

Evaluation Toolkit for Prospective Workforce Innovation Fund Grantees

May 2014



Developed by:
Abt Associates Inc.
National Evaluation Coordinator for the Workforce
Innovation Fund
Employment and Training Administration
Department of Labor

WORKFORCE *Innovation* FUND



Contents

1. Introduction to Evaluation	2
1.1 Evaluation and the Workforce Innovation Fund (WIF).....	3
1.2 Establishing the Evidence Base	4
1.3 Evaluation Designs and Expectations for Evaluation Strength.....	5
1.4 Selecting a Project Type.....	7
2. Planning for Your Evaluation Phase 1	9
2.1 Creating a Preliminary Evaluation Plan for an Outcomes or RCT Study.....	9
2.2 Creating a Preliminary Evaluation Plan for an Implementation Study.....	29
2.3 Creating a Preliminary Evaluation Plan for a Cost Study	30
3. Planning for Your Evaluation Phase 2	32
3.1 Creating an Evaluation Design Report	32
3.2 Linking the Evaluation to the Intervention	34
3.3 Reporting on the Evaluation	37
3.4 Protecting Participants Rights.....	39
4. Evaluation Timeline and Budget.....	48
4.1 Evaluation Timeline.....	48
4.2 Evaluation Budget.....	52
5. Selecting Your Evaluator	55
5.1 What Do You Want in an Evaluation Team?.....	55
6. Working with the WIF National Evaluation Coordinator (NEC).....	60
6.1 Providing Technical Assistance	60
6.2 Review of Evaluation Processes and Deliverables.....	62
6.3 Collecting and Synthesizing Final Results	63
Appendix.....	64
Glossary of Terms	64
Evaluation References	68
Blank Template for Development of a Logic Model	73

DISCLAIMER: This toolkit is designed to assist prospective applicants. If there is any inconsistency between items in this toolkit and the Workforce Innovation Fund Solicitation for Grant Applications (SGA/DFA PY-13-06), and/or any Department of Labor (DOL) regulation or guidance, the SGA or DOL regulation or guidance will prevail.

1. Introduction to Evaluation

This toolkit was created for practitioners, service providers, and others who want to learn more about program evaluation. It was designed to provide a basic understanding of evaluation concepts and principles. While created specifically for prospective Workforce Innovation Fund (WIF) grantees and evaluators funded by the U.S. Department of Labor (DOL) and containing some content specific to WIF evaluations, it presents broad evaluation concepts and lessons that can be useful for evaluating other similar programs.

To get the most from this toolkit, it should be read in its entirety, as earlier chapters provide information that later chapters build upon. Each chapter provides critical information about either evaluation fundamentals or expectations specific to WIF evaluations.

Chapter 1 presents an overview of evaluation and the WIF program, and includes a description of how evaluation fits into the WIF program. In addition, this chapter describes the most common types of evaluation, including those with which prospective WIF applicants should become familiar.

Chapter 2 reviews the steps for creating a preliminary evaluation plan, an essential part of developing an application for a WIF grant. The chapter covers concepts such as creating a “logic model,” developing research questions to guide the evaluation, and reviewing the “evidence base” for an intervention. This chapter also provides guidance on “outcomes”: how to determine the most appropriate outcomes for a program and how to measure them effectively with the evaluation. Finally, it describes the type of evaluation appropriate for WIF, including the appropriate research methodology to use as well as data sources, and data collection methods.

Chapter 3 covers a number of evaluation-related issues and activities that a potential grantee hiring an evaluator should understand. These include a review of what is to be included in an Evaluation Design Report, how the evaluation can and should link to the program, what kind of reporting to consider from an evaluator, and protecting participants’ rights.

Chapter 4 reviews key considerations for developing a timeline and budget for an evaluation. Chapter 5 covers factors to take into account when selecting an evaluator.

Chapter 6 describes the role of the National Evaluation Coordinator. Finally, the appendix to the toolkit provides a glossary of the most common evaluation terms, and lists a number of resources on various evaluation topics. If at any point in reading the toolkit you encounter a term that you are unfamiliar with, check to see if it is in the glossary.

1.1 Evaluation and the Workforce Innovation Fund (WIF)

Program evaluation is more important today than ever before. Federal, state, and local policy makers increasingly look for evaluation results when making decisions about program investments. In this era of limited public resources, information on program effectiveness can be critical for garnering or maintaining support for an initiative. With this in mind, evaluation was designed into the WIF program, requiring promising workforce development programs to document what works. Specifically, WIF requires that all grantees contract with a third-party evaluator to conduct an evaluation of the funded initiative that will build on and expand the relevant base of evaluation literature. The importance of evidence-based practices is reflected in WIF's grant structure, where the amount of funding for WIF grantees is linked to the availability and quality of existing evaluation evidence supporting the proposed program.

Program evaluation is not only important to policy makers at the federal and state level, but also to direct service providers. High-quality program evaluations can provide multiple direct benefits to the program and participants served by the program. Not only can an evaluation tell you whether your program produces positive outcomes (e.g., were program participants able to find a job, increase their earnings), but it also can tell you—and the broader workforce community—how you were able to achieve results (e.g., what activities or actions produced the results).

Below are several ways that evaluations can positively affect your program and ultimately the individuals you serve:

- Program evaluations can help you **improve your program**. Evaluation can tell you whether your program produces positive outcomes or if it is not working as well as you had hoped. Learning that your program does not produce the results you envisioned may be just as valuable as learning that your program had positive results. You can then make changes that may result in improved outcomes for the individuals you serve.

- Program evaluations can result in **better outcomes for program participants**. Program improvements made based on evaluation findings result in a program that is better able to serve your participants and produce desired outcomes.
- In addition to benefiting your program directly, your evaluation findings can also provide **benefits to the larger workforce development community**. Program administrators and policy makers in the workforce system can benefit by learning about your program's outcomes. If there are positive outcomes, it could provide support for similar programs in other communities. Conversely, positive outcomes may prompt other localities operating different programs to adapt a program (or program elements) similar to yours. The implementation study component of the evaluation can help others replicate your program in their setting.
- Evaluations can help you **secure funding needed to sustain and scale up your program**. As noted, policy makers and other funders generally are seeking to invest in strategies that have proven effective. Should your evaluation produce positive or promising findings, it may result in increased support, as there would be an interest in continuing and even scaling up your program given the track record of effectiveness.



1.2 Establishing the Evidence Base

One of the goals of WIF is to promote effective and innovative strategies through evaluation and build a set of proven strategies that have positive participant and systems outcomes for the workforce community. WIF achieves this goal, in part, by requiring grantees to evaluate their workforce service provision approaches, also known as interventions, with a higher standard of rigor than previous evaluations of the approach you are implementing, known as the “evidence base.” This ensures greater certainty about whether a program truly produces positive outcomes for those it serves. You can think of the evidence base as the existing literature that documents the implementation of an intervention, how that intervention has been systematically evaluated, and the findings related to how the intervention influenced participant or systems-wide outcomes. The evidence base of existing studies should serve as the “point of departure” for your intervention, helping you define how and at what level you should implement the intervention and select the appropriate WIF project type. The evidence base also helps to determine the appropriate evaluation type for your intervention.

As a part of the WIF grant application, you will be asked to document the evidence base for your proposed intervention. In addition to being an application requirement, documenting the evidence base can help ensure that:

- You can incorporate into your planned activities the lessons learned concerning implementation best practices.
- Your proposed evaluation does not just replicate previous evaluations, but instead builds on what has already been learned.
- You are familiar with the range of issues experienced by previous program operators and evaluators of the proposed intervention (e.g., recruitment challenges, opportunities for partnership).

To document your proposed intervention’s evidence base, you will need to research published evaluation results that examine similar interventions in other contexts. Published evaluations can be found through a number of sources, including the websites of research firms and universities. There are also a number of government websites that have valuable evaluation resources. Once those resources are gathered, a next step is to create a system for archiving lessons learned from each evaluation. This archiving system, also known as a literature review, should document the findings from each study as well as its overall design, including the types of data collected, the methods for collecting and analyzing these data, and the overall level of rigor of the evaluation design and implementation. Additional discussion of how to identify and evaluate the evidence base for your intervention can be found in chapter 2 of the toolkit.

1.3 Evaluation Designs and Expectations for Evaluation Strength

This section of the chapter provides an overview of the two primary evaluation approaches supported by WIF—the pre-post outcomes study and the randomized controlled trial (RCT) study—to help familiarize you with the concepts. This section also describes what to expect in terms of the strength or “rigor” of these types of evaluations for understanding the effect or impact of the program on participants. In addition, this section describes implementation studies as well as cost studies, both required in WIF evaluations.

Pre-Post Outcomes Study

In these studies, data are collected at two points in time: (1) at “**baseline**,” before the intervention is implemented or before a participant begins the program, and (2) at “**follow-up**,” after the intervention is implemented or after program

*A **pre-post study** describes changes that occur during and after the intervention, but does not allow you to attribute any changes to the program itself.*

participation has ended. Essentially, the evaluator is taking a snapshot of selected aspects of individuals' or the systems' well-being (also known as “**outcomes**”) before the intervention and comparing it to the same aspects of the same individuals' or systems' well-being after the intervention. Common outcomes measured in workforce evaluations include (but are not limited to) employment rates, wages, training completion rates, coordination of activities (for evaluations of system-based interventions), employer engagement rates, number of participants engaging in a new online platform. Comparing the baseline and follow-up outcomes data in this way allows the evaluator to measure changes over time and describe participants before and after the program (for service-based interventions) or to describe the system before and after the intervention (for system-based interventions).

Outcomes studies are a good first step in developing an evidence base for a new intervention. They can help set the stage for later, more rigorous evaluations by generating early evidence of effectiveness. Importantly, compared to randomized controlled trials (discussed below), outcomes studies are less costly and time-consuming to conduct, which is appropriate for newly developed or untested interventions.

The main limitation of outcomes studies is that they do not allow you to link the intervention to any changes in outcomes. This is because outside factors other than the intervention may have contributed to the changes. For example, an improving economy, not just program services, may be at least partially responsible for increases in employment among program participants. As such, an outcomes study provides descriptive information on program effects but is less rigorous than a randomized controlled study.

Randomized Controlled Trial (RCT) Study

A randomized controlled trial (RCT) research design measures the “impacts” of the intervention or program on individuals or systems. An impact is an estimate of the direction (positive or negative) and magnitude (by how much) of the change in outcomes that can be directly attributed to the intervention.

*With a **random assignment** design, you can learn about the difference the intervention made, and be able to say with confidence that the intervention caused the change(s).*

The key to this design is random assignment. Eligible applicants are randomly assigned, as if by lottery, to the treatment group that receives the services provided by the intervention or to a control group that does not. This approach assures that the two groups are identical in all respects except that one will participate in the intervention (program services) and the other will not. Therefore, any

differences in outcomes between these groups (e.g., different rates of employment) can be directly attributable to the intervention.

RCTs are considered the “gold standard” in evaluation because this method allows programs to claim with a certain degree of confidence that participants have improved their employment outcomes solely because of that program. Although studies using this design can require more effort to design and implement, if random assignment is conducted correctly, the results provide clear, rigorous evidence of program effectiveness and carry a lot of weight with policy makers. Additionally, the results from an RCT evaluation will provide important contributions to the evidence base for your intervention. Results from this evaluation approach are also valuable to workforce development stakeholders and scholars in determining whether the expected impacts were realized, and in developing approaches that build on this evidence to refine and expand programs.

Implementation Study

An implementation study illuminates and explains “what is happening and why” in the design, implementation, administration, operation, services, and outcomes of social programs. This type of study can provide context and information that makes evaluation results more useful for improving program implementation. In addition, findings from implementation research can be used to inform future program development or replication.

Cost Study

Cost studies can come in various forms (e.g., a cost allocation analysis, a cost-effectiveness analysis), but the element common to all is that they provide information on the costs of the program, and some go further to provide information on how effective the program was compared to its cost. Cost studies are useful tools in allowing future practitioners to determine if they have the capacity to implement the intervention, and if the cost of the intervention is “worth it.”

Chapter 2 provides a more detailed discussion of implementation studies and cost studies.

1.4 Selecting a Project Type

For all WIF-funded evaluations, the evaluation design type must be specified in the grant application and is based on the type of project proposed, as defined by WIF. For each proposed intervention, the project types (A, B, and C) are defined by the level of innovation involved, the existing evidence base, the quality of the intervention’s logic model, and for Type C projects, the intent for use of the funding. The table below presents the WIF project types in summary format, highlighting requirements for each type. The remaining sections

of this toolkit will provide information to assist you in working through the elements within the table to determine which project type fits your intervention.

Table 1-1: Characteristics of WIF Project Types

Project Type A	Project Type B	Project Type C
<ul style="list-style-type: none"> ✓ The program offers a new and more effective strategy for addressing widely shared challenges. ✓ The program has not been systematically studied before, but the intervention is supported by a strong logic model. ✓ The proposed intervention is a departure from existing workforce strategies (i.e., it is innovative). 	<ul style="list-style-type: none"> ✓ The program has been implemented and evaluated before, and the evaluation results indicate some potential for positive impacts on participant or system-wide outcomes. <li style="text-align: center;"><i>OR</i> ✓ The program has been implemented but not by your organization, and the program is supported by strong evidence of positive change. The program has been evaluated before using an outcome, quasi-experimental, or random assignment design, and evaluation findings indicate some statistically significant positive change. ✓ Funding for this project will expand knowledge about the project's efficacy and provide more information about the feasibility of implementing proven projects in different contexts. 	<ul style="list-style-type: none"> ✓ Your organization has implemented the program before. ✓ The program has been evaluated before using a random assignment design, and evaluation findings indicate some potential for positive impacts on participant or system wide outcomes. ✓ Funding for this project would support a significant expansion of structural and/or service delivery reform ideas.

Additional Resources

Please note that useful references and resources on the topics covered in this toolkit can be found in the references table in the appendix.

2. Planning for Your Evaluation Phase 1

This chapter will lead you through the steps of creating a preliminary evaluation plan as part of your WIF application. The steps include understanding the evidence base particular to your intervention, developing a logic model, selecting the most appropriate evaluation design and study methodology, and identifying additional evaluation components that should be considered as part of the study. Note that a final evaluation plan will be developed after (and if) your project is funded in consultation with a professional, independent evaluator; the final evaluation plan will refine and expand upon your preliminary evaluation plan.

2.1 Creating a Preliminary Evaluation Plan for an Outcomes or RCT Study

The preliminary evaluation plan helps conceptualize evaluation activities and is useful for developing a clear understanding of how the evaluation and its related activities link with programmatic activities. The preliminary evaluation plan is required as part of the WIF grant application and can be used as a roadmap for implementing the evaluation and preparing a Request for Proposals (RFP) to procure an independent, third-party evaluator.

