfalls to avoid.

"Chapter 4: How Do You Hire and Manage an Outside Evaluator?" U.S. Department of Health and Human Services

http://www.acf.hhs.gov/programs/opre/other_resrch/pm_guide_eval/reports/pmguide/chapter_4_pmguide.html

This webpage, an excerpt from a more comprehensive guide to evaluation for program managers, discusses key issues to consider in finding and working with an outside evaluator for social service programs, including basic steps for finding the right evaluator and potential evaluation responsibilities for both the independent evaluator and program staff.

"The Critical Need for Program Accountability and Evaluation," Facilitation and Process

<http://facilitationprocess.com/the-critical-need-for-program-accountability-evaulation>

This blog post discusses the importance of incorporating accountability and evaluation activities into nonprofit programming. The author discusses three key barriers that prevent the implementation and practical use of evaluation activities as well as highlights the importance of evaluation to nonprofit growth and management.

"Evaluation Primer: An Overview of Education Evaluation," U.S. Department of Education

<http://www2.ed.gov/offices/OUS/PES/primer1.html>

This primer gives an in-depth overview of evaluation processes used by the federal government in evaluating educational programs, including descriptions of evaluation designs, recommended steps in planning an evaluation, and interpreting evaluation results.

“Evaluation Research,” Social Research Methods

<http://www.socialresearchmethods.net/kb/evaluation.php>

This resource offers an online textbook discussing the broad range of topics involved in social research, discussing evaluation in detail. Topics of discussion include an overview of evaluation terminology and research designs, the interaction of project planning and project evaluation processes, and the impact approach and culture make in conducting an evaluation.

“Guidelines for Selection of Evaluators,” UNESCO

<http://unesdoc.unesco.org/images/0015/001583/158394E.pdf>

This document offers technical and qualification criteria recommended by the United National Educational, Scientific, and Cultural Organization (UNESCO) to select an evaluator, such as appropriate expertise and diversity of evaluation teams.

“Outcome Indicators Project,” the Urban Institute and the Center for What Works

<http://www.urban.org/center/cnp/projects/outcomeindicators.cfm>

The Outcome Indicators Project, a collaboration between the Urban Institute and the Center for What Works, is intended to provide a framework to monitor and improve the performance of nonprofit organizations and initiatives. The Web site offers resources specific to building a common outcome framework, outcome and performance indicators specific to 14 program areas, and generic outcomes that can be used across nonprofit programs.

“Selecting the Right Independent Grant Evaluator”

http://www.findgrantevaluators.com/step_6.php

This resource provides guidance for professionals searching for an independent evaluator that is appropriately suited to conduct a project’s evaluation. It offers a general approach to locating an evaluator that meets the needs of the program, and the requirements of the grant funder, including general questions to frame the search, a list of potential interview questions, and indicators of competency to look for in potential candidates.

Resources About Logic Models

Logic Model Development Guide. W.K. Kellogg Foundation, October 2000. To order a print copy of the guide, call 1-800-819-9997 and request item #1209.

Measuring Program Outcomes: A Practical Approach. United Way of America, 1996.
<http://web.pdx.edu/~stipakb/download/PA555/OutcomeMeasurementAtUnitedWay.pdf>

Everything You Wanted to Know about Logic Models but Were Afraid to Ask. Connie C. Schmitz, Professional Evaluation Services, Minneapolis, Minnesota and Beverly A. Parsons, In Sites, Boulder, Colorado. <http://www.insites.org/documents/logmod.htm>

Resources About Evaluation Professionals

The American Evaluation Association has a useful reference, “Find an Evaluator”
http://www.eval.org/find_an_evaluator/evaluator_search.asp

U.S. Department of Health and Human Services, Administration for Children and Families, Chapter 4: How Do You Hire and Manage an Outside Evaluator?

http://www.acf.hhs.gov/programs/opre/other_research/pm_guide_eval/reports/pmguide/chapter_4_pmguide.html

Juvenile Justice Evaluation Center, “Hiring and Working with an Evaluator”

<http://www.jrsa.org/pubs/juv-justice/evaluator.pdf>

Endnotes

1. Heckman, J., Heinrich, C., & Smith, J. (2002). *The Performance of Performance Standards*. Retrieved from http://www-personal.umich.edu/~econjeff/Papers/pstand_final.pdf.

2. A logic model is a planning tool that clarifies and graphically displays what the program intends to do and what it intends to accomplish. According to the W.K. Kellogg Foundation’s *Logic Model Development Guide*, the components of a logic model vary, but most often they articulate resources, activities, outputs, outcomes, and goals.

3. OVC’s Technical Assistance Guide Series. *Guide to Hiring a Local Evaluator*. Retrieved March 17, 2010 from <https://ovcttac.gov/taResources/OVCTAGuides/HiringLocalEvaluator/wherelook.html>.

About the Authors

Kathy Krepcio (krepcio@rutgers.edu) is the Executive Director of the John J. Heldrich Center for Workforce Development.

William Mabe, Ph.D. (billmabe@ejb.rutgers.edu) is the Director for Research and Evaluation at the Heldrich Center.

Charyl Yarbrough, Ph.D., (cyarbrou@ejb.rutgers.edu) is a Research Project Manager and oversees the workforce and education evaluation portfolio for the Heldrich Center.

About the Heldrich Center

The John J. Heldrich Center for Workforce Development, based at the Edward J. Bloustein School of Planning and Public Policy at Rutgers University, is a dynamic research and policy center devoted to strengthening the nation’s workforce. It is one of the nation’s leading university-based centers dedicated to helping America’s workers and employers respond to a rapidly changing 21st Century economy. The Center’s motto — “Solutions at Work” — reflects its commitment to offering practical solutions, based on independent research and evaluation, that benefit employers, workers, job seekers, and the nation’s network of workforce development professionals. Evaluations conducted by Center researchers make use of a systematic process to estimate a workforce program’s value and/or identify which of its components contribute to and detract from the program’s value. All evaluations are designed and implemented to gain an understanding of the program to be evaluated and to producing information that can inform program and policy decisions. Heldrich Center evaluation work has included a range of quantitative/outcome and qualitative/process evaluations such as an evaluation of postsecondary training providers and pre-apprenticeship programs, as well as various workforce program efforts to connect youth, dislocated workers, minority males, individuals with disabilities, and incumbent workers to employment opportunities. Learn more at <http://www.heldrich.rutgers.edu>

Did You Know?

You can use your smart phone to take a photograph of the barcode on the right and immediately visit the Heldrich Center Web site? All you need is a QR (or Quick Response) Reader, a smart phone, and an Internet connection. Learn more at: <http://www.mobile-barcodes.com/qr-code-software/>



RUTGERS

Edward J. Bloustein School
of Planning and Public Policy