This section of the toolkit reviews a number of essential elements for the plan. The plan should lay out a path for both developing and implementing a rigorous evaluation, and should include information on the following topics, each of which will be discussed in detail below:

- Graphical and narrative description of the program's logic model
- Purpose of the evaluation
- Aspects of the program to be evaluated (some evaluations test particular services or program activities rather than the entire program)
- Key guiding research questions
- Existing literature review on the same or similar subjects
- Expected program outcomes that the evaluation will measure
- Evaluation type (e.g., pre-post outcomes study, randomized controlled trial)
- Research methodology planned, including discussion of potential data sources and data collection methods

To begin the process of creating your preliminary evaluation plan, you should start with a logic model.

Developing a logic model

Early in the evaluation (and program) planning process, you should create a logic model that describes the entire proposed intervention. The W.K. Kellogg Foundation describes a logic model as clarifying *what* you are doing in the intervention or program—the expected results, intermediate and long-term outcomes—and identifies how the project’s activities will contribute to achieving those outcomes.¹

A logic model communicates how the program operates and what it “looks like” when implemented.

A logic model is grounded in a specific “theory of change.” Examining the theory of change behind the program intervention helps to clarify the context in which the logic model exists. A first step in developing a theory of change, after identifying the social or economic issue to be addressed, is to identify the theoretical solution(s) for the issue based on available data. A next step is to describe the desired outcomes and impacts in addressing the issue. Finally, you develop a plan for moving from the identified issue to the desired outcome. In this final step, the resources, activities, and other outputs leading to the desired outcome are clarified.

A logic model communicates how the program operates and what it is expected to “look like” when implemented. It should be sufficiently detailed to allow for replication, even in another context with a different group of participants. Logic models should provide a detailed account of the intervention’s content and organization, its duration, the amount of staff training required to implement the intervention, and the services provided or system change activities undertaken. It should also contain a clear depiction of the relationships between intervention elements and the intermediate- and long-term outcomes those elements are expected to affect.

Why Is a Logic Model Important?

A well-designed logic model serves as a blueprint for the implementation and evaluation of the program it represents. Whenever possible, program developers, evaluators, and practitioners should invest significant time at the program planning stage thinking through the following:

- Elements of the program or intervention and their specific characteristics

¹ See the organization’s website, www.wkkf.org, for more information and examples.

- Resources and activities needed to best support implementation of those program elements
- Changes that are expected (whether they be at the system- or participant-level)
- Effective methods to assess the outcomes of interest

A comprehensive logic model for the program should specify (both graphically and in a narrative format) *all* of the hypothesized pathways to improved outcomes that are being supported under the program. This could include, for example, (a) changes in the instructional/training environment or professional or job-related leadership activities; (b) changes in the training/support in terms of content, strategies, use of instructional time, materials, technology, formative assessment; (c) participant changes in attitudes, motivation, skills, knowledge; and potentially (d) changes in the relationship between participants, program staff, and possible employers.

Components of Logic Models

A variety of frameworks are used to describe the parts of a logic model. A complete logic model should include the key components of the intervention, including inputs and activities, outputs, intermediate outcomes, and long-term outcomes. Each of these is described in turn below. Note that if the intervention has more than one focus, then it may also be appropriate to develop two or more logic models (for example, if you plan to implement a systems change intervention along with a training program, those two elements would be best represented in two different logic models).

Key components of the intervention are the **activities** and **inputs** that are under the direct control of the individual or organization responsible for program implementation (e.g., program developer, grant recipient)² and that are essential in implementing the program. They may include financial resources, professional development for trainers, curricular materials, or technology products. The key components should be described in a way that helps others understand what the intervention entails and, at a summary level, what is required to implement the intervention. You should include all key components of the intervention in your logic model. It should be clear from your logic model how the key components are related to or expected to produce the outputs that ultimately lead to the intervention's intermediate and longer-term outcomes. The key components of the intervention can be

Key components of the intervention are those activities and inputs that are essential in implementing the program.

² Note that “the individual or organization responsible for implementing the intervention” does not include recipients of the intervention (e.g., trainees, program participants)

viewed as the ingredients or resources of the program that are intended to achieve the desired changes.

Outputs can be described as products of the key components. In other words, outputs are what occur when the inputs and activities come together and the intervention is implemented. Outputs can be the services provided to participants that are expected to lead to outcomes. Some examples include:

- Program services; training/professional development activities or other supports for trainers (e.g., group training, on-site coaching, distance training, learning/practice teams, curriculum materials);
- Instruction or skill development (e.g., technology, formative assessment(s), use of instructional time, participant groupings)
- Participant, employer, and community engagement activities
- Systems of coaching, advising, or referrals

Outputs are what occur when the inputs and activities come together and the intervention is implemented.

Intermediate outcomes are goals that the intervention is expected to help achieve that are steps leading to achievement of long-term outcomes. You should include in the logic model all of the intermediate outcomes through which you expect the intervention to affect participant long-term outcomes. Some examples include:

- Participants complete training program
- Participants obtain credentials
- Participants are placed in jobs
- Number of overlapping services are reduced
- Number of complete records entered into new MIS

Long-term outcomes are the changes in behavior, attitudes, aptitude/skill, and knowledge for trainers, participants, environments, or larger systems.

Long-term outcomes can be described as the expected changes in behavior, attitudes, aptitude/skill, knowledge, and so on for trainers, participants, environments, or larger systems. For workforce interventions, specific long-term outcomes might include changes in employment and earnings, and the receipt of credentials. All outcome domains that the program is expected to affect should be included in the logic model.

Logic Model Example

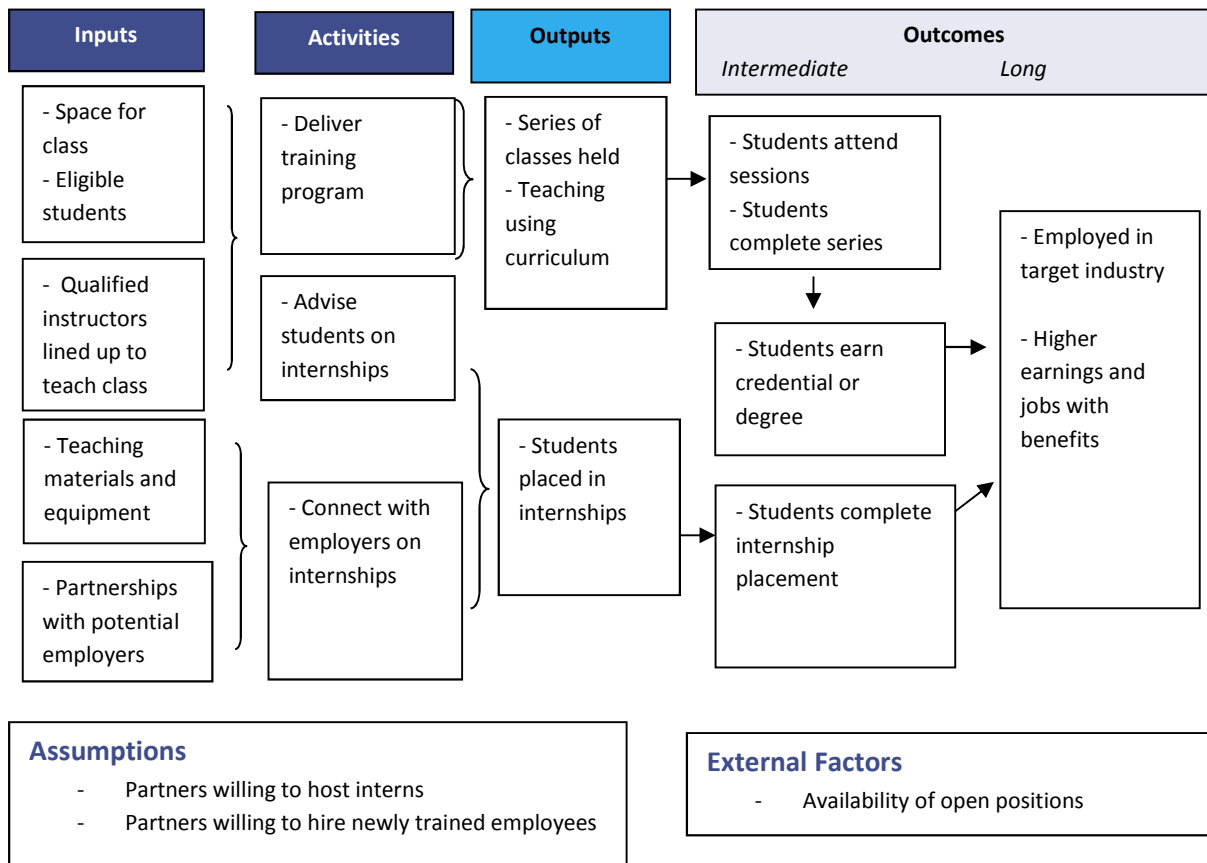
Logic models can be described in various formats: visual/graphical, tabular, and narrative. However, a graphical portrayal is most effective when combined with a detailed narrative description. Figure 2-1 below provides an example of a graphic representation of a logic

model. This example describes a simple intervention that offers a training program for a specific type of machine operator. The inputs are listed in the left-hand column and include space to hold the training sessions, eligible students, instructors to teach the sessions, materials, and partners who will provide slots for internships during the course of the training. The next column shows the activities that comprise the work being conducted in the intervention. These activities include conducting the training course, advising the students on internships, and connecting regularly with employers regarding the internships. The outputs are completing the delivery of the training course, including the use of the intended curriculum, and placement of students in internships simultaneous to the training. In this example, intermediate outcomes are students actually attending and completing the series of sessions, earning the related degree or credential, and successfully completing the internship placement. Long-term outcomes include employment in the target industry, and higher earnings and obtaining jobs with benefits.

This model also includes a space for consideration of assumptions made as well as any external factors that may bear on the intermediate- and long-term outcomes. These elements provide context for the intervention. Assumptions for this program include that the program will be able to secure partners who are willing and able to place students in internships and in entry-level positions. An external factor is the availability of the appropriate open positions for the students who complete the program. The program developers have researched the industry and know that current trends are positive for employment in this area, but turnover for positions may not necessarily continue at the historic pace.

Lastly, you will see in this example that specific inputs are linked to activities, activities are linked to specific outputs, and outputs are linked to specific outcomes. Rather than just a long list of each with unidentifiable links, there are several arrows showing which inputs and activities are expected to affect which outputs and outcomes. A clear understanding of these links will allow for better evaluation planning and, when needed, the ability to refine the evaluation and expected results based upon the reality of program operations.

Figure 2-1: Example of a Graphical Logic Model—A Training Program for Machine Operators



Assessment of Your Logic Model

Using the blank template found in the toolkit appendix, you can create a graphically displayed logic model. While you are required to provide both a graphical logic model and narrative description for the WIF grant application, beginning with the graphical template will enable you to identify the essential elements.

Once you develop an initial logic model, you can refine it by assessing its comprehensiveness, including its quality, breadth, and specificity. Using the questions below, you can further refine elements of your logic model, filling in where there are gaps in the responses to these questions.

Self-Assessing Logic Models: Guiding Questions

1. Does the model include critical inputs required for the implementation of the service activities? (e.g., recruiting and training trainers, developing a set of industry partnerships for internships)
2. Are there system-building activities that are part of the necessary foundation for the intervention? How are these indicated on the logic model?

3. Does the model include all of the key intervention activities to be provided to participants? Is there a planned or expected sequence to the participant activities that should be indicated?
4. Does the model include all key “first-level outputs” of the intervention? (For example, necessary but not sufficient conditions for achieving outcomes such as full participation, use of supportive services, and meetings with coach/advisor)
5. Does the model include all of the hypothesized immediate changes/outcomes expected for participants, across all relevant domains, and should these be linked to specific services that are assumed to lead to these changes? Does the theory underlying the design of the intervention specify specific participant outcomes for particular services?
6. Does the model suggest links between intermediate- and longer-term outcomes?
7. Are the longer-term participant outcomes likely to be measurable in the life of the evaluation?
8. As a complete visual or narrative text, does the logic model tell a clear and complete story about the unique intervention being tested and how it is hypothesized to lead to moderate and long-term outcomes for participants?
9. If using a visual representation, does supplementary narrative text provide a clear and complete story?
10. Are there assumptions about external conditions or other external factors that could affect the successful implementation of the intervention and if so, are these (or should these be) shown anywhere on the model?

As you think through the answers to each question, review your model and add detail or make adjustments. Be as specific and detailed as possible. Having a detailed, comprehensive logic model (or two if the intervention has more than one focus) is a critical foundation for guiding your evaluation design. Now that you have clearly specified what your intervention entails, you can move on to the next step of the preliminary evaluation plan: determining the evaluation purpose and scope.

Stating the evaluation purpose and scope

After creation of the program logic model(s), the next key step in developing a preliminary evaluation plan is to understand and build consensus around the **purpose of the evaluation**. Look back to your logic model and consider what you want to and can learn more about. Decide what you are evaluating, and why. Do you want to learn if your program saves money? Do you want to learn if your program helps participants find employment? Do you want to learn about completion rates and credentialing for program participants? Do you want to figure out if your new system or management structure is beneficial? Do you want to learn if particular services result in better outcomes than others (for example, do participants who receive an “extra” package of services have higher earnings or longer retention rates than those who receive the “standard” package of

services)? Clearly articulating exactly what it is you will learn from the evaluation will also help you to gain stakeholder support for the evaluation activities, as stakeholders will understand the purpose and potential benefits to your organization of the evaluation. During this process, you should continually refer back to your logic model to ensure that you consider all aspect of the program and what outcomes you expect to produce.



An integral part of deciding upon the purpose of the evaluation is determining **which program component(s) you want to evaluate and how you want to evaluate them**. In many cases, not all aspects of a program or grant-funded activity lend themselves to evaluation. It is also possible that certain components can be rigorously evaluated with data and quantitative analyses whereas other components more readily lend themselves to evaluation using qualitative information, such as interviews, document reviews, and focus groups (see section 2.2 for how to create a preliminary evaluation plan for an implementation study). One way to think about which components to evaluate and which methods and data sources to use (quantitative and/or qualitative) is to ask: *Can the (expected) results be quantified numerically?* For example, the effects of a training program on its participants can be quantified; it is quite feasible to collect numerical data on the amount or type of services being provided and on the characteristics or behaviors of participants receiving the services. Some system reforms can also be quantitatively measured. For example, when evaluating the implementation of a new management information system (MIS), it would be feasible to capture information on data accessibility and reliability, as well as on how the data are used in program operations. Other aspects of program operations, such understanding how different funding sources are brought together or how new partnerships are created cannot be enumerated but generate qualitative information. And, in many cases, program components can be evaluated using both qualitative and quantitative methods. For example, you may want to measure the results of a training program on participant earnings quantitatively but also learn about the participants' experiences and opinions qualitatively. Refer to section 2.2 for an in-depth discussion of the implementation study plan once you have decided which program elements you want to evaluate qualitatively.



Crafting your research questions

Once you have identified the goals of your evaluation and which program component(s) to evaluate, you should craft the key **research questions that will guide** your evaluation. Strong research questions identify distinct areas of program performance that your team would like to assess in a systemic and credible way. Strong research questions share the following characteristics:

- **They are specific and measurable.** While your team may have identified general goals and program components that you wish to evaluate, your research questions should identify the specific program components or outcomes you wish to test. For example, you may be interested in learning about the intervention’s impact on employment. A specific and measurable research question might be: “Are participants who complete the program in its entirety more likely to be placed in full-time unsubsidized jobs within three months of program completion than those who do not?” Research questions that incorporate a granular level of detail ensure that grantees, evaluators, and other stakeholders are clear about the outcome of interest and how it is measured. Including this level of detail also makes it easier to measure outcomes and ensures that results on employment mean roughly the same thing across all participants.
- **They are reasonable and appropriate.** Research questions that are reasonable examine outcomes that can realistically be achieved given the time available and expected level of effort of an intervention. For example, if your intervention is providing pamphlets detailing job search services to unemployment insurance claimants, it is probably reasonable to ask: “Did unemployment insurance claimants who received pamphlets detailing the job search services offered at American Job Centers (AJC) report a greater knowledge of AJC job search services than when they entered?” It is probably not as reasonable to ask: “Did the citywide community report a greater knowledge of AJC job search services after pamphlets were given to unemployment insurance claimants?” because the entire community did not receive the intervention. Research questions that are appropriate are built using past experience and findings reported in previous research. For example, in a job search assistance intervention, it would be appropriate to ask about intervention impacts on job placement, job retention, and earnings.
- **They are answerable.** Research questions must be answerable. There are many reasons why a research question may not be answerable; data may not exist to answer the question or that the outcome of interest may not be sufficiently defined.

For example, many workforce development programs are interested in their program's impacts on participant self-sufficiency. However, self-sufficiency does not have a standard unit of measurement and may mean many different things to many different people. To answer this question, you would need to define the term self-sufficiency and come up with a way to measure it (which would likely be done in consultation with your third-party evaluator).

Your research questions should be as specific as possible, focused on generating evidence of effectiveness of the WIF intervention, and focused on the outcomes and effects of interest to DOL and the WIF program. Strong research questions are rooted in a firm understanding of your program intervention's activities, likely level of effort, timeline, and likely target group as well as an understanding of past interventions' demonstrated impacts. Having realistic and specific expectations will help you design research questions that are specific, measurable, reasonable, appropriate, and answerable. Once you have drafted your research questions, you should review the existing scholarly evidence, and use the research questions you find in that review to refine the research questions that you will ask during your own evaluation.

Review of scholarly evidence, building on the evidence base

Once you have created your program logic model(s), decided upon the purpose and scope of the evaluation, and drafted your research questions, the next critical step in creating your preliminary evaluation plan is to review the existing research-based evidence related to your intervention. Reviewing the evidence will enable you to plan for and design an evaluation that provides useful, timely results and builds on the current knowledge base. The existing evidence will help you do the following:

- Determine what program design type may be appropriate (and which WIF grant type to apply for) by building off of and improving upon the existing evaluation work that has been done.
- Determine what evaluation methods to employ by looking at what methods researchers previously used (e.g., along with the outcomes or RCT study, will you also conduct an implementation study? A cost study?).
- Determine what aspects of the program to evaluate using which evaluation design, data sources, and methodology by looking at how components of other program were evaluated.
- Identify appropriate program outcomes and how best to measure them.
- Refine your research questions by looking to past research questions and answers, and expanding upon what has already been learned.

- Ensure that your evaluation builds upon the existing evidence, and contributes additional information to the current base of evidence (that is, your evaluation goes beyond what has already been done and sheds new light on the program/issues/question; also known as a departure from existing evidence).
- Consider how to best integrate evaluation-related activities into program operations.
- Look ahead to how you may want to disseminate and inform others of your eventual evaluation results.

The evidence review should reference scholarly studies of interventions similar to the one you are planning (for example, evaluations of other job training programs) and provide a summary of the findings of each study and how it relates to your program and evaluation plans. You should discuss the studies' methods: the overall design, the types of data, the methods for collecting and analyzing them, and the overall level of rigor of the design, implementation, and findings. Sometimes, this type of information is not available in a study's public report. In this case, as much information as can be gleaned should be described.

The evidence base for your intended intervention should not be limited to exact replicas of your intervention. You can and should use research on previous interventions that are *related* to or have implications for your intervention. Research on interventions focused on different populations, with some variation in design or services can be considered part of your evidence base. For example, if you wish to implement a job training program for ex-offenders, it would be appropriate to cite in your evidence review studies that have examined the efficacy of both job training programs and programs designed for ex-offenders. Both job training and ex-offender studies will provide valuable information on how your program should be adapted to serve this population and will offer important insights into why your proposed intervention may improve participant outcomes. Any scholarly literature you used to help conceptualize your intervention should be considered to be part of your evidence base.

The evidence review will not only guide you in your choice of intervention and evaluation design type, research questions, and methodology, but it will also guide you in your choice of which WIF grant type (A, B, or C) to apply for. Table 2-1 details the key elements of each reviewed study that should be examined so that you can choose the appropriate WIF grant type (note that this table does not cover all of the elements you should review in creating your preliminary evaluation plan; instead, it covers the elements that apply when choosing your WIF grant type).

Table 2-1: Guidance for Developing the WIF Evidence Base

Element to Review	Description
Implementation History of the Intervention	<p>The first step in choosing the appropriate WIF grant type is to search websites, journals, and workforce-based forums for the implementation history of your intervention. That is, in the past have other workforce agencies implemented the intervention or a closely-related version of it? <i>(If there is <u>no</u> implementation history of the intervention or a closely-related intervention, then the remaining elements below will not apply and you will instead turn to your proposed logic model to determine which grant type to apply for.)</i></p>
Evaluation History of the Intervention	<p>The next question to answer is whether the intervention in some form (an exact replica or similar version) has been previously evaluated. You should consider independent evaluations conducted by a third-party evaluator that collected and analyzed data, as well as evaluations conducted by state agencies. Anecdotes about how the implementer <i>perceived</i> the effectiveness of the program do not count as evaluations and should not be considered in this context. You should document published evaluation results that:</p> <ol style="list-style-type: none"> (1) Refer to at least two data collection time-points (e.g., “baseline and follow-up,” “pre-post”); (2) Discuss statistical testing methods that were used (e.g., chi-square tests, t-tests, ANOVA tests, ANCOVA, MANOVA tests, linear regression models, multiple regression models, hierarchical linear modeling (HLM), logistic regression models, cost-benefit, cost-effectiveness); (3) Refer to the “significance level” of the analysis results (e.g., p-value, alpha level, “statistically significant,” “p less than 0.05”). <p>Articles, posts, or reports that use these methodologies are considered to provide “scholarly” evidence and can be used as “evaluations” for your review purposes.</p>
Evaluation Design	<p>The type(s) of past evaluations conducted for your proposed or similar intervention is crucial to determining your WIF grant type. You will need to determine if the design was one of the following (listed in descending order of “rigor”):</p>

Element to Review	Description
	<p>(1) Randomized Controlled Trial—individuals are randomly assigned to a treatment group that receives the intervention under study and a control group that does not.</p> <p>(2) Quasi-Experimental Design (QED)—this design type can come in many shapes and forms. The key difference between QED studies and RCT studies is that QED studies use a <i>comparison</i> group that is <u>not randomly formed</u>. Also, while the comparison group should be referred to as such, you may sometimes see it incorrectly referred to as a “control” group, so you cannot determine design type based simply on word choice. Instead, look at how the groups were formed. Was the comparison group formed from existing administrative data? Was it formed purposely (that is, people were purposely selected to be in that group so that they could “match” the treatment group), or was it “randomly formed” (referring to the random assignment). Note that QED evaluations may use the following terms: Difference in differences; nonequivalent control groups design; repeated treatment designs; regression discontinuity designs; case-control designs; time-series designs; interrupted time series designs; propensity score matching; instrumental variables; panel analysis.</p> <p>(3) Pre-Post Outcomes Design—these evaluations have <u>no</u> control or comparison group. Only the intervention participants in the program under study are included in the evaluation, and outcomes will be reported for this group only. The evaluation should include data from at least two time points, typically referred to as “baseline” and “follow-up,” or even simply just “pre” and “post.” Studies or evaluations that measure outcomes/data at only one time point (such as after program participation) do <u>not</u> count as a pre-post design and cannot be considered for a WIF evidence base.</p>
Evaluation Results	<p>Lastly, you should examine the evaluation results. Were they positive or negative? In what areas did the program produce effects? In what areas did it not? Does the report/journal article/website mention that the results were “positive,” “showed gains,” or some other term referring to a desired</p>

Element to Review	Description
	<p>outcome? Did the report discuss that these positive results were “significant” or “statistically significant,” or for random assignment studies, that the “treatment group” had better/larger/bigger results than the “control group?” If so, then you can consider the evaluation results to indicate some potential for positive impacts on participants or the system.</p> <p>Note that many evaluations have positive results in some areas/outcomes but not in all. You will likely find a mix of positive and negative results when reviewing the evidence base. When creating your WIF application, you will want to look for some potential for positive impacts on participants or the system, and explain your reasoning behind viewing that potential as positive.</p>

Sometimes it can be difficult to differentiate between anecdotal evidence and scholarly evidence. Your evidence review should focus exclusively on scholarly evidence—that is, evidence from studies conducted by independent evaluators that use rigorous methods and publish their findings in a report or peer-reviewed journal article. Anecdotes from program staff about how their program worked, or how they think it helped participants, does not qualify as scholarly evidence. Scholarly research evidence will discuss statistical significance and provide numerical estimates of the size of the changes related to the program (its “effects” or “impacts”). The following two examples illuminate the differences between scholarly evidence and anecdotal evidence.

Anecdotal evidence quote:

“Our program was very successful. We served 630 participants. The program really helped people find jobs and improve their resumes. 328 participants had jobs at the end of the program.”

Scholarly evidence quote:

“The program produced positive, statistically significant effects on participant employment. Participant employment increased by ten percentage points over the course of the intervention.”

Table 2-2 provides is a list of common terms used in scholarly evaluation research literature. Keep an eye out for these to help you find scholarly sources of evidence regarding your intervention or a similar intervention.

Table 2-2: Terms to Look for When Identifying and Reviewing Scholarly Evidence

Key Terms		
Evaluation Types: Cost-Benefit Implementation Study Interrupted Time Series Pre-Post Propensity Score Matching Quasi-Experimental Random Assignment Randomized Controlled Trial Regression Discontinuity	Details about Evaluation Participants & Methods: Comparison group Contamination Control group Cross-over Population Sample Treatment group	Data Collection and Measurement: Baseline Follow-Up Outcome Measure Scale Variable
Analysis Methods: ANOVA Chi-square test Cronbach’s Alpha Regression Regression model Significance testing T-test	Evaluation Results: Alpha Bias Confound Counterfactual Generalizability	
		P-value Reliability Significance testing Statistically significant Validity

Sources of scholarly evidence

Scholarly evidence sources can be found in peer-reviewed journals, on federal agency research webpages, and on the websites of professional evaluation firms and universities. For WIF evaluations, DOL ETA’s Research Page (http://wdr.doleta.gov/research/eta_default.cfm) is a good resource, as is the Government Accountability Office’s website (www.gao.gov). There are numerous other research firms and universities that have publicly available reports, and you are encouraged to search those as well.³ Lastly, the *Journal of Labor Economics*, *Social Science Review*, and the *Journal of Public Policy and*

³ These cannot be recommended directly here due to concerns about endorsements.

Analysis may be resources as well. It may be helpful to identify other reviews of scholarly research in the area of your intervention, and these can be used to identify key studies.

Selecting and justifying the evaluation design type

Once you have decided on and described the evaluation's purpose and scope, and you have reviewed and shown an understanding of the evidence base, your next task in writing the preliminary evaluation plan is to describe the type of evaluation you will conduct (i.e., randomized controlled trial, pre-post outcomes study). While the WIF grant type for which you apply dictates the evaluation design that you will use, you will want to consider the following factors when choosing your WIF grant type (and thus the evaluation design type):

- **Prior evaluations conducted on the same or a similar intervention.** In order to build upon and expand the current evidence base, and contribute to your community of practice, your evaluation should use methods of higher rigor than prior evaluations. For example, if there are descriptive (pre-post outcomes) studies on an intervention that you are undertaking, you must conduct an RCT evaluation.
- **Methods to best and most accurately answer your research questions.** Some questions, such as questions about who is participating in your program and the characteristics of their participation, may be best answered with a pre-post descriptive study, whereas other questions about the effectiveness of your intervention are likely to be best answered with an experimental (randomized controlled trial) design. So, you



should also consider *what* you want to learn, and how sure you want to be about the eventual findings.

- **Organizational capacity to participate in the evaluation.** Consider how the evaluation activities will “blend” into the programmatic activities. The feasibility of carrying out the selected evaluation type should be discussed with program managers and stakeholders, and then should be discussed in the evaluation plan document. For example, for RCT evaluations of service delivery interventions, the plan should

describe how the random assignment process can be integrated into the participant intake and enrollment processes. Additional questions, such as how you will over-recruit for a control group, should also be addressed. For example, you will want to consider whether you expect your program to be oversubscribed (i.e., to have a waiting list). RCTs are easiest to implement when there is oversubscription.

This section of the preliminary evaluation plan should include a short discussion of the evaluation design type (RCT or pre-post outcomes) that demonstrates your understanding

of that kind of evaluation. You should also discuss the feasibility of carrying out that type of evaluation, and discuss why that evaluation design is appropriate for your proposed intervention.

This may be an appropriate point in your planning process to also think about an implementation study and a cost study. Consider how the RCT or pre-post outcomes study could be supplemented by an implementation and/or cost study. Then, craft additional preliminary evaluation plans for these supplemental studies. See sections 2.2 and 2.3 for further information.

Deciding upon and describing the research methodology

When planning for an evaluation, you must not only select an evaluation design, but also dig into the details of how the research will be carried out. These details are your **study methodology**. (Note some factors are only relevant to randomized controlled trials and are marked as such).

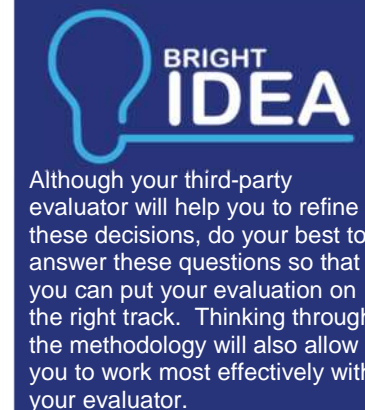
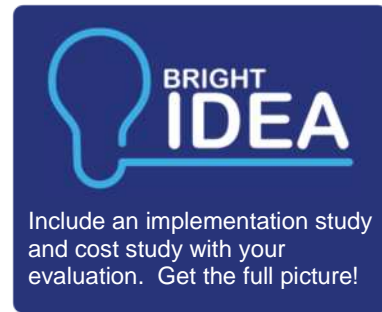
Although your third-party evaluator will help you to refine these decisions, do your best to answer these questions so that you can put your evaluation on the right track. Thinking through the methodology will also allow you to work most effectively with your evaluator.

Methodology Element 1: Population and Sample

First, you want to think about your target population for the intervention, and thus the study. Who is the overall population targeted for the intervention? What population will your study sample(s) represent?

Next, consider who your study sample will be; that is, who are the people who will actually participate in the evaluation. You should consider at this point if certain segments of the population may be excluded from the evaluation (e.g., veterans), and what the eligibility criteria for participation in the program and the study may be. For example, the study could include all WIA-registered individuals with limited-English ability that will enroll in a certain timeframe while the intervention continues to serve individuals who enroll after this period. Then, think about the sources for recruitment of the study sample.

Also, if you are conducting an RCT, consider how you will create a control group and still ensure sufficient program enrollment numbers. Programs often overestimate the



number of individuals who will be interested *and eligible* for the program. This can cause difficulties further down the road when half of the eligible applicants are assigned to the control group (control group members always have to be deemed eligible for the program before being randomly assigned). So, think about by how much the program may be oversubscribed, and discuss how you will recruit enough individuals into the evaluation. While you are conducting your evidence reviews, compare these estimates to the sample sizes of previous evaluations to get a sense of whether you can enroll enough people in the study for the results to be informative and meaningful.

Methodology Element 2: Random Assignment Protocols [For RCT evaluations only]

For randomized controlled trial evaluations, a critical element is the random assignment process. Here, you should provide an overview of how you envision carrying out random assignment. While you will work with your third-party evaluator on the details, you should be able to discuss in your preliminary evaluation plan the following questions.

- At what point in the program flow could participants be randomly assigned?
- How many different groups will you form using random assignment? Random assignment studies must have a control group that does not receive the intervention, but they can have one or potentially more than one treatment group. Will you assign participants to a treatment or control group, or perhaps to a treatment 1, treatment 2, or control group (this would apply when you want to test if different variations of a program work better or differently from one another)?
- What strategies do you plan to use for informing individuals of the study and obtaining their informed consent to participate?
- What strategies will you use for maintaining each individual's random assignment status?
- In what ratio will you assign participants to the treatment and control groups? Many RCTs employ a one-to-one ratio of treatment to control group assignment, but alternatives are possible. For example, you could design a study that randomizes participants to either a treatment group and a control group in a 2-to-1 ratio so that for every three participants two would be assigned to the treatment and one to the control group. This might be preferable in circumstances where the population being served is considered particularly vulnerable or in need of the intervention being examined or where you felt certain you could recruit a large sample to the study. The details of the random assignment ratio will be worked out with your evaluator, but at this point in your preliminary evaluation plan you should provide a rough estimate of your treatment-to-control group ratio.

Methodology Element 3: Determine Outcomes to be Measured

What outcomes do you plan to measure? The outcomes selected should directly link back to your research questions, study objectives, and logic model (see section 2.1). Identifying the outcomes you will measure will help inform your decisions on what data to collect and the methods for data collection. Note that outcome goals (e.g., the desired change in employment levels) do not need to be specified in an evaluation plan. The plan should focus on how the outcomes are defined rather than any specific target goals.

Note that you should specify the outcomes you plan to measure with each study component. That is, what will you measure with the impact or outcomes study? What will you measure with the implementation study? What will you measure with the cost study?

Methodology Element 4: Determine Data Sources for the Study

What data sources do you anticipate using? For some outcomes (like credentials, job placement, or job retention), the data may already be collected as a matter of business-as-usual. For others, data collection instruments like surveys need to be created and implemented. For RCTs, you will have to consider data sources for the control group who will be not going through the program.

Potential data sources include participant surveys, agency administrative records, state or federal administrative records, interviews, focus groups, and document reviews.

- **Administrative data** come from existing databases and can be especially useful for collecting employment and wage data on a large number of participants, as well as data on service receipt and completion and operational costs and resources. If you plan to use administrative data, you should describe the administrative data source(s), the samples, and larger population included in the data, and the measures that could come from these data. It is important to consider the costs that could be associated with acquiring data. For example, the National Directory of New Hires (a national database of employment and unemployment information sponsored by the Department of Health and Human Services' Office of Child Support Enforcement) charges a fee for every



Social Security Number submitted for a match. State unemployment insurance (UI) data may also be a resource, but will have cost and time impacts. Many free sources of information require that requestors put together an application for data access. So in addition to accounting for fees associated with data collection, you need to consider staff time for putting together data request applications when planning your budget. (See chapter 4 for additional discussion of evaluation budget considerations.)

- **Survey data** are useful for gaining information from the sample unit of analysis on a variety of indicators beyond those available from other administrative sources, including detailed characteristics of participants, program experiences, and employment and educational outcomes, among others. If you plan to use surveys, you will also want to think about timing and frequency of the surveys and envisioned survey mode (e.g., telephone, web, in person).

For participant-focused evaluations, time begins when a participant is randomly assigned (and not at program exit). This method of counting time allows for the timing of data collection to be equivalent across treatment and control groups (as control group members will not have a time of program exit).

For each data source, you should also consider what information you plan to capture with that source. For example, with surveys you may plan to capture participant outcomes. With interviews you may plan to capture information on program administration, and with focus groups you may plan to capture the participant or staff experience.

Methodology Element 5: Data Collection Strategy

At this stage, you want to think about at what time points you will collect data. It is important to align the timing of data collection to ensure that the information presented

for each variable is useful to the evaluation. For example, if participant earnings are a critical outcome variable, then wage measurement must be conducted at reasonable time points to note change. You should primarily base this decision on current evidence related to each outcome, but you can also factor in your own expectations for when you think program effects may emerge. When creating your evaluation plan, you may decide to define the data collection timing as specific months and years, or based upon the intervention month, or based upon general ideas such as “at program entry” or “12 months after program entry.”

Evaluation Timeline and Budgeting

Lastly, when devising your program evaluation plan, you should consider the overall timeline for the evaluation and how that will fit in with the project timeline. Budgeting is also an important factor and will play a role in determining who your evaluator will be, the

evaluation design type, the sample size, the data collection methods, and which evaluation components you choose to implement. See chapter 4 for an in-depth discussion of evaluation timelines and budgeting for your evaluation.

2.2 Creating a Preliminary Evaluation Plan for an Implementation Study

The implementation study (also known as a “process study”) illuminates “what happened,” and makes evaluation results more interpretable by fully describing the context for the evaluation’s findings. Of particular importance are program implementation and fidelity (determining whether the program was implemented as designed), program participation rates, and program completion rates. All of these can affect, as well as help you clarify the evaluation results by helping the reader understand what happened “on the ground” and what the program looked like. The implementation study findings can also be used to inform future programming by describing the program features that worked well and those that did not work when put into practice.

When planning for an implementation study, you should consider the following elements:

Research Questions

- The research questions should focus on measuring the program delivery process and its fidelity to the program model (i.e., how the program was originally intended to be delivered). You may also want to capture information on participants’ level of participation (e.g., the number of sessions attended).
- As with the broader outcomes or RCT evaluation, the research questions should relate back to the logic model (particularly the activities and outputs listed) and should be stated as specifically as possible.

Data Sources and Collection

- Implementation evaluation data sources typically consist of individual interviews with stakeholders, focus groups with stakeholders and/or participants, program observations, and collection of programmatic documents and data.
 - For interviews and focus groups, think about whom you will speak with, why it is important to include them in the study, and how you will recruit them to participate.
 - Also consider your program administrative data on participation in program activities and completion. Program data can provide a wealth of information to answer questions about how your program works, and how participants utilize services.

- In your preliminary evaluation plan, discuss which data sources will be used to answer which research questions.

2.3 Creating a Preliminary Evaluation Plan for a Cost Study

Type A WIF applicants are required to propose a cost allocation analysis. Type B and C applicants are required to propose a cost-effectiveness analysis. Therefore, inclusion of a preliminary plan for the appropriate type of cost study is essential, and will provide useful information to future practitioners debating whether or not to implement your intervention.

A **cost allocation analysis** provides documentation on program operational costs at the per-participant or per-system level, and usually include the average cost per participant. These studies look only at the costs of a program. You could also estimate the cost per-unit/participant depending on such things as location or other factors. In your preliminary evaluation plan, you should discuss how you plan to track costs and what different types of costs are important to track (e.g., salaries, equipment, and support services provided). How will you learn about costs per participant at different participation levels? How will you learn about costs of the program at different sites?

A more advanced type of cost analysis is a **cost-effectiveness analysis**, which asks “What is the cost, per unit, for the results obtained?”^[1] This kind of analysis examines costs in terms of a single outcome. In the context of an employment and training program, the outcome could be placement, employment (ever employed), or employment meeting specific criteria (e.g., in terms of wages, benefits, retention). A cost-effective program is one that delivers its key outcome at a reasonable cost per outcome (i.e., at a cost that is similar to or lower than comparable programs).

For example, you could plan to examine how much was spent, on average, for each program participant placed in a job. To answer this question, you would compare the total spending to the number of participants who obtained a job. In the preliminary evaluation plan, you should lay out the outcomes you plan to include, and indicate how you will track spending. The specific analysis techniques and methods for quantifying costs will be developed at a later time in consultation with your third-party evaluator.

When planning for the cost allocation or cost-effectiveness study, you should consider the following elements:

^[1] Bell, S. “Improving the Evaluation of DOL/ETA Pilot and Demonstration Projects: A Guide for Practitioners.” U.S. Department of Labor, Employment and Training Administration, 2001, pg. 109.

Research Questions

- Decide what research questions you plan to answer with your cost study. Specify the types of costs you will examine, and, for cost-effectiveness studies, what outcomes you will include. Think about different groups of people for whom you want to calculate costs, and also about different “packages” of services for which you could estimate costs. Rather than only looking at “total cost,” think about various breakdowns of costs that would be interesting to examine and useful for future practitioners to learn about.

Measures

- Consider the various cost measures you will estimate. This could include staff time, equipment, and any other budgetary aspects that are expected to be changed by the intervention. Next, you will need to link each cost measure with its data source.

Data Sources and Collection

- Data sources for cost information should be clearly specified. This may include internal financial records or databases, programmatic records and budgets, or other administrative data. You should also think about the frequency of data collection and reliability of the data sources (such as budget or accounting reports). The availability of the data for the external evaluators should also be considered.

3. Planning for Your Evaluation Phase 2

The evaluation basics and steps for preliminary planning described in the previous chapters bring you to the point of actually implementing the evaluation. While the evaluation activities will be the responsibility of the third-party selected to conduct the evaluation, as the entity contracting with the evaluator, you will need to understand evaluation study's timeline, resources needed (staff and funding), and other factors that could affect its quality. This chapter describes four evaluation activities that program staff should be aware of and understand. These include: developing the Evaluation Design Report, which will guide all aspects of the evaluation; understanding how and when the evaluation activities link to program activities; reporting by the evaluator on interim and final study findings; and the protecting the rights of participants involved your study.

3.1 Creating an Evaluation Design Report

The first task of the third-party evaluator you select to conduct the evaluation of your program will be to develop a plan for carrying out the evaluation. Typically called an *Evaluation Design Report* (EDR), this document will serve as a guide for the evaluator and program staff to the unfolding of various evaluation activities and how they will link to or coordinate with program operations.

The EDR will build upon your preliminary evaluation plan (see chapter 2). The selected evaluator should expand and refine each element in the preliminary evaluation plan to create a comprehensive and feasible plan.

In addition to expanding upon the elements of your preliminary evaluation plan in the EDR, your evaluator will propose appropriate and reliable outcomes that can be measured with given resources and data. Additionally, the evaluator will ensure that the evaluation method(s) you envisioned are appropriate and will expand upon the methodological plans. The evaluator will also be responsible for designing suitable controls for reducing any threats to successful interpretation of findings, including preparing appropriate study designs and describing how they will be implemented so that any risks are mitigated and any limitations overcome to the maximum extent possible. Finally, the evaluator should

include a timeline for evaluation activities, as well as details about reporting on the evaluation.

The full list of EDR elements is included in Table 3-1. Most of these elements are described in chapter 2 with the exception of *Analysis, Sampling, and Methods*, and *Validity and Threats*. These two technical EDR elements will be addressed in future guidance directed to evaluators. The table below is provided for information only to help familiarize you with the purpose of the EDR and the areas it will cover. Development of the EDR will be the responsibility of the evaluator selected to conduct your evaluation, although it will build on the work completed for the preliminary evaluation design.

Table 3-1: Elements to be Included in the Evaluation Design Report

EDR Element	Summary Description
1. Intervention Purpose and Scope	The intervention’s purpose and scope are described. Each of the components of the intervention is spelled out (if more than one). This section also describes which aspects of the intervention are being evaluated, and which are not. Expected outcomes from the intervention are clearly described here. Possible effects of other activities or interventions on the variables of interest beyond those being evaluated are specified.
2. Literature Review, Evidence Base, and Evaluation Contribution	The evidence base is reviewed and synthesized. Relevant past interventions and evaluations are included (with discussion of research designs used, if applicable). The report describes how the evaluation will enhance the broader workforce system or contribute to the workforce evaluation literature.
3. Logic Model	The logic model reflects the inputs, activities, outputs, and outcomes for the specific intervention beyond general program operations.
4. Study Purpose	The purpose of the study is described. The study examines the effectiveness of an intervention (as opposed to a descriptive, process study).
5. Research Questions	The evaluation design includes research questions that are linked to the specific intervention and align with the logic model. There are hypotheses about expected outcome changes due to the intervention.
6. Participants, Samples, Population, and Units of Analysis	The report describes: the unit of analysis; the eligibility or exclusion criteria for intervention participants (i.e., the target population, if applicable); the overall population targeted for the intervention or from which generalizations will be made; whether the evaluation will be conducted on the entire population vs. a sample (and if a sample, whether the sample is representative).

EDR Element	Summary Description
7. Evaluation Design a) Type b) Rigor	The design type and justification for this approach is provided. The analytic methods are described and are appropriate to the evaluation design type and proposed intervention.
8. Data a) Definitions and Measurement b) Sources and Collection	This section defines the specific inputs or activities expected to produce primary outcomes, and proposed outcome measures that are valid, appropriate and reliable. The data sources are described for each measure. The EDR describes whether the data are from new sources (primary data collection), or existing administrative sources (secondary data collection (e.g., from an MIS or financial reporting), and details whether the sources are unique to the intervention and evaluation. The specific methods and schedule for collecting the data from each of the sources are delineated and are deemed reliable.
9. Analysis a) Sampling b) Methods	If there is a sampling plan, the EDR describes the purpose and method of sampling; provides anticipated sample sizes; and for RCT designs (if not discussed above) shows power calculations. The process for random assignment is described, as well as any impact formulas and other analytical assumptions. This section includes model specifications, and describes the analytical software or other tools that are appropriate for the kinds of data and evaluation design.
10. Validity and Threats	This section of the EDR discusses any issues of internal and external validity, as well as threats to validity, and considers their implications. Strategies to mitigate selection bias, if needed, are also described.
11. Implementation Study	In this section, the implementation study plans are described, including a discussion of research questions, data sources and collection, and analysis methods.
12. Cost Analysis	In this section, the EDR describes the plans for the cost analysis, including research questions, variables, data sources, data collection, and analysis methods.
13. Reporting	The plans for reporting the evaluation results are provided, as well as plans for providing evaluation data sets, including providing a de-identified data set to the National Evaluation Coordinator (NEC) at the conclusion of the study.

3.2 Linking the Evaluation to the Intervention

The evaluation should be closely connected to the intervention (or program) throughout the grant period. It is the responsibility of both the grantee and the evaluator to ensure that the evaluator has sufficient access to project



activities and staff to obtain information and conduct evaluation activities, while still maintaining sufficient independence to be objective. In practice, this means that coordination and communication with the evaluator will be important and can be operationalized through regular communication mechanisms, such as conference calls/meetings and ensuring the evaluator is informed of project design decisions. It is important for program staff to understand the significance of the evaluation and the need to be responsive and provide information as required by the evaluator. Finally, program staff people that are in communication with program participants must be able to clearly articulate the requirements of the evaluation (for example, they should be able to explain the random assignment process if an RCT design is being used). Training is often provided by the evaluator to inform program staff about the evaluation and ensure that they work with the evaluation processes and requirements.

There are typically several key time points during the evaluation when the evaluator will need access to program staff to obtain information about activities. These time points include when baseline data collection takes place; when implementation of random assignment design occurs; and when additional data collection (such as follow-up surveys) is conducted. At each of these points, you should work closely with your evaluator to find the best way to integrate the study activities into program activities. Close and sensible integration of evaluation activities into program activities will have many benefits, including reduction of burden on staff, informed program applicants or study participants, minimal effect of the evaluation on your program (ideally, the evaluation should have no effect on the program), and buy-in to the evaluation among both staff and study participants.

Sensible integration of evaluation activities into program activities will have benefits:

- *Reduced burden on staff and participants*
- *Clear information for participants*
- *Minimal effect of the evaluation on the program*
- *Buy-in from staff and participants*

Baseline data collection

Often conducted at the time of program enrollment, this data collection activity may require program staff to collect additional information from participants. If staff will be involved in conducting this data collection, they will need to be trained by the evaluator. You and your evaluator should discuss when baseline data should be collected, and how this data collection can best be integrated into your typical enrollment processes so as to reduce burden on both staff and study participants. For example, if you already collect some data on participants at enrollment, you may want to consider combining the evaluation data

collection and your standard data collection into one document; or, you may want to consider administering both forms simultaneously to ensure that you and your evaluator receive all of the information needed.

Random assignment process (for RCT design evaluations)

Assignment of individuals to the “treatment” condition or the “control” condition (those who will receive status quo services) is also conducted at program enrollment. Again, if program staff will be involved in this process, they will need to be trained by the evaluator. It will also be important for you and the evaluator to decide jointly on the best point in the program flow or intake flow for conducting random assignment. You will want to integrate the random assignment and other processes required for this type of evaluation into your standard intake procedures so as to lessen the burden on program staff and to adequately inform applicants about the processes and requirements for enrolling in the program.

Typical Points for Integration:

- *Baseline data collection*
- *Random assignment process (for RCT designs)*
- *Interim data collection*
- *Post program exit data collection*

For example, you could add the random assignment step to your enrollment process in this way: When participants come to an information session, they complete eligibility screening forms. Those who are eligible move on to another session during which the study is reviewed, the participant’s consent is obtained, and the baseline data collection form(s) is completed. Then, a staff member uses this information to conduct random assignment using a system designed by your evaluator. While there are many ways to integrate random assignment into your enrollment process, the key is to do it as smoothly as possible so as to not lose applicants’ interest and not burden program staff.

Data collection activities

The evaluator will determine an appropriate timeframe for data collection, and should coordinate well ahead with the program staff if needed. If the data collection involves surveying participants, you will want to remind study participants about data collection activities when they are on the horizon. The evaluator may ask you to endorse any data collection effort whenever you make contact with participants, or may ask to send a letter on your behalf. While you should not assume a large amount of burden for this data collection effort, it will benefit both your program and your evaluation if you do what you can to encourage participants to respond.

3.3 Reporting on the Evaluation

The reports that your evaluator will produce about your evaluation are the culmination of all of the evaluation activities; they convey the evaluation's findings to relevant stakeholders, including you, DOL, the larger workforce community, and other potential funders. As described in the first chapter of this toolkit, the evaluation findings can help you improve program operations, pursue cost efficiencies, and support program sustainability. They may also help other communities or programs implement best practices.



Not all reports come at the end of the evaluation, however. Some evaluations include a component that provides feedback at the beginning of a program to shape program design and implementation. **Interim reports** typically share findings about the intervention as it is being implemented. **Final reports** serve as the official record of the evaluation and can present outcome or impact and implementation findings.

The next section describes the types of reports you may want your evaluator to produce and provides guidance on how to clearly communicate reporting expectations in your RFP.

Types of Reports

- *Final Report*
- *Interim Report(s)*
- *Progress Reports*

Types of Reports

Aside from the evaluation design report, the only report that is *required* for the WIF grant is a final evaluation report. However, you may also want your evaluator to produce an interim report(s) to convey findings to you at evaluation mid-point(s), or progress reports to provide updates on evaluation activities.

- The **final report** serves as the definitive report on the study and may be published on the Department of Labor's website. Final reports are helpful to you, policy makers, and the workforce community because they give a complete snapshot of the program from start to finish, and include results on participant outcomes.
- You may also benefit from **interim reports**, which can convey findings about the intervention as it is being implemented. Interim reports will allow program staff to learn more about short- and medium-term participant or systems outcomes during program implementation. An interim report could be submitted at the evaluation's mid-point, or

you could require yearly reports that include some description of the program implementation, participant characteristics, and short-term outcomes or impacts.

- **Progress reports** are written reports provided on a regular basis (e.g., monthly, quarterly, yearly) that provide updates on evaluation activities during the given time period. These reports can help you (1) determine if an evaluator is on-track to complete the evaluation within the allotted time and resources, (2) understand what activities are being undertaken, and (3) maintain open communication throughout the course of the evaluation.

Key topics for evaluation reports:

- Intervention being implemented and tested by the program
- Economic, geographic, and/or political context that may have contributed to program implementation or results
- Information on the methodologies researchers have used to answer research questions and analyze data, including implementation and cost analysis
- Information on the sources of both quantitative and qualitative data
- Program operations as they have been planned and implemented
- Analysis of quantitative and qualitative data
- Interpretation of results
- Identification of best practices, including how other programs may use information in the report to replicate or scale-up programs like the one being studied

Incorporating Reports into your RFP

The RFP for evaluation services should clearly communicate your reporting expectations. To ensure that you receive appropriately budgeted proposals from prospective evaluators, you should explicitly state how many reports your team will expect and when you will expect them. Note that increasing reporting requirements will increase evaluation costs. You should carefully balance the value of the information desired with cost considerations.

Before writing your RFP, you should do the following:

- Identify what your team would like to learn. Do you want evaluation reports at an earlier phase in the project to inform practice? Is a final report sufficient?
- Identify how often your team would like reports. If your team determines that it will only request a final report, you will need to decide when that report should be delivered for your review. In establishing the reporting timelines, you should take into account the following:

- Final reports should allow for the follow-up period for tracking participant outcomes. For example, if evaluators will be tracking participant outcomes 12 months after program enrollment, you will want to ensure that the final report is not due until after the 12-month period is over for the last participants enrolled.
- Evaluators will need time for analysis, writing, and revisions. This process can take anywhere from two to five months, depending on the sample size, complexity of analysis, and number of revisions.

If you will be requesting interim reports, you should determine how many reports will be necessary, and when they should be delivered. In the past, grantees have requested a mid-project report (interim report), annual reports (one interim report for each year of the grant cycle), or some combination of the two, depending on the resources available. When determining when these reports should be delivered, you should think about your program intervention and when it might be reasonable to expect results. For example, if your program requires six months of training and your main interest is in participant outcomes, it probably does not make sense to request an interim report before the first group of enrollees has completed the training.

3.4 Protecting Participants Rights

An additional and significant aspect of planning for your evaluation and considering how it will be conducted is the issue of protecting the rights of the individuals your program serves when they are part of an evaluation study. As part of your program evaluation efforts, you and your evaluator may collect and store detailed information about individuals who agree to participate in your study. For projects involving individual-level data, the information study participants provide will allow you and your evaluator to measure the effectiveness of your program. As such, the study participants, or human subjects,⁴ are the heart of your research project. How you and your evaluator treat them and their information is of critical importance. In fact, protecting the rights of study participants will be one of your most fundamental responsibilities; it is also a requirement for federally funded research.

⁴ Defined in the Code of Federal Regulations 45 CFR 46.102(f), see: <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102>

A human subject is anyone about whom an investigator conducting research gains (1) data by way of intervention or interaction with the individual or (2) **identifiable private information**.⁵

Private information includes information that an individual can reasonably expect will not be made public, and information about behavior that an individual can reasonably expect will not be observed or recorded.

Identifiable means that the identity of the individual is or may readily be ascertained by the investigator or associated with the information.

Personally Identifiable Information (or PII) is information that would allow someone to identify, contact, or locate another person or to identify an individual in context. PII includes names, social security numbers, dates of birth, addresses, and other contact information. It also includes any piece of personal information that can be used alone or in combination with another source to uniquely identify a single individual.

Federal grantees should be mindful of the requirements for protecting human subjects from a project's inception to its completion. Even as early as the grant application stage, you should be aware of human subject considerations and plan accordingly. By doing so, you will support the smooth running of your project and avoid the negative consequences of failure to comply with federal and state regulations.

Human subject considerations have very practical implications for your evaluation timeline and budget, and they may also affect other major and minor features of your project such as:

- The evaluator you will choose (their concern for protecting human subjects, experience with human subjects review, data security capacity).
- How you and your evaluator will structure data collection (the types of data gathered and how).
- How much time you need to allow for Institutional Review Board (IRB) application (see below for more information) and review.

⁵ Defined in the Code of Federal Regulations 45 CFR 46.102(f), see: <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102>

- The additional staffing resources needed to conduct informed consent. The technology (encryption software, secure file transfer protocol (FTP)) needed to securely transfer and store participant information.

This section is designed to inform you about the responsibilities and requirements of conducting research with human subjects. At the end of this section, under the heading “Important Considerations,” you will find a summary of key considerations and essential steps for prospective grantees regarding human subjects research.

Human Subjects Protections

Federally-funded research and evaluation involving human subjects must comply with federal and state laws and regulations governing the ethical treatment and the rights of research participants. Most of these policies are based on the **Belmont Report** (1979)⁶, a summary of ethical principles and guidelines for protecting research participants. The Belmont Report acknowledges both the social benefit that research has produced as well as the ethical issues it has sometimes posed. It states that research studies must be guided by three ethical principles: **Respect for persons, beneficence, and justice**. Your evaluator must carry these principles into their research protocols and activities.

Respect for persons:
Consent to be in a study should be informed and voluntary.

The most commonly applied policy is the Federal Policy for the Protection of Human Research Subjects⁷, also known as “**the Common Rule**.” In order to comply, each organization that conducts research must have or hire an Institutional Review Board (IRB) to ensure the protection of human subjects. Other steps researchers must take to protect human subjects include:

Beneficence:
Do no harm; maximize benefits of research and minimize risks to participant.

- Seeking informed consent from study participants,
- Preserving participants’ privacy and confidentiality,
- Maintaining strict data security standards, and
- Reporting any adverse events and unanticipated problems.

Justice:
Fair and equitable distribution of the benefits and burdens of research.

What this means is that conducting research is not as easy

⁶ See <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>

⁷ See <http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html>

as gathering and analyzing data: it requires a plan for study recruitment, consent, data collection, data management, analysis, and reporting that ensures that human subjects are protected. Even studies that use administrative records and never directly survey or interview participants are subject to human subjects considerations.

Institutional Review Board

Even the most well-meaning researcher may overlook the effect of their work on study participants. To ensure appropriate protections are put in place and maintained, Institutional Review Boards (IRBs) serve as an independent and objective ethics committee to ensure the protection of human subjects. An IRB's job is to approve (or disapprove) human subjects research—or require modifications prior to approval. IRB review should occur prospectively—that is, before study procedures are implemented. IRBs are also responsible for providing continuing review by monitoring active research projects, reviewing any unanticipated problems, and reporting serious adverse events to regulators. IRBs have the authority to suspend or terminate research that does not comply with the Common Rule in order to protect research participants from harm. The main roles and responsibilities of IRBs are summarized in Table 3-2.

You should ensure your evaluator consults with the IRB about their study plans, and receives approval (if required) for the study before engaging human subjects. The IRB will let them know whether the study is “exempt” from review or is “non-exempt” and needs to undergo a formal review by the board. Federal regulations provide clear guidelines for this process. The decision about whether a study's protocol requires IRB review—and the level of review (expedited or full board)—depends on a variety of factors, including: the level of risk to participants overall and relative to the potential social benefit of the research, the vulnerability of the population under study, and the steps the evaluator will take to minimize risks and safeguard participants. Generally speaking, studies that may pose minimal or greater than minimal risk to participants will need to be reviewed by the IRB. What constitutes exempt research, minimal risk, or greater than minimal risk should be decided by an IRB—not the evaluator.

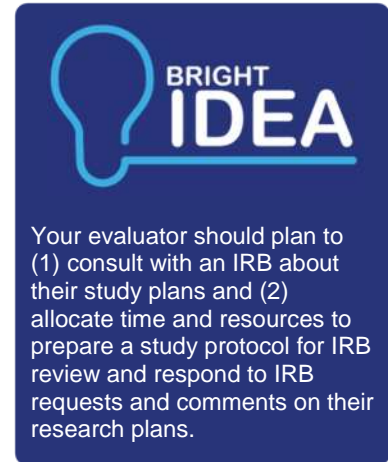


Table 3-2: IRB Responsibilities and Study Approval Criteria

Responsibilities of IRBs	Criteria Needed to Obtain IRB Approval
<ul style="list-style-type: none"> • Prospectively review and approve study procedures • Review “unanticipated problems” and adverse events • Observe and monitor studies (e.g., observe consent process, audit consent forms) • Suspend or terminate studies if needed to protect the safety of participants • Report serious adverse events to the appropriate regulators • Train researchers, evaluators, and other key team members on ethical standards to protect participants 	<ul style="list-style-type: none"> • Minimize risks to participants • Benefits to society and/or participants outweigh the risks • Select participants equitably to distribute burden • Obtain and document informed consent process • Monitor welfare of participants for safety and complaints • Minimize risks to privacy and confidentiality • Additional safeguards to protect rights and welfare of vulnerable populations (e.g., children, pregnant women, “persons who are economically or educationally disadvantaged”)

IRB Review: What to Expect

If IRB review is required, study activities involving human subjects (e.g., study recruitment, data collection) cannot begin until the IRB has approved the study. Thus, the evaluator will need to allocate adequate time and resources during the design phase to preparing an application for IRB review and respond to any IRB requests and comments on their research plans. Approval is communicated by a formal letter and an official IRB stamp for consent forms.

The evaluator will need to allocate adequate time and resources to preparing an application for IRB review and response to any IRB requests.

How much time will review take? The process could take anywhere from one to several months. The time it takes to prepare an application for IRB review, have it reviewed, respond to comments or requests for revisions, and receive approval depends on both the complexity of the research project and the level of risk it poses to participants. The evaluator can begin preparing the application as soon as the study design has been determined; the procedures for recruiting, consenting, and enrolling participants have been laid out; and the protocols for collecting and securing data have been defined. The evaluator will need to dedicate time up front to documenting these and other aspects of

their research plan and prepare an application for IRB review. They will also need to allow time for the board members to review and, if needed, discuss the application (they usually meet on a regular monthly or quarterly schedule).

Review will be ongoing. Collaboration with the IRB does not cease with the receipt of the approval letter and stamp. As mentioned, the IRB will monitor the study on an ongoing basis while it is active. This is done, in part, by requiring studies to re-up their approval on a regular (usually annual) basis through a “continuing review” submission.

Critical Factors in Protecting Human Subjects

Three major areas for consideration during the IRB review process include: the study’s procedures for informed consent; how the evaluator will protect privacy and confidentiality (including the plan for data security); and how the evaluator will handle adverse events and unanticipated problems. These factors have very real, on-the-ground implications for how the study will unfold. You will want to be aware of and may be involved in executing these aspects of the study protocol.

Informed Consent

In order to collect information from and about individuals participating in a research study, the evaluator will need to obtain their legally effective informed consent (or that of a parent or guardian if the participant is a minor (under 18)). Obtaining informed consent includes giving prospective study participants sufficient opportunity to consider participation and minimizing “undue influence” and “coercion.”

Undue influence:

Offering excessive or inappropriate reward for participation or causing someone to feel obligated to participate.

Informed consent is a *process* (not just a form) that begins with explaining the study. The information provided to prospective participants during the consent process must help them understand the implications of participating in the study they are being asked to be a part of. It is essential to disclose all relevant information honestly and to give each individual the opportunity to ask questions and receive answers to their questions.

Coercion:

An overt or implicit threat of harm such as telling a parent/guardian or young woman that they will lose access to services if they do not participate.

Privacy and Confidentiality

Once participants consent to the study, the study team is required to protect their privacy and the confidentiality of their information to prevent its accidental disclosure or loss.

- Protecting participants' **privacy** means collecting data in such a way that participants' information is not seen or overheard by others. (For example, they take surveys in a private space where others cannot view their answers.)
- Protecting the **confidentiality** of participants' information means: (1) not sharing information about participants except to those authorized to have it and (2) complying with a study-wide plan for secure collection, transfer, storage, and use of participants' information.

Data Security

Grantees and evaluators have a collective responsibility to protect participants' information. All parties will need to work together to create a climate of accountability and responsibility when it comes to data. Part of this includes establishing a plan for data security. This is important because there may be penalties for losing data—even if the data is not misused—and for failing to destroy data once it is no longer needed. Moreover, no one wants to inform another or be informed that their information was lost, stolen, or otherwise compromised.



Robust data security can help safeguard against the accidental loss or disclosure of participants' information. You and your evaluator will need to work out the specifics as they will apply to your evaluation.

Adverse Events and Unanticipated Problems

In research as in life, things do not always go as anticipated. For this reason, the evaluator should have a plan for handling adverse events and unanticipated problems. Examples of unanticipated problems include the following:

- A participant is visibly upset by the questions during the survey.
- You learn that a participant is at risk of harm.
- A participant or their parent/guardian has serious concerns about the study.
- Study procedures were not followed (i.e., participant consent, data security).
- Study data are lost (e.g., consent forms, paper and pencil surveys).

If these issues occur during the evaluation of your intervention, you should work with your evaluator, and in turn the National Evaluation Coordinator, to help resolve them.

Important Considerations for Prospective Grantees

Protecting the rights of study participants will be one of your fundamental responsibilities. It requires planning, documentation, coordination, and consistent attention. It means taking steps that may feel overly cautious or cumbersome to those new to the process. Looked at differently, it can be a point of pride knowing that the individuals volunteering to contribute to your study are being treated with the respect, beneficence, and justice that you would expect for yourself or someone you care about.

Protecting the rights of study participants will be one of your fundamental responsibilities.

When seeking an evaluator you will want to:

- Make sure the evaluator is knowledgeable about human subject protections and has experience with IRB review. A good question to ask is whether they have an IRB in-house (most universities and many independent evaluation firms do) or what external IRB they would consult about the study.
- Consider the evaluator's access to the technology and infrastructure that will allow them to securely collect, transfer, store, and maintain data files.

When preparing your research plans (either for your grant application or when working with the evaluator on the research design), you will want to:

- Make sure adequate resources and attention are devoted to IRB review and human subject protections.
- Make sure that your evaluator has an in-house IRB or has worked with one previously.
- Understand the evaluator's plans for collecting data and how it will be securely transferred, stored, and maintained.

Finally, while the study is ongoing you will want to:

- Ensure that your evaluator has an in-house IRB or has hired one for the study.
- Monitor your evaluator's progress in applying for, obtaining, and annually updating IRB approval.
- Understand the evaluator's protocol for protecting human subjects; handling adverse events; and securely collecting, transferring, and maintaining data (particularly PII).
- Ensure that all relevant staff understand and comply with these human subject and data security protocols to the extent that your organization and its partners are involved in

executing the study (e.g., gathering study consent, administering surveys, and transferring data).

4. Evaluation Timeline and Budget

Accurately planning the timeline and budget for your evaluation will ensure that enough time is allotted for critical evaluation activities, such as data collection or report development and dissemination; and that sufficient resources are available to successfully carry out the activities. This chapter focuses on these two key aspects of planning. To develop your timeline and budget, your team should determine (1) what kind of evaluation you will implement based on the type of WIF grant you are applying for, (2) the tasks that are likely to be required to conduct the evaluation, and (3) the time and level of staff effort associated with each of these tasks. This section includes common evaluation tasks that you may wish to incorporate into your evaluation timeline and provides guidance on factors that may affect your evaluation budget.

4.1 Evaluation Timeline

It is useful to begin the planning by creating the timeline. This will help you to identify the specific tasks required as well as the expected level of effort for each task, which in turn will help you to develop the budget. While every evaluation is unique, several activities are common to all evaluations and should be included in your timeline. The table below provides examples of common evaluation tasks to consider while planning your evaluation timeline and budget. The table also provides estimates for when the tasks occur and how long they typically last. These estimated timeframes are not set in stone and should be regarded only as a starting point to begin your consideration of timeframes. The length of time needed for many of these tasks will depend on numerous variables, such as the type of evaluation being conducted, the size of the project, the number of stakeholders involved, the length of your intervention, and how well-established the intervention is. Furthermore, this is not a complete list of evaluation activities; you may have additional evaluation activities in your timeline and budget, which will likely be further refined down the line by your evaluator. This set of sample activities and timing is presented to help you better understand your evaluator's likely scope of work and budget, and to help in creating a first draft of your timeline. Tasks listed in the table are ordered in typical chronological order, although many tasks will occur simultaneously.

Table 4-1: Sample Evaluation Activities and Timing

Task	Activities Associated With this Task	Who Is Responsible	Time It Takes
Intervention and Evaluation Planning Activities			
Planning your intervention	Finalize intervention details related to such aspects as program flow, service delivery, roles and responsibilities of partners, and data sharing.	Grantee with some evaluator assistance once procured	<1-6 months
Procuring the evaluator	Write and publicize the RFP; review applications; select an evaluator; write and negotiate the contract with the selected bidder; approve the selected evaluator and negotiate contract through a local procurement process.	Grantee	<1-5 months
Designing the evaluation	Conduct initial meeting to discuss intervention and evaluation goals; define intervention being tested; define research questions of interest; define measures, outcomes, and data sources of interest; determine appropriate follow-up period; develop and revise evaluation design report (EDR); develop data security and human subject protection plan; get Institutional Review Board approval.	Evaluator and Grantee	2-6 months
WIF NEC Review	Submit draft EDR for WIF NEC review and comment; review and respond to comments.	WIF NEC	1-2 months
Intervention and Evaluation Start-Up			
Recruitment	Create recruitment materials; obtain IRB approval of all recruitment materials (if applicable); implement recruitment activities.	Grantee and Evaluator	Initial Activities: 1-3 months Recruitment: Ongoing
Data Measures and Systems	Create shared measures and data collection forms with partners (if applicable); formalize data sharing agreements/create a plan for accessing the data; make necessary modifications to existing data systems; assign data entry responsibilities.		2-4 months

4. Evaluation Timeline and Budget

Task	Activities Associated With this Task	Who Is Responsible	Time It Takes
Participant Enrollment	Conduct eligibility screening and informed consent process; enroll participants into the study; conduct baseline data collection; make random assignment (if applicable); notify participants of random assignment status (if applicable); notify participants of any additional steps necessary to take part in the program (e.g., orientation, filling out paperwork, drug tests).	Grantee and Evaluator	Ongoing while new participants entering study
Ongoing Evaluation Activities			
Data Collection	Varies depending on design and research questions, examples include: conduct data collection from non-program sources (e.g., Unemployment Insurance records, National Directory of New Hires, secondary or postsecondary education records, public assistance); transfer program data from grantee systems to evaluator; conduct calls and site visits to learn more about the program as its being implemented; develop survey instruments and conduct of survey(s) and follow-up survey(s) as necessary.	Grantee (data entry and sharing program records) and Evaluator	Data entry: 2-6 hrs/wk Data collection: 30 hrs/yr Data transfer: 4 hrs/qtr Surveys and follow-up: 3-12 months ⁸
Quality Control	Ensure participants receive only assigned services; check data for completeness and quality; troubleshoot or clarify data issues.	Grantee and Evaluator	10 hrs/qtr
Evaluation Check-ins	Participate in regular (weekly/monthly) grantee-evaluator calls to discuss evaluation implementation activities and the intervention.	Grantee and Evaluator	1 hr/wk

⁸ Follow-up periods will be determined by grantees and evaluators. Typically, workforce research has used follow-up periods of 6-18 months after program entry or random assignment (for experimental evaluations).

4. Evaluation Timeline and Budget

Task	Activities Associated With this Task	Who Is Responsible	Time It Takes
Analyses			
Interim Reports	Check data for quality and comprehensiveness; analyze data collected to date; write and revise (as appropriate) report.	Evaluator	3-4 months
Final Report	Check data for quality and comprehensiveness; analyze all data collected during evaluation; include outcome/impact analysis, cost studies, implementation studies, and so on as applicable; write and revise (as appropriate) report.	Evaluator	3-4 months
Public Use Data Set	Check data for quality and comprehensiveness; remove all PII (e.g., names, social security numbers, addresses); write comprehensive codebook (i.e., a guide describing each of the variables in the data set).	Evaluator	2–6 weeks

As mentioned above, many of these tasks will occur simultaneously. For example, applicants should procure their evaluator while they are planning and finalizing their intervention.

The time needed to plan the evaluation design will also vary by project, but the design should be final before the end of Year 1. Note that evaluation design activities have consequences for the intervention implementation timeline. A WIF grantee cannot begin to implement its intervention until the Department of Labor approves the grantees Evaluation Design Report. This means that all areas of concern with the evaluation design, as highlighted by the National Evaluation Coordinator or the Department of Labor must be resolved prior to implementation of the intervention. Evaluators are often asked to revise and refine their design plans before receiving approval. Therefore, the evaluation timeline should allow one to two months for WIF NEC review as well as time for at least one subsequent revision of the evaluation plan. It will be important to inform all partners of any potential impacts that the evaluation design process may have on program implementation.

Evaluation design activities have consequences for the intervention implementation and timeline.

4.2 Evaluation Budget

Developing a realistic evaluation budget will ensure that evaluators have sufficient resources to carry out activities successfully and to provide you with a high-quality evaluation. The activities in your timeline will comprise the major cost of the evaluation. The staff time and level of effort that will be required to carry out the activities must be covered in the budget. A number of other items also need to be considered when budgeting for the evaluation. These items are described in the table below.

Table 4-2: Factors Affecting Evaluation Budget

Factors that May Affect Your Evaluation Budget	Considerations
Evaluator Qualifications	More experienced evaluators are more likely to have the knowledge and expertise to conduct a high-quality evaluation, and as a result can be more expensive than less-experienced evaluators. Because of the trade-off between cost and experience, many evaluators will propose an evaluation team that uses a mix of senior, mid-level, and junior staff to meet quality needs while maintaining cost-competitiveness. See chapter 6 of this toolkit for more information on what you should consider when selecting an evaluator.
Sample Size	Sample size refers to the number of participants included in the evaluation. A larger sample size can give you more precise findings. However, as your sample size increases, some of your evaluation costs will also increase, including costs related to recruitment, participant surveys, data collection and analysis, and staff time needed for evaluation procedures.
Internal Data Systems	All evaluations require some kind data collection to answer research questions. Grantees need to have systems in place to record program data. Evaluators need to have systems in place to collect data from grantees and other sources. In some cases, existing data systems can be used for the evaluation. In other cases, data systems will need to be modified to accommodate the evaluation (e.g., additional variables need to be captured in your MIS; the MIS needs to be modified to conduct random assignment). It is likely that your evaluation may need to work across some number of data systems to capture the information you need. If you know that you will need to change your existing data systems, you should budget the funds needed to modify them.
Data Collection—	Administrative data from sources beyond the program are often

Factors that May Affect Your Evaluation Budget	Considerations
Administrative Data	critical for the evaluation analysis (e.g., federal, state, and local workforce agencies, including Unemployment Insurance). Collecting data from these sources may have an associated cost.
Data Collection—Surveys	If the research approach includes conducting surveys (such as surveys of staff or participants), this work will need to be considered and planned for in the budget. Costs for conducting surveys and follow-up surveys include staff time to develop and revise the survey instrument, time for tracking people down, the financial incentives for survey completion (if used), postage (if the postal service will be used to deliver the survey), technology for creation of the instrument and staff time for synthesis of collected information. Survey costs will increase with sample size.
Travel	Evaluation teams typically visit grantee sites (once to several times) to better understand how grantees' programs work on the ground and to monitor evaluation procedures. If your evaluator is not based locally, you should plan to budget for at least some transportation, lodging, meals, and incidentals for one to two evaluators over the lifecycle of your grant.
Planning and Revisions	Planning your interventions and evaluation is critical to ensuring a successful project experience. Since much of program development and evaluation design is iterative, it is strongly recommended that you budget for a number of planning meetings and several rounds of revisions to major deliverables. This includes any round of revisions requested by the Department of Labor or its agents.
Coordination with the WIF NEC	As described in chapter 6 of this toolkit, the WIF NEC is a resource for all WIF evaluators. The NEC offers webinars and evaluation resources and conducts periodic check-ins with evaluators to assess progress. Your evaluator should budget at least some staff time to coordinating with the NEC and attending quarterly technical assistance events.

Final Note on Budgeting—Considering Tradeoffs

When you are budgeting for your evaluation, you want to ensure that you are making appropriate trade-offs between quality and cost competitiveness. If a potential bid seems too good to be true, do not hesitate to ask the evaluator more questions about the services they plan to offer you, who will be providing these services, and their experience providing these services in the past. While there will be some cost variation in potential bids, all bids

should be responsive to the key needs outlined in the RFP and their budgets should be reflective of the associated level of effort.

5. Selecting Your Evaluator

A high-quality evaluation will provide the most complete source of information on your program’s effect on participant and system-wide outcomes. However, as you have learned from reading this toolkit, not all evaluations are equal in terms of quality. To obtain a high-quality evaluation, you need to select an evaluator with the expertise and capacity to conduct the evaluation design type you are pursuing. This chapter provides guidance on the qualities to look for in potential evaluators and how to advertise your Request for Proposals (RFP) to attract the best possible bidders.

5.1 What Do You Want in an Evaluation Team?

An important first step in finding a high-quality evaluator is deciding what you are looking for in a potential evaluator. Before writing your RFP, your team should determine the level and type of experience and qualifications you want the evaluator to have. Below are some questions that may help you determine which qualities are most important when selecting your evaluator.

Table 5-1: Questions to Ask a Potential Evaluator

Topic	Questions To Ask
Evaluator Experience	<p>Has the evaluator demonstrated that the firm or its staff has experience conducting an evaluation of your design type?</p> <p>Has the evaluator conducted an evaluation in this content field (e.g., transitional jobs, job training, youth employment) or a closely related topic?</p>
Staff Qualifications	<p>Has the evaluator proposed a team that includes a mix of seniority levels (e.g., will senior-level, mid-level, and junior-level staff all be involved)?</p> <p>What are the qualifications, both in education and experience, of the proposed Project Director and Principal Investigators (or proposed team members in similar lead roles)?</p> <p>What skills will evaluators need to conduct the evaluation as you envision it (e.g., proficiency in data collection and analysis, observational techniques)? Has the evaluator demonstrated that either the firm or its staff possess these skills?</p> <p>If the evaluator has proposed subcontractors as a part of the evaluation</p>

Topic	Questions To Ask
	team, do they also have the qualifications necessary to conduct the evaluation?
Evaluator Independence⁹	<p>Is the evaluator an independent third-party that can assess the program and its effects on participants objectively?</p> <p>Are there any potential conflicts of interest that may be in violation of local, state, or federal procurement processes?</p> <p>If you have worked with this evaluator in the past, what firewalls or other mechanisms exist to ensure that this evaluation can be conducted in an objective manner?</p>
Capacity and Resources	<p>Is the proposed evaluation team of sufficient size to adequately carry out the tasks that are likely associated with your evaluation design type? For example, an evaluation team conducting a random assignment study in multiple sites would likely require more staff than an evaluation team conducting an outcomes study in a single site.</p> <p>Has the evaluator demonstrated that they have the capacity and resources to (1) collect data (e.g., secure data transfer sites); (2) analyze data (e.g., statistical analysis programs); and (3) provide technical assistance on any evaluation-related issues?</p> <p>What kinds of administrative, information technology, and publication and dissemination support does the evaluator have to aid in conducting the evaluation and producing deliverables?</p>
Specialized Knowledge	Does the evaluator have other specialized knowledge that makes it a strong candidate? For example, is the evaluator familiar with the geographic, political, or other context(s) that is important to the study?

It is reasonable to expect a qualified evaluator to have (1) experience implementing an evaluation of the proposed evaluation design type or in the proposed content field; (2) senior staff with a post-graduate education and 5+ years of experience that can demonstrate the technical skills necessary to implement the study; (3) the capacity and resources to facilitate meetings, data collection, data analysis, and report writing; and (4) other specialized knowledge that your team has determined to be important to understanding the program. Overall, evaluator and staff qualifications are often the single most important factors in ensuring a quality evaluation. Evaluation teams that use a mix of seniority levels will have the appropriate experience to conduct the evaluation but will also

⁹ Please note that in-house evaluations conducted by grantee organizations' staff will not be considered sufficient for the evaluation component of the WIF. Grantee organizations can participate in baseline and administrative data collection but not in follow-up data collection or analysis.

be cost-competitive. After deciding on the qualifications you expect your evaluator to have, you should report them in the RFP to ensure that only qualified applicants apply.

Proof of Qualifications

When responding to your RFP, evaluators will need to submit proof of their qualifications. Below are some examples of the types of documentation you can expect to receive, and may wish to require, as a part of the response:

Table 5-2: Evidence of Evaluator Experience and Expertise

Topic	Examples of Acceptable Evidence
Evaluator Experience	<ul style="list-style-type: none"> • Descriptions of previous studies that the evaluator has conducted using the proposed evaluation design type or in the proposed content field • Previous publications or other deliverables from these studies • Description of the company’s mission, history, and experience • Referrals from previous clients • Demonstrated record of re-hire by previous clients
Staff Qualifications	<ul style="list-style-type: none"> • Resumes or C.V.s • Publications written by senior staff demonstrating their familiarity or skills in a particular evaluation type or context
Capacity and Resources	<ul style="list-style-type: none"> • Description of the proposed team • Description of the firm’s administrative and IT support • Description of any data collection and/or analysis tools created and/or utilized by the firm in previous studies • Excerpts of products produced for previous clients (when publicly available).
Specialized Knowledge	<ul style="list-style-type: none"> • Description of specialized knowledge of your topic of interest

Assessing Bidder Applications

Ideally, you will receive several applications for your evaluation. To systematically assess and easily compare these applications, you may find it helpful to create a rubric, or point system, to rate evaluators’ qualifications. Rubrics allow your team to allocate more points to those qualifications or characteristics that your team values most and identify the evaluators that will likely be a good fit for your needs. For example, while evaluator and staff experience should always be among the most important factors, you may also value the evaluator’s content knowledge or experience in your geographic area. You can assign weight to each factor to help you find the strongest evaluator for your project.

Getting the Word Out—Publicizing your RFP

Once you have determined the scope of work and the characteristics of your ideal evaluator, you will need to write and advertise your RFP. Evaluators come in all shapes and sizes and so advertising to a broad audience will ensure that your needs are well publicized and that you will receive a good mix of proposals from which to choose. The table below lists likely evaluator types and where you might find more information about these kinds of evaluators:

Creating a point system, or rubric, to rate evaluator’s qualifications will help identify the bidders that meet your evaluation needs.

Table 5-3: Evaluator Types and Sources

Types of Evaluators	Where To Find Them
Research Evaluation Firms	<ul style="list-style-type: none"> American Evaluation Association (http://www.eval.org/), Association for Public Policy Analysis and Management (http://www.appam.org/)
University Research Centers, Faculty or Affiliated Academics	<ul style="list-style-type: none"> University websites, usually with a focus on relevant departments.
Independent Research Consultants	<ul style="list-style-type: none"> American Evaluation Association (http://www.eval.org/), Association for Public Policy Analysis and Management (http://www.appam.org/)

You may also be able to identify potential evaluators as you conduct your review of existing research-based evidence; take note of research authors, evaluation firms, and consultants noted in the relevant literature. You may be able to reach them with your advertising, as described below.

Ideally, your RFP will result in numerous qualified responses from which you can select the best candidate. You can increase your chances of getting several responses by advertising your RFP through multiple vehicles. Consider the following ways to advertise your RFP:

- Send letters or emails that announce the RFP to a likely group of evaluators.
- Post a notice about your RFP on a number of evaluation-focused websites.
- Post a notice about your RFP on your organization’s website.

- Announce the RFP on any social media sites associated with your organization, including but not limited to LinkedIn, Twitter, and Facebook.
- Advertising the RFP in local, state, or national evaluation newsletters or publications.

Cast a wide net in order to receive the most responses.

All advertisements should comply with your organizational, state, or local procurement processes.

6. Working with the WIF National Evaluation Coordinator (NEC)

Carrying out a high-quality evaluation means that the results will provide valuable information about the program that will help the grantee and the population it serves as well as the broader workforce delivery system. To help ensure that WIF grantees develop a quality evaluation that helps build strong evidence about effective workforce interventions, DOL engaged a National Evaluation Coordinator (NEC) for WIF.

The NEC is an important resource for WIF grantees and third-party evaluators. The primary goal of the NEC is to provide guidance and technical assistance so that evaluation methods are strong and findings are reliable. The NEC also helps to relay information to evaluators about WIF-specific evaluation expectations and deliverables. This section of the toolkit describes the role of the NEC and provides grantees with information about what to expect when working with the NEC, once a WIF grant has been awarded.¹⁰



6.1 Providing Technical Assistance

First and foremost, the NEC's goal is to help WIF grantees achieve reliable results from their evaluation. To do this, the NEC offers technical assistance to strengthen evaluation methods and practices. Assistance is delivered through several mediums, including webinars, guidance documents, references, and a discussion board that will facilitate peer-to-peer learning. Grantees can also contact the NEC directly anytime there is an issue, question, or problem with the evaluation.

Webinars

The NEC delivers webinars to all WIF grantees and evaluators that are designed to proactively provide guidance on issues and problems commonly faced in evaluations. The NEC also develops webinars to address issues or problems that WIF evaluators identify over

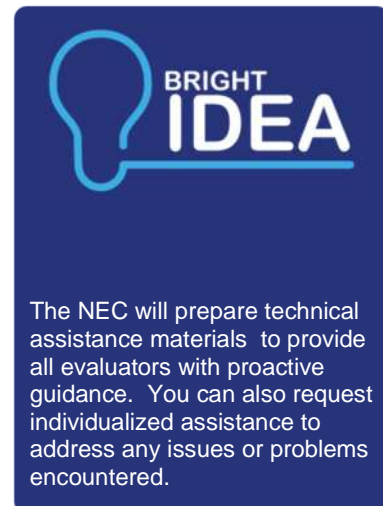
¹⁰ The WIF NEC is primarily a resource for WIF grantees post-award. However, the NEC also develops general guidance, such as this toolkit.

the course of the evaluation. Webinars are also used to communicate WIF-specific expectations (e.g., what should be included in the Evaluation Design Report, the process for NEC deliverable review, timelines for deliverables). Evaluators are expected to attend the webinars, and grantees are also welcome to participate. In the past, the NEC has delivered webinars on topics such as the following:

- Existing evidence for various workforce intervention types
- Common data elements for workforce evaluations
- Choosing the right evaluation type
- Guidance on logic models, data collection, random assignment protocols, and power calculations
- Components of the WIF Evaluation Design Report

Guidance Documents

The WIF NEC also provides evaluators with written guidance documents and resources. Examples include lists of evaluation resources, literature reviews for various workforce development strategies, and outlines for the Evaluation Design Report specific to the evaluation type (e.g., outcomes evaluation, random assignment evaluation). You can also request additional resources from the NEC (e.g., recommended references for a given analysis method, guidance on creating an informed consent form and process). All technical assistance resources are made available on a SharePoint site accessible by WIF grantees, third-party evaluators, Federal Project Officers (FPOs), and the DOL National Office. While the target audience for these resources is typically WIF evaluators, the resources may also help ensure a common understanding of evaluation activities among WIF grantees, FPOs, and other stakeholders.



Online Discussion Board

WIF grantees and their third-party evaluator also have access to an online discussion board established by the WIF NEC. The discussion board is hosted on a password-protected SharePoint site and is designed to facilitate peer-to-peer learning. On the board, grantees may pose questions to the NEC or to other evaluators, share lessons learned or successes, or have discussions about general evaluation topics. All third-party evaluators are encouraged to participate in the discussion board. The NEC also uses the online discussion board to deliver announcements (e.g., upcoming webinars, approaching deadlines).

Individual Communication

The NEC is always available to provide individualized evaluation technical assistance to all WIF grantees and third-party evaluators. Grantees can contact the NEC by email; all questions, comments, or issues sent to the NEC email are assigned to the team member who is best-suited to respond. Because questions are triaged (e.g., questions about random assignment are sent to an experimental evaluation expert, questions about analysis methods are sent to methodologists), the NEC requests that all questions/issues be sent via email. The NEC may either answer the question by email or may request a conference call to discuss the issue further.

6.2 Review of Evaluation Processes and Deliverables

As noted above, the WIF NEC's role is to make sure grantees end up with an evaluation that is high in quality and provides reliable results by following industry standards for evaluation. As such, the NEC works directly with each third-party evaluator to make sure the proposed evaluation plans and formal deliverables do not include issues or problems that would devalue the evaluation. To do this, the NEC reviews and provides feedback on evaluation plans and deliverables. While the NEC can review all evaluation materials upon request, at a minimum the third-party evaluators should submit the following documents to the WIF NEC for review:

- **An Evaluation Design Report (EDR).** After NEC's review of the EDR, the third-party evaluator will revise and resubmit a Final EDR that addresses any NEC comments, as appropriate. Chapter 3 of this toolkit provides introductory information about the components of the EDR. Forthcoming information will provide additional, detailed guidance for evaluators about preparing EDRs.
- **A performance data template.** This document lists the performance and evaluation measures and key milestones of the project, which should be consistent with the evaluation plan submitted with the grantee's application. The performance data template should be submitted to both the DOL Federal Project Officer (FPO) and the NEC.
- **An Analysis Plan and Data Collection Plan.** This document details plans for collecting and analyzing data from the WIF program.
- **Evaluation data collected by third-party evaluators.** This data should be submitted at the end of the grant period. Before submission, the data should be cleaned of all identifying information (i.e., names, social security numbers, or other identifying information for the participants).
- **The final evaluation report.**

6. Working with the WIF National Evaluation Coordinator (NEC)

Reports and other deliverables sent to the NEC are reviewed by multiple team members to ensure that the feedback is comprehensive and accurate. Review feedback is sent to third-party evaluators, who are expected to make revisions as necessary. The NEC review process is designed to strengthen the WIF evaluation by ensuring that the methods used and interpretation of results are appropriate.

The NEC is also responsible for observing and documenting each evaluation's progress and providing status updates to DOL. To prevent issues with evaluation design, implementation, analysis methods, and timing, the WIF NEC stays in close contact with all WIF evaluators (through periodic conference calls or email exchanges) to keep track of evaluation progress.

6.3 Collecting and Synthesizing Final Results

At the conclusion of the grant period, the WIF NEC will provide an overall analysis of the WIF third-party evaluations. To do this, the NEC will collect all data, analyses, and final reports prepared by the third-party evaluators and develop a final report that discusses lessons learned from the overall effort and from the third-party evaluations.

Appendix

Even if you are not conducting an evaluation yourself, working with an evaluator requires an understanding of several terms and concepts that many practitioners may not be familiar with. The glossary provided here contains definitions for common evaluation terms and concepts that you may encounter when discussing your program evaluation. It should provide you with a basic understanding of evaluation concepts, which will allow you to work effectively with the evaluator you select. The glossary is not comprehensive, but the terms and concepts included are some of the most frequently used in evaluation design, implementation, and analysis.

While this toolkit is designed to provide a broad overview of evaluation concepts, should you wish to develop a deeper understanding of evaluation concepts and analysis methods, you can use the reference section after the glossary as a starting point. You can use these references yourself or you can share them with your evaluator. It is important to note that neither the WIF NEC nor DOL endorse the cited materials or their authors. However, the WIF NEC has reviewed the references and recommends them for self-study.

Glossary of Terms

Attrition: Loss of subjects from the study sample over the course of the evaluation. There may be many causes for attrition including, for example, program drop-out or relocation.

Baseline Data: Information collected about study participants prior to program participation or random assignment. Baseline data can be used to describe the study sample and measure participant progress.

Comparison Group: A comparison group is a group of study participants whose outcomes and experiences are compared to the treatment group. In an experiment, the comparison group is either exposed to a different treatment or to no treatment (a no-treatment comparison group is called a control group). A control group is created randomly.

Cost Allocation Analysis: Cost allocation is a management tool that involves establishing a budgeting and accounting system with which program managers can determine a unit cost, or cost per unit of service. This type of analysis includes documentation on program operational costs at the per-participant or per-system level. Cost allocation analysis looks only at the costs of a program, and in most cost analyses of employment and training programs, the analysis focuses on unit costs (e.g., per participant, enrollee, or FTE position). Unit costs might be compared across sites, or used in a regression analysis to estimate if and to what extent such things as program scale, geography,

or program approach might predict (or affect) unit cost.

Cost-Effectiveness Analysis: A type of evaluation research that compares program costs to program outcomes. Cost-effectiveness analysis examines costs in terms of a single outcome. This outcome is not monetized. In the context of an employment and training program, the outcome could be placement, employment (ever employed), or employment meeting specific criteria (e.g., in terms of wages, benefits, or retention). A cost-effective program is one that delivers its key outcome at a reasonable cost per outcome (i.e. at a cost that is similar to or less than comparable programs).

Generalizability: The extent to which the study's conclusions based on the sample can be said to represent results for the entire population from which the sample was drawn.

Implementation Study: An implementation study illuminates and explains “what is happening and why” in the design, implementation, administration, operation, services, and outcomes of social programs. This type of study can provide context and information that makes impact evaluation results more useful. Findings from implementation research can be used to inform future program development or replication.

Informed Consent: The agreement given by study participants to take part in the study after having been informed of the nature of the research.

Inputs: Resources that go into a program, such as grant funds, personnel, and equipment.

Institutional Review Board (IRB): A review body consisting of researchers, representatives of the research subjects, and individuals knowledgeable in the rights of human subjects, established or designated by an entity to protect the welfare of human subjects recruited to participate in research. IRB review of all study materials is required in studies that directly affect the participants in some way, such as in a randomized controlled trial study, and may be required in studies that do not involve the participants directly.

Interrupted Time Series: A non-experimental design in which outcomes are measured for a group of participants multiple times, both before and after the intervention. This approach is similar to a pre-post-test design except that measurements are taken at multiple points both before and after the intervention, which provides greater confidence that the outcomes after the intervention resulted from the intervention and not random fluctuation.

Intervention: The program, project feature, or innovation that is being studied.

Logic Model: A description of a program/process that includes a conceptual framework showing the activities and methods being used to achieve relevant outcomes. It provides an overview of a program/process and identifies key components of the process, product, strategy, or practice (i.e. the active “ingredients” that are expected to be critical to achieving the relevant outcomes). The logic model also describes the relationships among the key components and outcomes and can be displayed in the form of graphic and/or by textual descriptions.

Outcomes: The intended results of a process or program (including changes in conditions, such as employment, earnings, or income, as well as changes in attitudes, values, and behaviors).

Outcome Study: Examines the changes in targeted conditions, attitudes, values, or behaviors between baseline measurement and subsequent points of measurement. Changes can be immediate, intermediate, or long-term. An outcomes study seeks to provide information on how individuals fared in the program without attributing causality.

Outputs: What is produced that can be easily described and quantified as a result of program activities (for example, numbers of workshops held or people trained).

Power: Power refers to the ability of a study to detect meaningful program impacts at a given level of statistical certainty.

Power Analysis: A power analysis is used to determine the required sample sizes necessary to reach statistical conclusions (also known as statistical significance). Usually, the results of a power analysis are expressed as Minimum Detectable Impacts (MDI) or Minimum Detectable Effects (MDE). The MDI allows the researchers to know the level of impact the new intervention must have on an individual's desired outcomes, such as earnings and employment, for the impact to be detected with a given sample size and specified probability of error.

Pre-Post Data Analysis: A type of outcomes study where behavior before a program (or a subject's participation in it) began (pre-program) is compared to behavior at a point after the program was completed (post-program).

Qualitative Data: Non-numerical data that provides detail and description (e.g., data from interviews or focus groups).

Quantitative Data: Numeric data that can be analyzed using statistical methods (i.e., data that can be counted, scored, and categorized).

Quasi-Experimental Design (QED): A research design with a comparison group that is similar to the group receiving the intervention in important respects but that does not receive the services being tested. QED designs attempt to approximate an experimental design by using a comparison group, but they do not use random assignment to create a control group that is identical to those in the treatment group.

Randomized Controlled Trial Studies/Random Assignment Impact Evaluation: A research design in which participants are randomly assigned by lottery to a treatment group that receives services or a control group that does not receive services (or to one of two or more treatment groups). The difference between the average outcome for the treatment group(s) and for the control group is an estimate of the effectiveness of the intervention. Most social scientists consider random assignment to be the only way to assure that observed effects are the result of a given program and not of other factors.

Reliability: The degree to which a measurement or measurement instrument produces consistent results over time.

Representative Sample: A sample that mirrors the population from which it was selected in all the respects potentially relevant to the study and its outcomes.

Sample: A subset of a larger population that is used to study the population as a whole.

Statistical Significance: The mathematical likelihood that an observed effect is due to chance. Statistical significance is usually expressed as a p-value, with a smaller p-value meaning that the outcome is less likely to be due to chance and more likely is a true change or effect.

Target Population: The group larger than or different from the population sampled to which the researcher would like to generalize study findings.

Theory of Change: A theory of change is a way to explain your underlying understanding of the issue you are addressing—it clarifies why you are doing what you are doing. It is a description of a program that includes a clear identification of the population for which it is intended as well as the theoretical basis or description of the expected causal mechanisms by which the intervention should work. Theories of change are often represented visually.

Treatment Group: In an experiment, the treatment group is the group that receives the intervention(s) being tested. Also called the experimental group.

Unit of Analysis: The unit of analysis is the major entity (the “what” or “who”) that is being analyzed for the study. The unit of analysis can be, for example, individuals, groups, geographical units (e.g., cities, states, countries), or social interactions.

Validity: The degree to which a test accurately measures what it intends to measure.

Evaluation References

General Program Evaluation Resources

- S. Bell. (2001). "Improving the Evaluation of DOL/ETA Pilot and Demonstration Projects: A Guide for Practitioners." Research and Evaluation Report Series 01-A. U.S. Department of Labor, Employment and Training Administration, Washington, DC. Available at: http://wdr.doleta.gov/research/eta_default.cfm?fuseaction=dsp_resultDetails&pub_id=2445&bas_option=Author&start=1&usrt=4&stype=basic&sv=1&criteria=bell
- S.R. Khandker, G. B. Koolwal, and H.A. Samad. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. World Bank, Washington, DC. Available at: http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUBOEPI1101Official0Use0Only1.pdf
- P. Rossi, M. Lipsey, and H. Freeman (2003). *Evaluation: A Systematic Approach*. 7th Edition. Sage Publications, Thousand Oaks, CA.
- What Works Clearinghouse, "Procedures and Standards Handbook (Version 2.1)." Available at: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf
- J.S. Wholey, H.P. Hatry, and K.E. Newcomer, eds. (2010). *Handbook of Practical Program Evaluation*. Jossey-Bass, San Francisco, CA.

Cost Analysis

- H.M. Levin and P.J. McEwan. (2001). *Cost-Effectiveness Analysis: Methods and Applications*, Second Edition. Sage Publications, Thousand Oaks, CA. [Includes information on both cost-effectiveness and cost-benefit analysis.]
- P. Z. Schochet, J. Burghardt, and S. McConnell. (2006). "Impact and Benefit-Cost Findings Using Survey and Summary Earnings Records Data: Final Report." Mathematica Policy Research, Princeton, NJ.
- H. M. Levin, C. Belfield, F. Hollands, A. B. Bowden, H. Cheng, R. Shand, and Y. Pan. (2012). *Cost-Effectiveness Analysis of Interventions that Improve High School Completion* Teachers College, Columbia University, New York.

	<ul style="list-style-type: none"> • H. M. Lewin, E. Garcia, and J. Morgan. (2012). <i>Cost-Effectiveness of Accelerated Study in Associate Programs (ASAP) of the City University of New York (CUNY)</i>. Columbia University, New York.
Implementation Study	<ul style="list-style-type: none"> • J. Century, M. Rudnick, and C. Freeman. (2010). "A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge." <i>American Journal of Evaluation</i> 31:2. • M. Lennon and T. Corbett. (2003). "Policy into Action: Implementation Research and Welfare Reform." The Urban Institute Press, Washington, D.C. • M. Scheirer and E. Rezmovic. (1983). "Measuring the Degree of Program Implementation: A Methodological Review." <i>Evaluation Review</i> 7:5. • A. Werner. (2004). <i>A Guide to Implementation Research</i>. The Urban Institute Press, Washington, D.C. • The National Implementation Research Network (NIRN) is a center that supports research and prepares resources on measuring implementation. See: www.fpg.unc.edu/~nirn.
Interrupted Time Series	<ul style="list-style-type: none"> • H.S. Bloom. (1999). "Estimating Program Impacts on Student Achievement Using 'Short' Interrupted Time Series." Manpower Demonstration Research Corporation, New York. • H.S. Bloom. (2003). "Using 'Short' Interrupted Time-Series Analysis to Measure the Impacts of Whole-School Reforms: With Applications to a Study of Accelerated Schools." Manpower Demonstration Research Corporation, New York. • D. Ho, K. Imai, G. King, and E.A. Stuart. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." <i>Policy Analysis</i>, 15, 199-236. • V.J. Hotz, G.W. Imbens, and J.A. Klerman. (2006). "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Re-Analysis of the California GAIN Program." <i>Journal of Labor Economics</i>, 24:2 (July):521-566

Logic Model

- Knowlton, L.W. and Phillips, CC. The logic model guidebook: Better strategies for great results. Los Angeles: Sage; 2009.
- McLaughlin, J., and Jordan, G. (1999). Logic models: A tool for telling your program's performance story. *Evaluating and Program Planning*, 22, 65-72.
- W. K. Kellogg Foundation. (2004). Logic Model Development Guide. Available at: <http://www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx>
- J. Century, Freeman, C., and Rudnick, M. (2008). A framework for measuring and accumulating knowledge about fidelity of implementation (FOR) of science instructional materials. Presentation at the Annual Meeting of the National Association for Research in Science Teaching, March, 2008.
- A. Cassata-Widera, J. Century, D.Y. Kim.. (2011). Rigorous measures of implementation: A methodological framework for evaluating innovative STEM programs. Presentation at the SREE Fall 2011 Conference, Washington, D.C.
- B. Lawton, P.R. Brandon, L. Cicchinelli, and W. Kekahio. (2014). Logic models: A tool for designing and monitoring program evaluations. (REL 2014–007). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific. Available at: <http://ies.ed.gov/ncee/edlabs>.
- M. Nelson, D. Cordray, C. Hulleman, C. Darrow, and E. Sommer. (2012). "A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services & Research*.
- L. Knowlton and C. Phillips. (2012). The Logic Model Guidebook: Better Strategies for Great Results, Second Edition. Sage Publications, Thousand Oaks, CA.

Outcome Study

- R. DeVellis (2003). *Scale Development: Theory and Applications*. 2nd Edition. Sage Publications, Thousand Oaks, CA.
- H. Hatry (1999). *Performance Measurement: Getting Results*. The

Urban Institute, Washington DC.

- L. Martin and P. Kettner (1996). *Measuring the Performance of Human Service Programs*. Sage Publications, Thousand Oaks, CA.
- J. McDavid, I. Huse, and L. Hawthorn. *Program Evaluation and Performance Measurement: An Introduction to Practice*. 2nd Edition, Sage Publications, Thousand Oaks, CA.
- J. King, L. Morris, and C. Fitz-Gibbon. (1987). *How to Assess Program Implementation*. Sage Publications, Thousand Oaks, CA.

Power Analysis

- Bloom, H. (2006). "The Core Analytics of Randomized Experiments for Social Research" MDRC Available at: http://www.mdrc.org/sites/default/files/full_533.pdf
- S.W. Raudenbush, Martinez, and Spybrook (2007). "Strategies for Improving Precision in Group-Randomized Experiments." *Educational Evaluation and Policy Analysis*, 29(5).
- S.W. Raudenbush. (2011). *Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01)* [Software]. Available from www.wtgrantfoundation.org.

Quasi-Experimental Design

- T. Cook and D. Campbell. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin, Boston, MA.
- R. Dehejia and S. Wahba. (1999). Causal Effects in Non-Experimental Studies: Re-Evaluating The Evaluation of Training Programs. *Journal of the American Statistical Association* 94:448 (December 1999), 1053–1062.
- S. Guo and M. Fraser (2009). *Propensity Score Analysis: Statistical Methods and Applications*. Sage Publications, Thousand Oaks, CA.
- G. Imbens and T. Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142:2.
- D. Lee and D. Card (2008). "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics* 142 :2.
- D. Rubin and N. Thomas. (March 1996) "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52:1.

	<ul style="list-style-type: none">• W. Shadish, T. Cook, and D. Campbell. (2002). <i>Experimental and Quasi-Experimental Designs for Generalized Causal Inference</i>. Houghton Mifflin, Boston, MA.• J.D. Singer and J.B. Willett. (2003) <i>Applied Longitudinal Data Analysis</i>. Oxford, U.K.: Oxford University Press.
Randomized Controlled Trial	<ul style="list-style-type: none">• V. Alferes (2012). <i>Methods of Randomization in Experimental Design</i>. Sage Publications, Thousand Oaks, CA.• T. Cook and D. Campbell. <i>Quasi-Experimentation: Design and Analysis Issues for Field Settings</i>. Houghton Mifflin, Boston, MA.• W. Shadish, T. Cook, and D. Campbell (2002). <i>Experimental and Quasi-Experimental Designs for Generalized Causal Inference</i>. Houghton Mifflin, Boston, MA.• J.D. Singer and J.B. Willett. (2003) <i>Applied Longitudinal Data Analysis</i>. Oxford, U.K.: Oxford University Press.
Theory of Change	<ul style="list-style-type: none">• J. Reisman and A. Gienapp. (2004). Theory of Change : A Practical Tool For Action , Results and Learning. Organizational Research Services, Seattle, WA. Retrieved from http://www.aecf.org/upload/publicationfiles/cc2977k440.pdf.• D. H. Taplin and H. Clark. (2012). Theory of Change Basics (pp. 1–8). ActKnowledge, New York, NY. Available at: http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/ToCBasics.pdf.

Blank Template for Development of a Logic Model

Program: _____ **(name)** _____ **Logic Model** (uses text boxes: add/change boxes and arrows as needed)

Situation:

