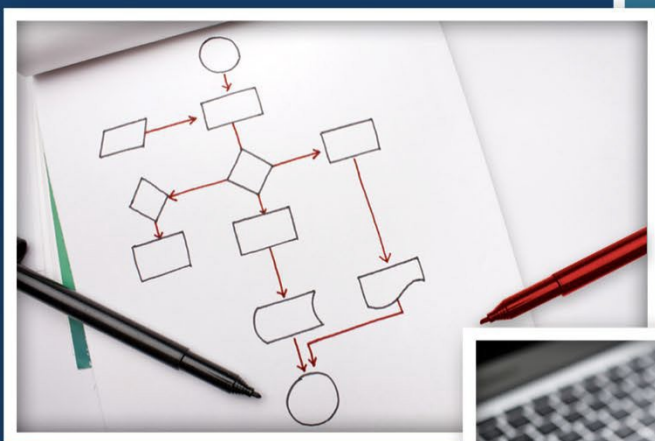




U.S. Department of Labor
Employment and Training Administration
Office of Policy Development and Research
Division of Research and Evaluation

Evaluation Toolkit: Key Elements for State Workforce Agencies

September 2020



Acknowledgements:

This toolkit provides a basic overview of evaluation elements for program management purposes and a two-part quick action planner to assess evaluation readiness. It also draws from information about other generally accepted and available evaluation resources that may be useful as an entry point for state workforce agencies who may not fully utilize evaluations in program planning or implementation, or who may need additional evaluation reference to expand evaluation activities.

Coffey Consulting, LLC, with support from the U.S. Department of Labor's Employment and Training Administration (DOL-ETA), Office of Policy Development and Research staff, developed this toolkit under Contract #1630DC-17-U-00005. Wayne Gordon, Dan Ryan and Gloria Salas-Kos offered initial guidance and provided evaluation reference resources and materials, including foundational evaluation tools developed for the Workforce Innovation Fund grantees. Randall Denison, Michelle Ennis, and Charlotte "Sande" Schifferes also provided feedback related to the Workforce Innovation and Opportunity Act of 2014 and the subsequent evaluation-related regulations. Under Contract #1630DC- 17-U-00026, IMPAQ International, LLC, provided additional technical assistance and developed the WIOA Quick Action Planner (QSAPs) for the evaluation readiness and evaluation design and implementation assessments.

Updated by Safal Partners, October 2019, for the U.S. DOL-ETA, Office of Policy Development and Research under Contract# 1630DC-18-A-0005, Task Order#1630DC-19F00024. Special thanks to Dr. Margaret Patterson for her work in reviewing and providing content updates.

Disclaimer: The resources, templates, and information provided in the Evaluation Toolkit serve as a general guide for states and local areas. Although every effort is made to ensure that the material within this document is accurate and timely, we make no warranties or representations as to the accuracy or completeness of the contents, whether the contents are current, or free from changes caused by third parties. All information is provided "as is" without warranty of any kind. No information provided in this document may be considered legal advice and it is the responsibility of each user of the Evaluation Toolkit to ensure that their materials meet all Federal, state, and local requirements. Use of the materials does not imply compliance with ETA requirements.

Table of Contents

Introduction	1
1. State Evaluations and WIOA	3
1.1 Purpose of Evaluation	3
1.2 Context for Evaluations in State Workforce Agencies	5
1.3 Gaps in and Benefits of Evaluation Activities	6
2. Broaden State Research and Evaluation Capacity.....	8
2.1 Expand or Enhance Evaluation Capacity	9
2.2 Develop an Evaluation Plan or a Research Learning Agenda	14
3. Consider Evaluation Types and Timeframes.....	17
3.1 Four Major Types of Program Evaluation	17
3.2 Evaluation Combination and Continuum of Study	24
3.3 Create an Evaluation Timeline	24
4. Develop an Evaluation Design Plan.....	28
4.1 Develop the Logic Model	28
4.2 Determine the Evaluation Purpose and Scope	33
4.3 Develop the Key Research Questions	34
4.4 Review the Existing Evidence Base	35
4.5 Determine the Research Design and Data Collection Approach	38
4.6 Prepare the Preliminary Evaluation Plan	40
5. Select an Evaluator.....	43
5.1 What Approaches to Consider	43
5.2 In-House Unit/Staff Approach	45
5.3 Partnership Approach	45
5.4 Third-Party Approach.....	46
6. Implement the Evaluation	48
6.1 Create the Evaluation Design Report.....	48
6.2 Develop the Data Analysis Plan	50
6.3 Protect Participant Rights	51
6.4 Coordinate Evaluation and Program Activities	54
6.5 Evaluation Report Requirements.....	55
Conclusion.....	58

Appendices.....	A1
APPENDIX A: Performance, Research and Evaluation Framework.....	A2
APPENDIX B: WIOA Evaluation Regulations.....	A3
APPENDIX C: Evaluation Resources and References.....	A5
1. National and Federal Policy, Planning, and Guidance Resources.....	A5
2. Evaluation Resources and Online Toolkits.....	A7
3. Behavioral Insight Studies.....	A9
4. Cost Studies.....	A9
5. Data Analytics.....	A9
6. Implementation Studies.....	A10
7. Interrupted Time Series.....	A10
8. Logic Models.....	A10
9. Outcome Studies.....	A11
10. Power Analysis.....	A12
11. Quasi-Experimental Studies.....	A12
12. Randomized Controlled Trial (RCT) Studies.....	A13
13. Theory of Change.....	A13
APPENDIX D: The Learning Agenda Cycle.....	A14
APPENDIX E: Logic Model Template.....	A15
APPENDIX F: Evaluation Readiness Assessment—Overview and Instructions.....	A16
1. Overview of Assessment.....	A16
2. Instructions for Completing the Assessment.....	A16
APPENDIX G: Glossary of Terms.....	A27

Table of Exhibits

2.1 Department of Labor Evaluation Principles	9
2.2.1 Building a State Evaluation Plan or Research Learning Agenda.....	15
2.2.2 Examples of High-level State Research Questions	16
3.1.1 : Types of Program Evaluation—Implementation Studies	19
3.1.2 : Types of Program Evaluation—Outcome Studies.....	20
3.1.3 : Types of Program Evaluation—Impact Studies.....	21
3.1.4 : Types of Program Evaluation—Cost Studies	23
3.3 Sample Evaluation Activities and Timing.....	26–27
4.1.1 Graphic Logic Model Example.....	31
4.1.2 Logic Model Questions.....	32
4.4 Sources for Finding Workforce Evaluation and Research Studies	37
4.5 Data Requirements and Sources for Different Evaluations.....	39
4.6 Key Design Elements for Evaluation Plans	42
5.1 Evaluator Selection Questions and Evidence of Experience.....	44
6.1 Elements Included in the Evaluation Design Report, as Appropriate	49
6.2 Items to Cover in a Data Analysis Plan (RCT).....	50
6.3 IRB Responsibilities and Study Approval Criteria.....	52

Introduction

This publication presents information to help State Workforce Agencies (SWAs) implement provisions on evaluation in the Workforce Innovation and Opportunity Act (WIOA). Commissioned by the Employment and Training Administration (ETA) of the U.S. Department of Labor (DOL), it is part of a larger technical assistance (TA) effort to build evaluation capacity in state workforce systems.

Planning strategies and implementation tools, presented in the form of a toolkit, build upon technical assistance, research, and guidance previously developed by ETA's Office of Policy Development and Research, DOL's Chief Evaluation Office (CEO), and other sources. The information shared in this toolkit also includes a working definition of evaluation and describes the purpose, context, rationale, and types of evaluation as key elements to build upon or expand state evaluation capacity. In addition to the key evaluation elements, this guide also summarizes states' evaluation activities and themes documented in an assessment and scan conducted by the National Association of State Workforce Agencies (NASWA). The latter part of the toolkit details a "how-to" guide for designing evaluations, most of which comes from technical assistance material created to promote independent, high-quality evaluations under the Workforce Innovations Fund (WIF) grants.

Each toolkit section provides information about either evaluation fundamentals or expectations specific to state evaluation efforts, with each section providing information that later sections build on. The toolkit uses a broad-based perspective to maximize its usability. Some sections may be more useful than other sections, depending on the user's level of evaluation knowledge and expertise. For those who want to delve into evaluation design, additional resources and references organized by specific topics are included in the Appendix. The information in each toolkit section combined with the additional resources provides an easy-to-use, one-stop TA guide.

Section 1 provides an overview of evaluation purposes and key concepts, including differences between program performance, program monitoring, program research, and program evaluation. This section also describes WIOA state evaluation requirements, highlights common evaluation practices and presents a summary of research and evaluation gaps identified in the NAWSA scan. Section 1 sets the stage for the remainder of the toolkit.

Section 2 describes DOL-ETA evaluation policy, evaluation plans, and research learning agendas; it presents information to broaden state research and evaluation activities that include building relations, leveraging existing funding sources, expanding staff skills and capacity, using and sharing administrative data, and leveraging and integrating data. This section also describes the need to develop a state evaluation plan or learning agenda.

Section 3 outlines the types of evaluation, provides key considerations for evaluation designs, and reviews key elements for developing a timeline for an evaluation project to use under varying circumstances.

Section 4 delves into implementing evaluation activities, evaluation design reports, and data analysis plans; protecting participant data; and presenting interim and final study findings.

Section 5 describes the processes and steps to consider when selecting an evaluator—from utilizing in-house staff to using external partners, such as institutions of higher education or contracted third party research firms.

Section 6 covers factors to consider when determining who may conduct the evaluation.

The **Appendices** include a table that delineates the differences between performance research and evaluation and presents the WIOA regulations on evaluation at 20 CFR 682.220, with 13 different sets of resources and reference to consider in developing evaluation designs. The appendices also include a template to assess an agency's evaluation capacity, an example of the learning agenda process, a Logic Model template, and a glossary of the most common evaluation terms.

While this toolkit is a guide with key evaluation elements for SWA managers and staff, it also may be of interest to state and local Workforce Development Boards (WDBs), American Job Centers (AJCs), and state and local partner agencies. As WDBs and AJCs participate in partnership activities, an awareness of state evaluation planning, design, and data collection and analytics may help with utilizing evaluation findings to improve services and outcomes for the benefit of customers in the workforce system.

1. State Evaluations and WIOA

The Workforce Innovation and Opportunity Act (WIOA), signed into law in July 2014, includes strengthened provisions for state evaluations. The use of evaluation in the public workforce system aligns with an expanding movement in the U.S. toward evidence-based policy. A 2017 report¹ of the [Commission on Evidence-based Policymaking \(CEP\)](#) and a chapter in the 2018 Federal Budget² focused on evidence-based policy describe evaluation as a means to help government agencies “learn what works and what does not, for whom and under what circumstances, and how to improve results.”³

The sections below offer a working definition of evaluation and review requirements in WIOA on evaluation.

1.1 Purpose of Evaluation

The use of evaluation to measure the effective and efficient use of funds for services is a management function for successful government programs. Federal, state, and local policymakers are increasingly aware of the need for rigorous evaluation results to support evidence-based policymaking, and thus making well-informed decisions about program investments. In this era of limited public resources, information on program effectiveness is especially critical for state workforce administrators in setting the direction for state and local implementation of WIOA—what specific programs, services, or activities to prioritize within the context of formula grant, discretionary grant, and national programs, state economic development priorities, and state labor market dynamics.

Evaluation results, findings, or recommendations are also critical for garnering or maintaining support for a specific initiative. An evaluation explains whether a program produces positive outcomes (e.g., were program participants able to find a job or increase their earnings) and how the program achieved



Evaluation as Used in Toolkit

The term “evaluation” is broadly used to include the study of workforce programs, systems, strategies, services, activities, or interventions. An evaluation study:

1. Examines specific interventions, service elements, or activities within a program;
2. Measures the impact of services on a target population, such as youth with disabilities; or
3. Determines the effect or effects of services using administrative data from a statewide system, such as a career pathways system, or a statewide strategy, such as an industry sector strategy, to name a few possibilities.

¹ See the report at: <https://www.cep.gov/report/cep-final-report.pdf>

² “Building and Using Evidence to Improve Government Effectiveness,” Chapter 6, *Budget of the U. S. Government, A New Foundation for American Greatness, Fiscal Year 2018*, May 2017, p 55, https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/budget/fy2018/ap_6_evidence.pdf

³ More recently, the *President’s Executive Order on Expanding Apprenticeships in America, June 15, 2017, Section 10 Programs*, calls for rigorous evaluations of existing programs. The Office of Management and Budget (OMB) directive providing *Fiscal Year 2019 Budget Guidance, July 7, 2017* calls for agency proposals on building and using a portfolio of evidence and strengthening agency capacity to use evidence, evaluation, and data as tools to improve federal government effectiveness.

results from following an evaluation design plan with a defined methodology and data analytics (e.g., what activities or actions produced the results).

Evaluations can positively affect state workforce agency (SWA) program planning and implementation efforts, the individuals served, strategic policy planning, and funding efforts; as well as SWA coordination efforts with the larger workforce community. Evaluation may help with the following actions:

- Improve specific SWA programs, services, interventions, or activities. Evaluations can reveal whether program service components produce positive outcomes or whether there is a need for other program improvements. Learning that certain program components may not produce the envisioned results may be just as valuable as learning that program(s) have positive results. The SWA can then make changes that may result in improved outcomes for individuals served.
- Use tested or evaluated innovative interventions to increase or improve outcomes for program participants. Program improvements identified through evaluation findings may show more effective or efficient results that lead to better services for participants and produce desired outcomes.
- Determine which state policies to implement, planning activities to continue, and funding priorities to consider as part of WIOA program administration and management, both in the short-term and in the long-term development of a meaningful four-year state plan.
- Secure state or other funding needed to sustain and scale up a state priority initiative. Funders generally seek to invest in proven strategies that are effective. If an evaluation produces positive or promising findings, how the SWA disseminates those results may increase support. Many states, funders, and employers report an interest in continuing and even scaling up program(s) that demonstrate a track record of effectiveness.
- Demonstrate long-term impacts on individuals and communities. Impact evaluations may help emphasize the causal evidence or attributions of the services used in SWA program(s) to produce a change or changes in the existing service delivery area.
- Educate the larger workforce development community. State workforce administrators, evaluation managers, and policymakers may benefit by learning about participation in program evaluations. Positive outcomes or impacts may increase support for similar programs and services in other communities. Also, positive outcomes or impacts may prompt other localities operating different programs to adopt a similar program (or program elements).



Benefits of Using Evaluations

- Improve SWA program(s), services, interventions or activities;
- Increase or improve outcomes for the population served;
- Determine state policy, planning, & funding priorities;
- Secure other funding to sustain or scale up a state initiative;
- Demonstrate long-term impacts; and
- Educate the larger workforce community.

1.2 Context for Evaluations in State Workforce Agencies

The National Association of State Workforce Agencies (NASWA) conducted a state scan on research and evaluation capacity in May 2017.⁴ It includes a summary of an assessment from the responses of 41 states. Designed to collect information about current research and evaluation capacity within SWAs, the national scan summarizes the SWA responses and findings related to research agendas, state agency support, staffing, funding, and data capacity. The scan also includes two case studies—the NASWA report identifies Washington and Ohio SWAs as “high achievers” within the context of each state’s research and evaluation capacity. It also further identifies technical assistance and capacity needs by research skill area. While the NASWA scan allowed the Department of Labor (DOL) to understand the interest or demand for the types of research and evaluations produced, as well as the kinds of partnerships used to fund, conduct, or participate in research and evaluations, the scan found that “only a handful of state agencies reported having sufficient capacity to conduct research on program impacts.”

As restated in the scan, WIOA also emphasizes the use of data to inform workforce development planning and implementation decisions in three key program management areas, namely to: (1) develop state policies and programs; (2) support the front-line delivery of customer services; and (3) address state and local workforce agency accountability through performance outcomes. In essence, WIOA calls for states to use research and evaluation data to make evidence-based policy and program decisions.


However, articulating the differences between data uses for performance management, research activities, and evaluation is somewhat difficult given various state perspectives on terminology use and since the NASWA scan findings relate to the kinds of research and evaluation predominantly conducted by SWAs. The *Performance, Research and Evaluation Framework* presented in **Appendix A** may help SWAs think about program research and evaluation activities within the broader context of their work. This framework identifies the activities and components—with examples of performance tracking, measurement, accountability, monitoring, research, and evaluation—to demonstrate the differences in purpose and execution.

As observed by the NASWA scan, state-conducted research and evaluation activities tend to focus on reporting and analysis of performance outcomes; Labor Market Information (LMI) research, including special analyses; and other forms of descriptive or comparative research and analysis. An emphasis on descriptive and comparative research analysis and reports drives the need to manage state and local performance accountability systems, LMI analyses, and other related research (e.g., Bureau of Labor Statistics [BLS], Federal-State Cooperative Program; ETA Workforce Information Grants for States). However, program performance and labor economic and statistical activities build a foundation for state-conducted evaluations and support evidence-based policymaking.

To build or expand state research and evaluation capacity, this toolkit addresses evaluations for workforce development programs. As such, program evaluation within the broader context of the WIOA regulations, noted in **Appendix B**, describe allowable evaluation methodologies and components as follows:

⁴ “[Evidence-Building Capacity in State Workforce Agencies: Insights from a National Scan and Two State Site Visits](#)” Yvette Chocolaad, Stephen Wander, National Association of State Workforce Agencies

- States may consider a broad array of options in the types of studies considered evaluations. These types of evaluation may include “process and outcome studies, pilot and demonstration projects that have an evaluative component, analyses of administrative and programmatic data, impact and benefit-cost analyses, and use of rigorous designs to test the efficacy of various interventions.”
- State evaluations may include many other components, such as “multiple phases and such tasks and activities as necessary for an evaluation, such as a literature or evidence review, feasibility study, planning, research, coordination, design, data collection, analysis, and report preparation, clearance, and dissemination.”
- In addition to the evaluation design phases and tasks, states can coordinate with WDBs and other state agencies, cooperate with Federal evaluations, disseminate evaluation findings, and use state set-aside funds to conduct evaluations.



Three “Cs” of WIOA Evaluation

- **Conduct** evaluations to promote continuous improvement, new research efforts, and test innovative strategies or services;
- **Coordinate** and design evaluations with state and local workforce boards or other state agencies; and
- **Cooperate** with Labor or Education evaluations by providing data, participating in surveys, and encouraging site visits.

This toolkit highlights state evaluation responsibilities for three applicable approaches that ETA refers to as the “three Cs” of WIOA evaluation: conducting and coordinating evaluations and cooperating with federal partners who conduct evaluations (see inset). It describes in-depth key evaluation elements to consider as SWAs plan, procure, design, and implement evaluations for workforce development programs.

1.3 Gaps in and Benefits of Evaluation Activities

Evaluations, other forms of research, and data analytics exist in varying degrees of use within the public workforce system. The NASWA state scan found that, among the states responding to their assessment, there was, with the exception of one state, “a demand ... for the types of information that workforce agency research and evaluations can produce” and that these requests emerged from state agencies, as well as from the state legislatures and governors’ offices. Many of the “most pressing questions” identified as topics for research and evaluation focused on program outcomes or impacts and understanding the labor market in the state. However, some state agencies also posed questions “aimed at better understanding customers and their barriers” or “improving program operations and administration.”

Of the states that responded to the NASWA scan, 75% indicated the existence of at least one unit that initiates and advances research and evaluation efforts. Furthermore, a high percentage of state workforce agencies use or partner with “outside researchers to conduct at least one research or evaluation effort” over a four-year period. The scan findings indicate “staff capacity and funding, research and evaluation activities, and research and evaluation methods were less encouraging” with “more than half of the [SWAs] ... producing three or fewer in-house research and evaluation studies” over the same four-year period.

At the same time, as the NASWA study found, there were consequences for states that were not able to conduct adequate levels of evaluation and research due to resource constraints. State respondents provided comments that, without sufficient funding for such activities, it was difficult “to anticipate changes in trends and make timely data-driven decisions”; “harder to make sound policy decisions without proper research”; and that it led to “limited knowledge, unknown effectiveness, limited transparency, [and] reduced consumer choice.”

Beyond the SWAs’ capacity and efforts to produce research and evaluation products and concerns with resource constraints, “only a handful of the reporting agencies report having sufficient capacity” related to the “research skill areas most often associated with evidence-based policy-making—conducting experiments and employing quasi-experimental designs.”

Employing rigorous evaluations methods may thus have a vital and practical role in administering workforce programs. For example:

- Evaluations used within required WIOA planning processes may help to understand services patterns and outcomes, determine which participant groups or service areas need additional resources, and which populations might benefit from different interventions.
- Evaluations may be used to formulate policies or address funding priorities, both in the near term and within a meaningful four-year state plan.
- Evaluation can be used to test new interventions that improve outcomes for program participants.
- Evaluations with demonstrated program results may encourage the SWA and other agency or external funders to sustain and scale up the initiative. If an evaluation produces positive outcomes or impacts, or promising findings, it may result in increased support to continue or expand services.
- Evaluations can help build evidence-based practices to benefit the larger workforce community. In addition to SWA-supported evaluations, other state program administrators, evaluation managers, and policymakers can:
 - Benefit by learning how to address evaluation findings;
 - Adopt practices and program models demonstrated as being effective; or
 - Continue to experiment with similar approaches.

Evaluations that provide in-depth information on implementation practices can also help both state and local workforce entities to replicate or adapt a specific program to their settings.



Benefits of Evaluations

Evaluations can:

- Be used when developing state plans;
- Help to understand program operations and differences in outcomes for groups and/or areas;
- Provide data and insights to help in setting policy or funding priorities;
- Identify proven approaches to improve services or areas for experimentation;
- Secure funding to sustain or scale up initiatives; and
- Build evidence-based practices in the workforce community.

2. Broaden State Research and Evaluation Capacity

As noted in Section 1 above, SWAs must coordinate, to the extent feasible, with Federal evaluations conducted under WIOA, including DOL's own evaluations. DOL conducts a wide array of evaluations—from process and outcome studies, pilot and demonstration projects, programmatic and economic data analyses to impact and benefit-cost analyses. Section 169 of WIOA also describes several evaluation and research projects and provides details on the evaluations required for Title I programs, career pathways, and equivalent pay. The legislation also suggests seven other research projects for disconnected youth, business needs, nontraditional occupations, performance indicators, public housing assistance recipients, older workers, and credentials for prior learning that may present opportunities for coordination with an SWA in the future.

SWAs may also want to consider several options that may enhance their ability to inform policy and to conduct their own evaluations. For example:

- *Become familiar with the active and recently completed workforce program evaluations and the resources identified in **Appendix C** of this toolkit.* Utilize current and past Federal research and commonly used evaluation methods to create new opportunities that build upon and increase the body of evidence. The resources in Appendix C are organized under policy, plan, and guidance resources; evaluation resources and toolkits; behavioral insights studies; cost studies, data analytics; implementation studies; interrupted time series; logic and theory of change models; outcome studies; power analysis studies; quasi-experimental studies; and randomized controlled trial studies.
- *Consider an evaluation policy and become familiar with DOL's Evaluation Policy Statement.⁵* Developed by the Chief Evaluation Office (CEO), this policy statement identifies the following principles: rigor, relevance, transparency, independence, and ethics (in human subject protections)—all of which are relevant to state-conducted evaluations. These principles are similar to those in other Federal agencies, as well as common in standards promoted by such organizations as the American Evaluation Association.

States may want to follow similar principles to assure stakeholders that their own evaluations include valid and reliable data, are publicly available, and protect the privacy of individuals who are subjects in the research. The DOL Evaluation Policy also helpfully places evaluation in the context of a learning organization, stating that:

“Evaluations produce one type of evidence. A learning organization with a culture of continual improvement requires many types of evidence, including not only evaluation but also descriptive research studies, performance measurement, financial and cost data, survey statistics, and program administrative data. Although this policy focuses on evaluation, the principles apply to the development and use of other types of evidence as well.”

⁵ <https://www.dol.gov/asp/evaluation/EvaluationPolicy.htm>.

The policy statement also addresses rigor in evaluations, no matter what type of study is undertaken, consistent with WIOA regulations. The DOL evaluation principles described in Exhibit 2.1 use excerpts from the policy statement found on the CEO’s website.

Exhibit 2.1: Department of Labor Evaluation Principles

Principle	Brief Description
Rigor	“Rigor is required for all types of evaluations, including impact and outcome evaluations, implementation and process evaluations, descriptive studies, and formative evaluations. Rigor requires ensuring that inferences about cause and effect are well founded (internal validity); requires clarity about the populations, settings, or circumstances to which results can be generalized (external validity); and requires the use of measures that accurately capture the intended information (measurement reliability and validity).”
Relevance	“Evaluation priorities should take into account legislative requirements and the interests and needs of leadership, specific agencies, and programs; program office staff and leadership; and DOL partners such as states, territories, tribes, and grantees; the populations served; researchers; and other stakeholders.”
Transparency	“DOL will make information about evaluations and findings from evaluations broadly available and accessible, typically on the Internet. DOL will release results of all evaluations that are not specifically focused on internal management, legal, or enforcement procedures or that are not otherwise prohibited from disclosure. Evaluation reports will present all results, including favorable, unfavorable, and null findings. DOL will release evaluation results timely ... and will archive evaluation data for secondary use by interested researchers (e.g., public use files with appropriate data security.”
Independence	“Independence and objectivity are core principles of evaluation. Agency and program leadership, program staff, stakeholders, and others should participate in setting evaluation priorities, identifying evaluation questions, and assessing the implications of findings. However, it is important to insulate evaluation functions from undue influence and from both the appearance and the reality of bias.”
Ethics	“DOL-sponsored evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. Evaluations will comply with both the spirit and the letter of relevant requirements such as regulations governing research involving human subjects.”

2.1 Expand or Enhance Evaluation Capacity

While a handful of states have developed a robust evaluation capacity, most state workforce administrators face challenges in building their state’s capacity to conduct evaluations. The following subsections identify some key elements and capacity-building tips that may be helpful to the user in addressing the challenges based on a synthesis and adaptation of multiple sources.⁶

Leadership and Agency Relationships: Building relationships fosters a state culture that supports evaluation and evidence-based policymaking and ensures that evaluations conducted are relevant to all key stakeholders. To build state relationships, consider the following action items.

⁶ NASWA, February 2017; ETA Listening Sessions, February 2015 and December 2016; CEP, September 2017; Workforce Information Advisory Council, 2017

- Cultivate buy-in—leadership and support—from the Governor’s office, State Workforce Board and agency heads, and State Legislative staff;
- Develop cross-agency relationships among workforce, education, social services, and economic development; identify common research and evaluation goals to allay concerns about funding competition and other “turf” issues;
- Determine and employ strategies to develop and maintain trust regarding information sharing among state agencies and staff;
- Use phased approaches to produce data “wins” that provide the evidence needed for policymakers to make informed decisions and for agency heads to improve programs;
- Produce objective research products upon which policymakers and agency heads can rely;
- Explore legislation to institutionalize cross-state coordination and collaboration concerning data sharing (e.g., establish centralized cross-agency longitudinal administrative data sets and key roles and responsibilities for those engaged in data and research efforts).



State Evaluation Capacity-Building

- Build leadership and agency relationships;
- Leverage existing funding sources;
- Expand staff skills and capacity;
- Establish “big data” organizational structures for data sharing;
- Leverage and integrate federal workforce and related “big data”;
- Develop a state evaluation plan/learning agenda; and
- Track major developments affecting state evaluation capacity-building.

Leverage Existing Funding Sources: To make sound strategic investments, states may want to invest in evaluations. SWAs could consider options to invest, such as funding in-house research or evaluation studies or procuring external partners or research firms to increase the level of independence and objectivity. What follows are funding options or opportunities that may be of interest:

- *Use or continue to use discretionary grants from the DOL Workforce Data Quality Initiative (WDQI) and ED State Longitudinal Data System (SLDS) programs to develop the data infrastructure needed to make research and evaluation possible and efficient.* These state longitudinal databases include information on programs that provide training and employment services and are linked longitudinally at the individual level. WDQI funding is made available through competitive grants administered by DOL in support of a parallel and much larger effort, SLDS grants administered by the U.S. Department of Education. These two programs encourage the development of state education and workforce longitudinal administrative databases. Through analysis, these data will demonstrate the relationship between education and training programs, as well as the additional contribution of the provision of other employment services.
- *Leverage other DOL grants to states for the development of labor market information and analyses to conduct program evaluations.* Pursue grants from Bureau of Labor Statistics (BLS) Federal/State Cooperative Program and Workforce Information Grants to States (WIGS). Despite flat or declining funding levels, leverage resources from these grants for evaluation activities, as appropriate.
- *Use Governor’s statewide set-aside funds to support the conduct of evaluations for Title I core programs, as required by WIOA.* To be responsive to Federal funders, the conduct of evaluation demonstrates efforts toward continuous improvement.

- *Leverage the use of specific program funding for evaluations, where possible; embed evaluation requirements as part of program design and delivery.* As efforts to demonstrate what works increase, the need to employ evidence-based practices also increases. As funding streams allow, develop evaluation requirements into funding opportunity announcements, set expectations for funding recipients, and provide evaluation technical assistance.
- *Collaborate with research universities to conduct evaluations in common interest areas.* Partnerships with universities allow states to harness their intellectual, research, community engagement, and capacity- building expertise to address a multitude of issues. Universities employ skilled evaluators, support graduate student research opportunities, secure “big data” computing facilities, and offer other evaluation services they may be able to provide at less cost than private firms or in-house services; and given mutual interests, they may be willing to provide in-kind resources.
- *Seek foundations with common evaluation interest areas; they may be willing to fund and conduct a particular state study.* The Council on Foundations stresses, “no matter how different foundations can be, they all share the need to know what works, and especially what works well. The better that foundations can demonstrate how their grants are making a difference, the more value they will bring to their communities. To know what works, foundations must [also] evaluate their grants.”

Expand Staff Skills and Capacity: Many states do not have a research and evaluation unit in their workforce agencies or otherwise lack staff capacity to conduct research and evaluation (e.g., skills, especially in experimental research, and staff numbers). To build staff capacity, consider the following action items:

- Work with relevant state entities to allow for pay grades that can support hiring experienced evaluators;
- Partner with universities that have strong social science research centers or capabilities;
- Explore hiring graduates from public policy, public affairs, or data analytics degree programs (programs which appear to be expanding across many universities); and
- Create a data and research staff work environment that is mission driven and promotes innovation to retain talented staff.

Establish “Big Data” Organizational Structures: Creating the capacity to house and use “big data” will allow access to data that can be shared as necessary for evaluation activities. The two case studies that describe centralization of administrative data and access—Ohio and Washington— in the NASWA February 2017 report are considered “high achiever” models. Such organizational structures can help multiple state agencies, external researchers, and others readily use workforce, education, and other data to conduct evaluations. Key actions to create such a data capacity include:

- Develop a cross-agency longitudinal administrative data set covering a range of public programs including wage record data (such as WDQI and SLDS grants as seed funding).
- Establish a neutral, centralized entity to collect data across agencies. A neutral agency can set and enforce requirements regarding data collection, data sharing, data access procedures, and security standards, including such elements as:
- Standardized application and approval processes for data access/data requests;
- Development of metadata, documentation, and codebooks for use of the data in later research planning;
- Confidentiality forms and certification procedures, as well as a system for Institutional Review Board (IRB) notifications;
- Custom data extracts that are de-identified, longitudinal, and linked across data sources (e.g., employment linked to education);
- Extracts of approved data transferred via secure protocols; and

- Data destruction standards and affidavits.
- Collaborate with universities to establish a centralized entity that utilizes workforce information and program data for research. Universities can typically offer a full-service research center focused on education, workforce development, and human services policy and practice.

Leverage and Integrate Federal Workforce and Related “Big Data:” In addition to state-level “big data” integration via WDQI and SLDS, as noted above, there are other federal-level efforts to factor into state capacity building, including the following examples:

- **Workforce Innovation Performance System (WIPS):** The system used for the collection and reporting of required state and local workforce program data under WIOA (such as the common measures and other program administrative data). While all states are fully familiar with WIPS, of special note are the following key elements:
 - All core WIOA programs currently use WIPS with a phased-in plan for the rest of ETA-funded programs such as discretionary grants and national programs, and others such as Veterans’ Employment and Training Service programs;
 - Most Federally funded workforce programs now use WIOA common measures;
 - WIPS will have and/or be able to provide access to individual-level data.
- **Wage Record Interchange System (WRIS and WRIS2) and the State Wage Interchange System (SWIS):** WRIS is a data sharing agreement and process that facilitates the interstate exchange of Unemployment Insurance (UI) wage records data between participating states for the primary purpose of assessing and reporting on WIOA state and local workforce program performance, and for selected other workforce programs not under DOL jurisdiction. States signing the SWIS agreement will be able to exchange interstate quarterly wage records with any other state signing the agreement. SWIS replaces WRIS and WRIS2 agreements previously executed by most states. Although similar to the WRIS and WRIS2 agreements, the SWIS agreement incorporates all six WIOA core programs. The SWIS Agreement also addresses the requirements of WIOA, the confidentiality requirements of the Family Educational Rights and Privacy Act (FERPA) for education records, and the VR program regulations for VR records.
- **Integrated Post-Secondary Education Data System (IPEDS):** A dataset based on inter-related ongoing surveys of degree-granting institutions to report on undergraduate education program participation and outcomes (e.g., enrollment, completions, degrees, and certificates).
- **Federal Statistical Research Data Centers (FSRDCs):** The U.S. Census Bureau, on behalf of the federal statistical agencies, operates a network of FSRDCs in collaboration with research organizations (hosts) in 18 states and DC to provide secure access to a range of federal restricted-use micro-data for statistical purposes.
- **Center for Administrative Records Research and Applications (CARRA):** Through this center, the U.S. Census Bureau uses administrative data from federal, state, and third-party providers for statistical activities in support of demographic and socioeconomic research; as of 2017, 12 pilot research and evaluation projects spanned housing, health, welfare, education, and labor agencies.

Develop a Statewide Strategic Evaluation Plan and Annual Learning Agenda: A statewide strategic evaluation plan and subsequent learning agenda provide the foundation for building research evaluation capacity. These types of planning efforts provide momentum and direction to state evaluation priorities, identify completed research efforts, and identify any current studies underway. A multi-year evaluation plan allows SWAs to consider targeted research priorities, and the annual research and evaluation-learning agenda focuses on the state's key and immediate research questions. A multi-year evaluation plan typically requires input from a range of external stakeholders. On the other hand, the learning agenda identifies current research questions that arise from annual program operations and management concerns. Section 2.2 provides additional processes and details for a learning agenda and an example of the Learning Agenda Life Cycle in **Appendix D** provides some highlights from the process used by the CEO at DOL.

Track Major Developments Affecting State Evaluation Capacity-Building: The Congressional Commission on Evidence-Based Policymaking made many recommendations affecting federal and state data access, integration, and sharing, all of which have implications for state evaluation capacity-building (CEP, September 2017). Bi-partisan legislation to codify an initial set of these recommendations was enacted through the U.S. Congress—*Foundations for Evidence-Based Policymaking Act of 2018*. In addition, the Workforce Information Advisory Council (WIAC) established by WIOA to advise the Secretary of Labor on improving the workforce and labor market information system has recently completed its draft report of recommendations that includes several recommendations affecting state evaluation capacity-building (WIAC, draft, January 2018). State workforce administrators and evaluation managers may find it useful to examine the recommendations and track the results of these efforts to support their capacity building plans.

2.2 Develop an Evaluation Plan or a Research Learning Agenda

When planning to conduct evaluations under WIOA, development of a comprehensive, multi-year strategic research and evaluation plan, along with annual research “learning” agendas to lay out specific evaluation activities, helps to support or address state priorities.

The evaluation section of the state strategic plan can be a starting point for further development of the state’s workforce learning agenda. To further a state program’s research and evaluation plans, the following resources and tools may be helpful:

- **Examples of Learning Agenda to Examine Current Priorities.** ETA uses a Five-Year Research and Evaluation Strategic Plan⁷ that identifies high priority topic areas used for planning research and evaluation and develops an internal learning agenda to prioritize research questions by program. Mississippi and Ohio plans are featured in the NASWA report (NASWA Case Studies, February 2017). Exhibit 2.2.1 describes the seven elements of a state evaluation plan or research learning agenda.
- **High-Level Research Questions / The Big Questions.** One of the keys to building a state evaluation plan/research agenda involves determining high-level research questions—in other words, over the next year or more, what questions does the SWA want to answer? SWAs may benefit from examining the high-level research questions identified through the NASWA scan. Exhibit 2.22 identifies states’ most pressing questions about workforce agency research and evaluations as they relate to program impacts and effectiveness, program performance and outcomes, labor market data, customers and their barriers, as well as program administration and operations.
- **Completed State-Conducted Studies.** The NASWA report also provides a list and hyperlinks to recently conducted state studies that may also be useful to examine (NASWA, Appendix B, February 2017). Another source for identifying state-conducted studies is to review the Workforce System Strategies website, <https://strategies.workforcegps.org/search>, and search for studies tagged by geographic location.
- **How to Develop a Research Learning Agenda.** The Office of Evaluation Sciences at the General Services Administration has developed a series of toolkits to provide guidance to federal agencies on activities related to the implementation of the Foundations for Evidence-Based Policymaking Act of 2018 (“Evidence Act”). The Evidence Act Toolkit: A Guide to Developing Your Agency’s Learning, published in early 2020, although written for federal agencies can be used by states to learn from best practices and processes.
- **Annual Evaluation Planning Support.** The OES at the GSA also published tools for the development of an annual evaluation plan including an [Evaluation Planning Tip Sheet](#) and an annual evaluation plan workbook that could be customized by the state agency.

⁷ The [ETA Five-Year Evaluation Plan](#) describes the types of demonstration and pilot, multi-service, research, and multi-state projects that will focus on the current administration’s employment and training priorities.

⁸ The [Evidence Act Toolkit: A Guide to Describing Your Agency’s Learning Agenda](#) contains information on the benefits of a learning agenda, types of questions in a learning agenda, the people involved in the development process, how to gain leadership input and support, a process for learning agenda development, stakeholder engagement in research question development, writing the learning agenda, and links to additional resources on learning agendas.

⁹ The [Evidence Act Toolkit: A Guide to Developing Your Agency’s Evaluation Plan](#) describes the benefits of an annual evaluation plan, the people involved in the process, a definition of “significant evaluation”, the contents of the plan, and links to additional resources on annual evaluation plan resources.

Exhibit 2.2.1: Building a State Evaluation Plan or Research Learning Agenda

Element	Tips
Stakeholder Involvement	Establish or use an existing process for involving all key state agencies, state workforce board, local boards, and other stakeholders to coordinate the development of the plan—overall evaluation goals, study priorities, funding mechanisms, and roles/responsibilities.
Evaluation Principles and Practices	<ul style="list-style-type: none"> • Develop a set of evaluation principles. (As an example, see the DOL principles in Exhibit 2.1.) • Develop a set of core standards or practices to bring quality and standardization to state evaluations, especially regarding coordinating and collaborating with local workforce boards.
High-Level Research Questions (the “BIG” questions)	<ul style="list-style-type: none"> • Identify high-level research questions that start with what do state workforce and local boards want to know? What research questions did agency program managers identify in the WIOA state plan? (See Exhibit 2.2.2 for the most pressing state research questions as examples.) • Gather input from the state governor’s priorities, legislative requirements, and partner evaluation priorities; and utilize Federal evaluation priorities, external requests from academic or research organizations, and others as appropriate. • Organize key research questions into subject areas.
Building an Evidence Portfolio	<ul style="list-style-type: none"> • Based on the high-level questions, develop an evidence-based agenda in a particular subject area to think through what is known and not known. What evidence exists to build on, such as literature reviews, evidence reviews, and meta-analyses. • Conduct exploratory studies to address the state of the evidence. • Use existing administrative data for descriptive research to understand the relationship between services and outcomes regarding key research questions, such as data analytics projects. • Build evaluation milestones or activities into program designs. • Incorporate the concept of tiered evidence into the building process.
Tiered Evidence Approach	<ul style="list-style-type: none"> • Build the evidence portfolio using the concept of tiered evidence, increasing the state evidence base over time—delineated as a continuum of preliminary evidence on one end of the spectrum and strong evidence on the other. • Add evaluation/evidence requirements into state grant programs, with incremental funding linked to the availability and quality of existing evaluation evidence supporting the proposed program. (See, for example, DOL-ETA use of this concept in its funding of Trade Adjustment Assistance Community College and Career Technical [TAACCCT] and Workforce Innovation Fund [WIF] grants.)
Partnership and Governance Structure	<ul style="list-style-type: none"> • Develop or strengthen an existing state partnership and governance structure involving all key entities responsible for or involved in conducting state workforce research and evaluation, including the state and local workforce board. (See Ohio and Washington case studies for examples.) • Address governance issues such as agency roles and responsibilities, evaluation priorities, data sharing, and legal or cultural issues. • Develop partnership practices or procedures for data collection, data sharing, and handling differences in IT systems and data quality.
Plan, Document, and Disseminate	<ul style="list-style-type: none"> • Develop the plan document(s)—multi-year and annual. • Implement processes for ongoing stakeholder engagement and communication. • Determine if and how to publish and disseminate the evaluation plan/learning agenda.
Adapted from multiple sources: NASWA, February 2017; ETA Listening Sessions February 2015 and December 2016); CEP, September 2017; Corporation for National and Community Service/Social Innovation Fund, 2013 and 2016; DOL/CEO policy guidance from CEO website.	

Exhibit 2.2.2: Examples of High-Level State Research and Evaluation Questions

States' Most Pressing Questions for Workforce Agency Research and Evaluation	
Program Impacts/ Effectiveness	<ul style="list-style-type: none"> • Are the educational programs provided to offenders by the state prison helpful in obtaining employment after release? • What is the effectiveness of Unemployment Insurance (UI) profiling? • What are optimal policies or incentive mechanisms that encourage the greatest return on investment? • What is the relationship of TANF participation to successful workforce outcomes? • What is the effectiveness of the refugee training services?
Program Performance/ Outcomes	<ul style="list-style-type: none"> • Are participants making family-sustaining wages? • What are the workforce outcomes from training programs? • What are the employment and wage outcomes of degree and certificate program completers? • What are the outcomes of Department of Labor and Department of Public Health and Human Services workforce programs (e.g., WIOA, RESEA, TANF)? • What are the wage and employment outcomes of apprentices? Does the increase in wages result in sufficient tax revenue to justify an employer tax credit?
Development of Labor Market Data	<ul style="list-style-type: none"> • How are demographic changes impacting the labor force? • What are the demographics of minimum wage workers? • Where can employers find qualified workers? • What are the new industry clusters (e.g., advanced manufacturing)? • Who are the long-term unemployed?
Customers and Their Barriers	<ul style="list-style-type: none"> • What can be done to encourage higher labor force participation rates in targeted populations? • What tools should we create to evaluate client education and skills gaps? • Why are participants not successful, or why do they drop out? • What are the barriers to changing jobs for those currently employed? • What can be done to improve commuter transportation issues?
Program Administration/ Operations	<ul style="list-style-type: none"> • What is the accuracy and utility value of WIOA performance measures? • How do we address declining research budgets but increasing demand for data and insight? • How can we build on programs that are working? • Are we maximizing services across programs, particularly across multiple agencies? • How can we increase the number of apprenticeships?

Source: NASWA, February 2017, page 18.

3. Consider Evaluation Types and Timeframes

While evaluation involves systematic data collection and analysis to answer crucial research questions from policymakers or program managers, the methods used in any type of research study come from the questions posed. Common types of evaluations include impact studies that seek to determine the effects of an intervention relative to a control or comparison group; outcome studies that use performance data (and may involve statistical methods); cost benefit or cost effectiveness studies; and implementation or process studies.

Similarly, the data used in evaluations also flow from the questions the research seeks to answer; as such, data collection and analyses may be qualitative (e.g., derived from individual or focus group interviews with state or local officials or customers, as well as on-site observations and document reviews) or quantitative (as found in program data, wage records, or survey data, for example). However, note that research methods for programs may change over time. Two exciting and relatively recent developments in evaluation methodology include the use of new forms of data analysis and use of behavioral science.

Insights from behavioral science explore how tailored communication or program services on customer decision-making can lead to effective public services. Using behavioral science insights as a field study is also a new effort by DOL to measure how using such techniques may improve the performance and outcomes of DOL programs (Mathematica Policy Research, April 2017). Making specific changes in the structure of programs using behavioral insights may affect participant enrollment and retention and ultimately improve program outcomes.

This section provides an overview of the major types of program evaluation and describes their purposes, features, and limitations. It essentially serves as a concise “Evaluation 101” overview and provides a foundation for the content addressed in subsequent toolkit sections. Depending on the user’s level of knowledge about evaluation, this section may be useful as a:

- Refresher or review tool;
- Training or technical assistance tool for staff or stakeholders; or as a
- General tool to help fine-tune a learning agenda.

3.1 Four Major Types of Program Evaluation

There are essentially four major types or categories of program evaluation—implementation, outcome, impact, and cost studies—with various subtypes within each. Each type of program evaluation is defined and discussed in this section and is followed by four tables that detail the differences in the subtypes of studies and their associated practical considerations, including cost and levels of rigor.

Implementation Studies: An implementation study documents program operation or compares it against goals, across locations, or over time. It describes and analyzes “what happened and why” in the design, implementation, administration, and operation of programs and is generally used to determine whether a program is being carried out in a manner consistent with its goals, design, or other planned aspects. Implementation analyses can serve as stand-alone studies, especially to document new program

processes not yet studied. Implementation studies, as part of more comprehensive evaluations, may also include outcome, impact, and/or cost studies. Implementation studies provide context for other or subsequent evaluation findings and results and make the findings or results interpretable and useful for the programs, services, or interventions studied. Four broad subtypes—case studies, fidelity studies, performance studies, and site comparison studies—are examples of the implementation analysis described in Exhibit 3.1.1 on page 19.

Outcome Studies: An outcome study compares individual outcomes against goals, across programs or locations, or over time. Outcome studies differ from impact studies in one key area of comparative data analysis. Essentially, outcome studies determine if programs achieve the desired results or assess the effectiveness of programs to produce change. Nevertheless, outcomes are often thought (by program staff, not program evaluators) to indicate measurable change or “impact” when outcomes are compared over time or across comparable programs. Two broad subtypes—longitudinal studies and cross-sectional studies—delineate the purpose and practicality of the multiple types of comparative analyses to consider in Exhibit 3.1.2 on page 20.

Impact Studies: An impact study estimates the difference in individual outcomes attributable to a specific program or policy. Impact studies determine whether programs or policies measure the intended impacts—that is, the program causes the differences of the outcomes that it is designed to influence. If the purpose of an evaluation is to determine whether an occupational training program has the desired impacts on the employment and earnings of the individuals it serves, an impact study is the ideal type of evaluation to choose. The best specific type of impact study to carry out depends on considerations such as the budget for the evaluation, the desired level of confidence in the evaluation results, and the practical constraints on conducting an evaluation of a given program. It is important to note that experimental studies (randomized control trials or RCTs) are considered the most rigorous form of evaluation and are often called the gold standard, given that they provide the best scientific evidence of what works or does not. However, they are also the most intrusive type of impact study in that they intervene with program processes. Various types of implementation studies are usually part of impact studies, such as site comparisons and fidelity studies, and of course, outcomes are measured. Four broad subtypes—experimental (RCT) studies, quasi-experimental (QE) studies, rapid-cycle studies, and theory-based studies—delineate the purpose, describe the uses, and explain the practicality of the various types of impact studies to consider in Exhibit 3.1.3, starting on page 21.

Cost Studies: A cost study estimates program costs, makes cost comparisons, or weighs costs against outcomes or impacts. Cost studies involve analysis of the costs of a program, and some weigh program effectiveness against overall program cost. Sometimes cost documentation, estimation exercises, or simple cost calculations are considered a cost analysis, but not by program evaluators. While they are common elements of all cost studies, they are not considered cost analyses. A cost study draws conclusions about program costs based on systematic cost comparisons (particularly between programs and over time) or statistical analysis of cost differences or responses to changes in program features or inputs. The specific comparisons and statistical analyses depend on the program and the quality and detail of available cost data. Three broad subtypes—cost analysis studies, cost-effectiveness studies, and benefit-cost studies—are described in Exhibit 3.1.4 on page 23.

Exhibit 3.1.1: Types of Program Evaluation—Implementation Studies

Purpose: An implementation study documents program operation or compares it against goals, across locations, or over time. It describes and analyzes “what happened and why” in the design, implementation, administration, and operation of programs. Data Requirements: A defined set of qualitative and quantitative program and participant data related to the topics covered in the study.		
Type & Description	Practical Considerations	Cost & Rigor Considerations
CASE STUDY A study that provides a detailed examination of program operation and/or aggregate program outcomes at a single location. It may include fidelity or maturity assessments.	<ul style="list-style-type: none"> • Single local program case studies are generally done because a program is unusual (innovative and high-performing operation or a troubled, poor performing program). • Multiple case studies are generally done when an objective is more broadly about program operations, including identifying best practices. • Varied analytical components are sometimes purely descriptive; more often systematic comparisons which may include fidelity or maturity assessments. • Case study findings can be used to inform future program development or replication by describing program features that worked well and those that did not. 	<ul style="list-style-type: none"> • Relatively inexpensive, although costs depend on methods used, the number of sites or programs, and availability of pertinent data. • Not rigorous from the standpoint of impact (assessment is generally pre-post and employs a small sample). • Can provide a valuable first step before moving on to more rigorous evaluations.
FIDELITY STUDY A study that compares program elements and operation measures to the program model or objectives and involves determining if programs are implemented as designed or legislated.	<ul style="list-style-type: none"> • As a stand-alone study, it is used to assess whether program operations are consistent with legislative intent, regulations, agency objectives, and/or a program model. • As a common component of comprehensive evaluations including impact studies at multiple sites, it allows comparison of impacts at sites with high fidelity (i.e., operating a program consistently with intentions across all key elements) to those with lower fidelity. 	<ul style="list-style-type: none"> • Inexpensive and highly useful in many contexts. • Does not assess impacts, so not judged to be rigorous or not in the usual sense of the term. • An essential component of many impact evaluations.
PERFORMANCE STUDY A study that assesses program performance over time and involves assessing the progress made in scaling up and achieving program fidelity and smooth operations.	<ul style="list-style-type: none"> • Generally used for new programs focused on a single program in one place or a few locations. • Assessments may address fidelity and operational issues during an early phase or scaling up processes (since new programs often operate on a very small scale at first and begin to operate with greater fidelity and fewer problems at a larger scale as they mature). 	<ul style="list-style-type: none"> • Low to moderate study costs. • Low to relatively high study quality.
SITE COMPARISON STUDY A study that compares program operations across locations.	<ul style="list-style-type: none"> • Generally used when a study involves multiple related programs or a single program with many locations. • Comparisons are usually between program operation metrics and/or aggregate outcomes which may be considered performance measures. 	

Exhibit 3.1.2: Types of Program Evaluation—Outcome Studies

Purpose: An outcome study compares individual outcomes against goals, across programs or locations, or over time. Data Requirements: All outcome studies require individual-level outcome data on the population or sample of interest. Longitudinal outcome studies need baseline data on characteristics of sample members and pertinent outcomes before and after program enrollment.		
Type & Description	Practical Considerations	Cost & Rigor Considerations
LONGITUDINAL STUDY A process study examines outcomes over time for specific cohort(s) of individuals when focused on trends in program participation. A pre-post outcome study, a less rigorous form of impact analysis, assesses outcomes before and after program participation.	<ul style="list-style-type: none"> • Process analysis addresses program participation by following a cohort (or cohorts) of applicants, accepted applicants, or enrollees through a program’s chain of outcomes—that is, application, enrollment, assessment and/or orientation, program activities, and completion or graduation; it typically produces metrics such as enrollment or participation rates, the proportions of enrollees who participate in certain kinds of activities, and completion rates. • Pre-post analysis assesses outcomes before and after program participation; common outcomes measured in workforce pre-post studies include employment rates, average wage rates, and measures of education and training achievement such as credentials obtained; tracking progress of individuals before, during, and after program participation is the best way to understand interim or short-term outcomes, allowing for the measurement of changes over time and description of participants before and after the program. • Pre-post outcome studies may be a first step in developing an evidence base for a new intervention or laying the groundwork for a later, more rigorous impact evaluation. 	<ul style="list-style-type: none"> • Generally, longitudinal process studies are superior to, and more expensive than, studies providing point-in-time outcome estimates. • Pre-program longitudinal data can raise cost concerns, but program participation and immediate outcome data are generally collected as part of workforce and education program performance tracking. • Pre-post studies are a methodologically weak form of impact study, but a practical way to look at potential impacts; these studies are not as rigorous as an RCT or QE study, but less costly and shorter in time.
CROSS-SECTIONAL STUDY A study that analyzes the characteristics and outcomes for a population at a specific, single point in time and may involve comparisons to goals or across programs, population subsets, or locations.	<ul style="list-style-type: none"> • A common study of this kind is an analysis of all individuals served in a state training during a given year where the analysis may involve any number of comparisons of outcomes (often “single point in time” comparisons) between groups or locations. • Another common study of this kind is a comparative analysis of outcomes to program goals to determine whether local programs have met performance objectives, such as program graduates holding jobs at least three months. 	<ul style="list-style-type: none"> • Inexpensive to conduct given that data are readily available through routinely collected program performance data. • These studies provide insights but not rigorous impact estimates.

Exhibit 3.1.3: Types of Program Evaluation—Impact Studies

<p>Purpose: An impact study estimates the difference in individual outcomes attributable to a specific program or policy.</p> <p>Data Requirements: An impact study needs baseline data on the characteristics of sample members and on pertinent outcomes before random assignment or program enrollment, and follow-up data on pertinent outcomes after random assignment or program enrollment.</p>		
Type & Description	Practical Considerations	Cost & Rigor Considerations
<p>EXPERIMENTAL STUDY (RCT)</p> <p>A study that uses an experimental research design, randomly assigning individuals to treatment and control groups. Estimated impacts are treatment-control differences in outcomes. An experimental study in the field of program evaluation is usually called a randomized control trial (RCT).</p>	<ul style="list-style-type: none"> • RCTs are the clearest and most rigorous evidence of program effectiveness; they permit confidence that the program made a significant difference and caused the change (i.e., scientific proof). • In the RCT assignment process, assignment to the treatment and control groups is random; therefore, there are no systematic differences between the two groups, except for access to the program, and as a result, all measured differences in outcomes that occur after random assignment can be confidently attributed to the program. • RCTs deny program access to a control group, making their use controversial or infeasible, especially if adequate resources exist to serve an entire eligible population. • RCTs cannot be implemented retrospectively unless, prior to the evaluation, a lottery-like process has been implemented to allocate access to the program due to budget constraints; once random assignment has started, administrators must wait, possibly for many years, for impact results to become available. • Must weigh the challenges of conducting an RCT against the value of the clear, rigorous evidence of program effectiveness that can be produced. • RCT results: provide “proof of program effectiveness”; contribute to the evidence base; are valuable to workforce stakeholders and scholars in determining whether expected impacts were realized; and support development of approaches that build on the evidence to refine and expand programs. 	<ul style="list-style-type: none"> • Most rigorous form of impact evaluation—gold standard. • Expensive, although costs can be contained through use of records data. • Usually take multiple years to conduct (behavior insight studies are an exception). • Usually more challenging and expensive than other impact study options.
<p>QUASI-EXPERIMENTAL STUDY (QE)</p> <p>A study that seeks to replicate an experimental research design using comparison groups, sites, or conditions. The goal for estimating a set of “counterfactual” outcomes to the outcomes of the evaluated program’s eligible applicants or participants are compared.</p>	<ul style="list-style-type: none"> • QE methods tend to be used for impact studies of relatively small and new programs; then, particularly if there are favorable impact results and the program becomes a larger enterprise, an RCT may be conducted. • In general, maximizing QE study credibility involves: (1) identifying a comparison group that closely matches the program’s applicants/participants; (2) securing extensive data on pre-program outcomes for both the program and comparison groups; (3) using statistical matching techniques to select comparison group members; and (4) using statistical estimation methods for impact estimates that control for individual traits and pre-program differences. 	<ul style="list-style-type: none"> • QE methods such as CITS and RD are almost as rigorous as RCTs. • QE methods—attractive alternatives to RCTs—take less time and are less obtrusive and expensive. • CITS and RD are only feasible under certain conditions.

Purpose: An impact study estimates the difference in individual outcomes attributable to a specific program or policy.

Data Requirements: An impact study needs baseline data on the characteristics of sample members and on pertinent outcomes before random assignment or program enrollment, and follow-up data on pertinent outcomes after random assignment or program enrollment.

Type & Description	Practical Considerations	Cost & Rigor Considerations
<i>[QE continued]</i>	<ul style="list-style-type: none"> • Unlike RCTs, QE studies can often be implemented retrospectively, be carried out without denying services to a control group, take less time, and be easier and less expensive to conduct. • Comparative Interrupted Time Series (CITS), one of the most rigorous QE approaches, is a viable option when a new program or substantial policy change is implemented suddenly for one location, but not for other locations with essentially the same population and economic conditions. • Regression Discontinuity (RD), another rigorous QE design, can only be used in cases where eligibility for a program is determined by a count or score (such as hardship index or an education test score). • Other QE options—usually using comparison groups or sites—are less rigorous, but can be applied under a much wider range of evaluation circumstances. 	
RAPID-CYCLE STUDY A system for continuously conducting impact and outcome studies using research methods similar to those of a comparative interrupted time series (CITS), a rigorous form of quasi-experimental (QE) analysis.	<ul style="list-style-type: none"> • Rapid-cycle approaches involve creating a system for conducting numerous high-quality impact and outcome studies, where individual studies can be continuous (i.e., open-ended); once a system is in place (e.g., software, data files, routine data updating and statistical procedures, other elements), studies can be completed at relatively low cost. • Rapid cycle methodology is similar to CITS in that studies detect significant changes in the trajectories of strings of outcomes after the introduction of a new program or policy; it is an easy-to-repeat research procedure once the system is in place. • Rapid-cycle studies generate growing interest and use in recent years; potential for wide applicability, including evaluation of organization and system changes, new programs, and altered policies. 	Requires time and resources to put a system in place; once in place, high-quality impact and outcome studies can be carried out continuously at a relatively low cost.
THEORY-BASED STUDY A theory of change study involves mapping out the causal chain from inputs to outputs, outcomes and impacts, and testing the underlying assumptions linking these elements.	<ul style="list-style-type: none"> • Theory-based studies generally extend the use of a logic model. • These types of studies do not produce impact estimates unless supplemented by an impact or outcome study but can make a powerful case that impacts are taking place, including how and why. 	<ul style="list-style-type: none"> • Low to moderately high cost, the latter resulting from the inclusion of supplementary outcome and impact components. • Persuasiveness of these studies depends on pertinent research (leading to the theory), data quality, and supplementary components.

Exhibit 3.1.4: Types of Program Evaluation—Cost Studies

<p>Purpose: A cost study estimates program costs, makes cost comparisons, or weighs costs against outcomes or impacts.</p> <p>Data Requirements: All cost analyses require program cost and outcome information. Cost-effectiveness and benefit-cost studies also require results from impact or outcome studies and data from published sources.</p>		
Type & Description	Practical Considerations	Cost & Rigor Considerations
<p>COST-ANALYSIS STUDY</p> <p>A study that makes cost comparisons (between programs, over time) or estimates cost changes resulting from changes in program features or inputs.</p>	<ul style="list-style-type: none"> • Cost analysis studies are dependent on what specific types of comparisons and statistical analyses can be done, which are dependent on the program, and quality and detail of available cost data. • Cost analysis studies range from very modest cost-estimation exercises to sophisticated research projects involved in or separate from programevaluations. 	<ul style="list-style-type: none"> • Least expensive type of cost study unless it involves specialized, technically demanding analysis of cost issues.
<p>COST-EFFECTIVENESS STUDY</p> <p>A study that compares the relative costs of similar programs in achieving a single outcome or impact.</p>	<ul style="list-style-type: none"> • Cost-effectiveness comparisons involve ratios; the numerator is program cost and the denominator is an outcome or impact of program or service provided. • Common ratio calculation; the numerator is gross cost of program in an accounting period (typically a year) and the denominator is total number of outcome units achieved in same year. • Limitations: (1) an outcome is not an impact, so the lowest cost program producing a specific outcome may not be the program providing the “most bang for the buck”; can be overcome if net costs are used in ratio numerator and impacts in denominator; (2) only one outcome or impact can be used in a single ratio, and programs often have more than one; can be overcome by comparing multiple ratios, by constructing an outcome with two dimensions such as trainees who receive a certificate and obtain a job in the field for which they were trained, or by using benefit-cost analysis; (3) program performance can only be judged in relation to other programs; can be overcome by using benefit-cost approach. 	<ul style="list-style-type: none"> • Low to moderate cost. • Study quality depends on quality of outcome data (see outcomes studies above) and the appropriateness of cost-effectiveness comparisons.
<p>BENEFIT-COST STUDY</p> <p>A study that assesses the economic value of a program by assigning dollar values to various outcomes and impacts and comparing the aggregate value of these items to program costs, including return-on-investment or pay-for-performance model assessments.</p>	<ul style="list-style-type: none"> • Benefit-cost analysis is designed to address multiple outcomes and impacts simultaneously; for most workforce evaluations, it can only be done credibly using impacts measured through an RCT or QE design. • Benefit-cost studies monetize tangible program effects and allow for total benefits and costs of a program to be directly weighed against others in a manner analogous to comparing private sector revenues and costs; the value of program as an investment is judged by its net present value (analogous to private sector profit), internal rate of return (analogous to private sector rate of return), and benefit/cost ratio. • Assigning dollar values; many workforce program outcomes/impacts such as participant earnings already denominated in dollars; dollar values must be assigned to others using well defined methods; program costs and benefits occurring at different times must be converted to base-year dollars. 	<ul style="list-style-type: none"> • Most complete and rigorous type of cost study. • Dependent on credible impact estimates, which usually requires a rigorous impact study. • Low to moderate cost so long as there is good program cost accounting and credible impacts are available.

3.2 Evaluation Combination and Continuum of Study

Evaluations often combine several methods. For example, an evaluation of a new program might involve an implementation study combined with an outcome study. The implementation study might include cost estimates or a fidelity assessment as well. Similarly, an outcome study might collect and analyze performance in the near term, for example a snapshot or a comparison to outcomes of more established programs. Such a combined study is useful to demonstrate the potential of a particular program and provide credible information to improve or re-design it. Agencies with the capacity for rapid-cycle evaluation could likely conduct the outcome assessment component routinely at relatively modest cost. In contrast, a program previously studied or extensively piloted may be ready for definitive testing. Such an evaluation would include a rigorous impact study of long- and short-term outcomes, such as an RCT, a QE, or a rapid-cycle study covering a longer period. The evaluation would likely include other components, such as a comprehensive cost analysis, particularly if the program model is costly; a fidelity assessment, if the model is demanding or imperative; and an implementation analysis, if multiple models exist.

It is useful to think about evaluation as a continuum of study with low to high levels of confidence related to program outcomes and their context. At one end of the continuum are implementation studies with limited objectives—often exploratory or motivated by narrow policy concerns. At the other end of the continuum are rigorous impact studies with complementary and confirmatory evaluation components to identify impacts and further make broader policy judgments. In between the low and high levels of confidence, studies of varying methods and rigor may provide the evidence to support program decisions, strategies, and practices.

3.3 Create an Evaluation Timeline

A timeline for an evaluation plan creates a timeframe for critical evaluation activities, such as implementation of the evaluation design, procedures for data collection, and steps for report development and dissemination. SWA administrators and evaluation managers will want to set timeframes to give the selected evaluator opportunities to work together to refine the evaluation design, develop research protocols and data analysis plans, and identify additional evaluation requirements, tasks, and deliverables provided in a detailed timeline.

At a minimum, preliminary planning activities will provide a timeline that demonstrates a clear overview of the:

- Type of evaluation to implement;
- Methodologies the study plans to employ;
- Tasks needed to conduct the evaluation;
- Project management activities,
- Deliverables produced,
- Data resources and agreements needed to collect and analyze data; and
- Interim and final report completion dates.

This section provides broad timing estimates for common tasks to consider in building a high-level evaluation plan timeline.

While every evaluation is unique, the table in Exhibit 3.3 provides examples of common evaluation tasks and associated activities with each task. The table also identifies calendar timeframe estimates for each of the common task examples. The type of evaluation that the agency plans to conduct drives the tasks to include in a statement of work. Some of these tasks may or may not be appropriate for all research studies. The agency may also identify other research or evaluation tasks and activities to include in the study requirements, or the selected evaluator may expand the activities needed for the identified study tasks in the evaluation design. The estimated timeframes in this table are general and just a starting point for the agency's consideration.

Note, the sample evaluation activities and timing in the Exhibit 3.3 table are organized into three distinct sections:

- Evaluation planning activities;
- Evaluation start-up and implementation; and
- Reports and analysis.

Each task identified in the table also includes the associated section or sections for planning and implementing evaluations using this toolkit. Tasks in the table appear as they would typically occur in evaluation implementation activities, although many tasks maybe performed simultaneously.

Exhibit 3.3: Sample Evaluation Activities and Timing

Task	Activities Associated With this Task	Who Is Responsible	Estimated Calendar Time
Evaluation Planning Activities			
Logic Model Development (Section 4)	<ul style="list-style-type: none"> • Develop the logic model for the subject of the evaluation. • To evaluate a specific program or intervention in a logic model, consider the following details: <ul style="list-style-type: none"> ○ Program flow ○ Service delivery ○ Roles and responsibilities of partners ○ Data sharing activities 	<p>State sponsoring agency</p> <p>State agency and program operator for specific program evaluation</p>	1–6 months
Evaluation Planning (Sections 3, 4, 5, and 6)	<ul style="list-style-type: none"> • Conduct meetings/discussions: <ul style="list-style-type: none"> ○ Include relevant state officials and stakeholders ○ Identify overall evaluation goals and parameters • Develop preliminary evaluation plan • Develop timeline and budget • Determine evaluator selection approach 	State sponsoring agency	2–6 months
Evaluation Design Report, Including Data Analysis Plan (Section 4)	<ul style="list-style-type: none"> • Develop evaluation design report (EDR), including data analysis plan, for review/approval by state sponsoring agency. • Be sure to include data security and human subject protection protocols, obtaining Institutional Review Board (IRB) approval, and approval for surveys or other data collection needs, if needed 	Evaluator	<p>1–3 months</p> <p>1–6 months</p>
Evaluator Selection Process (Section 5)	<p>Selection of an evaluator may vary considerably; consider most efficient or effective approaches for:</p> <ul style="list-style-type: none"> • In-house unit; university or other partnership; third-party evaluator via RFP process • To select: <ul style="list-style-type: none"> ○ Third-party evaluators—write and publicize the RFP; review applications; select winning bidder; finalize/sign contract ○ Partnership configurations—identify partner, hold partnership development meetings, draft, negotiate and sign agreement • In-house evaluator—identify state unit and identify staff (FTEs) used 	State sponsoring agency	1–6 months
Evaluation Start-Up & Implementation			
Develop Data Collection Protocols and Instruments (Section 4)	<p>Varies depending on type of evaluation, research questions, and methods; examples include:</p> <ul style="list-style-type: none"> • Interview and/or focus group guides • Site visit guides • Survey instruments • Protocols, data sharing agreements, participant enrollment forms, random assignment protocols and related documents 	Evaluator	1–3 months

Task	Activities Associated With this Task	Who Is Responsible	Estimated Calendar Time
Obtain IRB Approval, if needed (Section 3)	Depends on the type of study conducted, determine if IRB approval becomes a factor for the timeline. Consider the following process activities: <ul style="list-style-type: none"> • Develop and submit informed consent, human study, and other privacy-related materials. • Estimate time allowed for review and approval 	Evaluator with State Sponsoring Agency	1 to several months
Data Collection (Section 4)	Consider the analysis needed, and: <ul style="list-style-type: none"> • Obtain existing program/administrative data (e.g., WIOA, wage records, post-secondary education records) • Obtain existing specific administrative data from program data systems • Conduct telephone interviews, site visits, focus groups, and other qualitative research; synthesize information/write internal reports for each interview, focus group, and site visit • Conduct surveys; compile and synthesize response results • Collect program participant data for pre-post outcome/RCT studies 	Evaluator	6 or more months, for multi-year studies, multiple rounds of data collection
Progress Reports (Section 4)	Evaluator submissions scheduled as required by state sponsoring agency (e.g., monthly or quarterly)	Evaluator	1 day per period
Interim Reports (Section 4)	<ul style="list-style-type: none"> • Check data for quality and comprehensiveness • Analyze data collected, to date • Write and revise (as appropriate) report based on state sponsoring agency feedback on draft 	Evaluator State sponsoring agency review and feedback of draft	3–4 months
Final Report (Section 4)	<ul style="list-style-type: none"> • Check data for quality and comprehensiveness • Analyze all data collected during evaluation • Write and revise (as appropriate) report based on state sponsoring agency feedback on draft • Publish report as required by WIOA 	Evaluator State sponsoring agency review and feedback of draft; publish report	3–4 months
Public Use Data Set (Section 4)	If “original data are collected,” a public use data set is standard protocol. Check data for quality and comprehensiveness; remove all personally identifiable information (e.g., names, social security numbers, and addresses); write comprehensive codebook (i.e., a guide describing each of the variables in the data set)	Evaluator State sponsoring agency publish data set	2–6 weeks

4. Develop an Evaluation Design Plan

This section outlines the key elements for creating a preliminary evaluation plan—the critical first step in designing and conducting an evaluation, regardless of the methods used. The preliminary plan defines the requirements for both developing and implementing a well-designed evaluation. Well-designed evaluations include clearly identified and articulated key research questions. Clearly articulated research questions, in turn, help to determine the study methods and data collection and analysis used to answer them.

Depending on who conducts the evaluation (see Section 5), the preliminary plan, also called a “statement of work,” leads the evaluation planner either to:

- Prepare a Request for Proposals (RFP) to procure an independent, third-party evaluator;
- Provide guidance to an in-house evaluator; or
- Develop a partnership agreement with a university or other organization.

Once selected or procured, the evaluation planner and the evaluator jointly administer and implement the requirements to conduct the evaluation.



Preliminary Evaluation Plan Elements

- A logic model to describe the program in the evaluation;
- The purpose and scope of the evaluation; The key research questions to guide the evaluation;
- Any existing evidence base (literature review) to support the study; and
- A research design and data collection plan.

4.1 Develop the Logic Model

It is a good idea to begin the preliminary evaluation plan with a logic model. A detailed logic model may not be necessary, particularly if the focus of the study is narrow rather than comprehensive. Even if the study focuses on a program activity or strategy, a simple logic model delineates program inputs, outputs, activities, and outcomes.

Importance of a Logic Model: A logic model relies on a specific “theory of change.” Examining the logic underlying a program (i.e., program, system, strategy, service, activity, or intervention) clarifies the subject matter and context of the evaluation. For example, within a logic model, program implementation describes the inputs and outputs, and the program results are expressed as outcomes and impacts. The logic model description is a detailed account of the program’s content and organizational structure, size, flow, staff support, the amount of staff training required to implement it, and the services provided or system change activities undertaken. It may also contain a clear depiction of the relationships between program elements and the intermediate- and long-term outcomes those elements are expected to affect. A well-designed logic model serves as a blueprint for the implementation and evaluation of the program it describes.

Components of a Logic Model: A complete **logic model** addresses the key components of a program, including inputs and activities, outputs, and both intermediate- and long-term outcomes. If more than one program, method of service delivery, or process exists, then it may also be appropriate to develop two or more logic models. For example, to evaluate a change management intervention along with an existing training program, consider how those two aspects of the program operate and depict the constructs in two different logic models. Inputs, outputs, activities, and outcomes are common to all logic models; however, these components also describe how a program operates or set expectations in the theory of change.

Key components in logic models are **inputs** and **activities** used to operate a program. The logic model components are the ingredients in the program implemented to achieve the desired outcomes. The key components—essential activities and inputs—may include, for example, financial resources, professional development for trainers, curricular materials, or technology products. The inputs and activities describe the program in action and summarize the required operations to attain fidelity to the model. Regardless if the logic model is linear or non-linear, it identifies the basic operational standards and structure in graphic form organized to demonstrate how the outputs lead to the program’s intermediate- and longer-term outcomes.



Benefits of a Logic Model

- Maps program management operations—identifies inputs, activities, outputs, and outcomes used to measure successful or achieved objectives; and
- Serves as a blueprint for the implementation and evaluation of the program, strategy, service, or intervention studied.

Outputs are products developed, deliverables completed, or milestones accomplished from the program activities and inputs. In other words, program outputs occur when the inputs and activities accomplish the intended objectives. Some examples of workforce program outputs include:

- Participant services (e.g., skill assessments, occupational training) delivered;
- Training/professional development activities or other supports for trainers (e.g., group training, on-site coaching, distance training, curriculum materials) completed;
- Instruction or skill development (e.g., technology, formative assessments, use of instructional time, participant groupings) completed;
- Participant, employer, and community engagement activities completed; or
- Coaching, advising, or referrals developed and delivered.

Intermediate outcomes are the expected program service delivery milestones or goals achieved that can lead to long-term outcomes. A logic model includes all intermediate outcomes through which the program is expected to affect participant long-term outcomes. Note that outputs and intermediate outcomes of workforce programs are often the same. Some examples include:

- Training program completions;
- Participant credentials;
- Participant job placements;
- Number of overlapping services reduced; and
- Number of complete records entered into a new management information system (MIS).

Long-term outcomes include the expected changes in behavior, attitudes, aptitude/skill, and knowledge for staff, participants, environments, or larger systems. Most importantly, workforce programs include changes in employment and earnings, employment retention, and the receipt of credentials as part of long-term outcomes. 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. Exhibit 4.1.1 provides an example of a graphic representation of a logic model based on a simple intervention that offers a training program for a specific type of machine operator. The inputs listed in the left-hand column 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 conducted in the program. 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 simultaneously with 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, higher earnings, and obtaining jobs with benefits.

This logic model example also includes a space for assumptions made, as well as any external factors that may bear on the intermediate- and long-term outcomes. The assumptions and external factors in a logic model provide context for the program's evaluation. Assumptions, for the program in this example, include the ability to secure employment partners committed to offering internships and entry-level positions for students. An external factor that may affect the evaluation is the availability of appropriate open positions for the students who complete the program. Such external factors often rely upon existing knowledge of employment trends in the industries that hire machine operators.

Lastly, this example illustrates the inputs linked to activities, identifies the activities linked to specific outputs, and shows the outputs linked to specific outcomes. Rather than just a long list of each program component without logical paths, the added arrows show which inputs and activities have an expected effect on outputs and outcomes. A clear understanding of the expected effects creates a more defined evaluation plan and, when needed, provides an opportunity to refine the study and articulates the expected results based upon the reality of program operations.

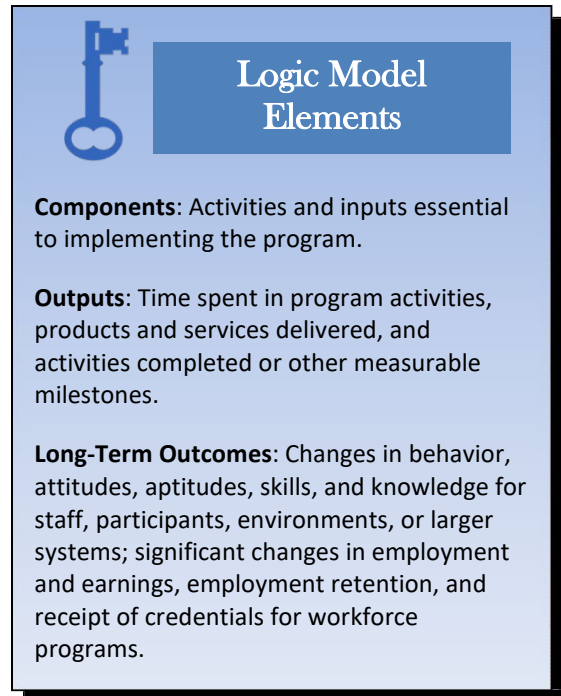
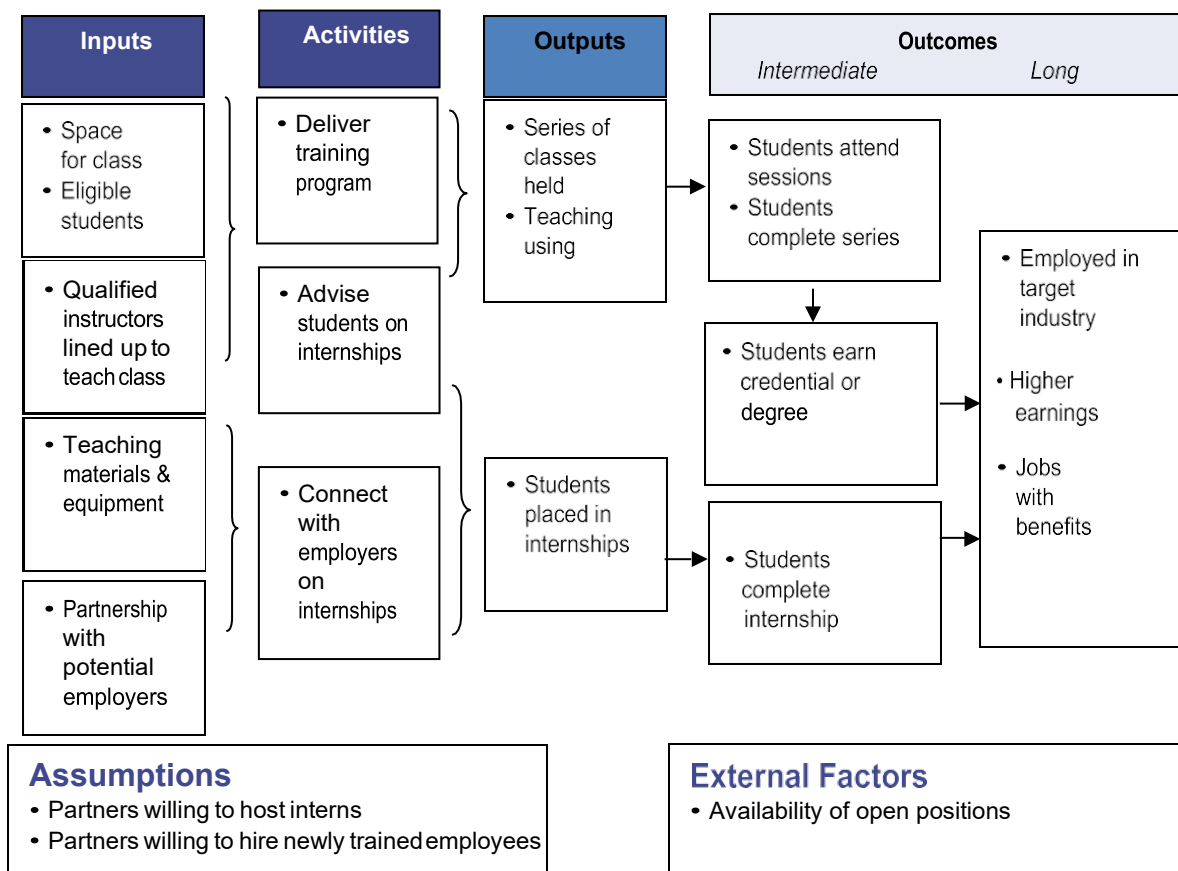


Exhibit 4.1.1: Graphic Logic Model Example—Machine Operator Training Program



Logic Model Creation and Refinement: The blank logic model template found in **Appendix E** is a template to create a graphically displayed logic model for an evaluation.

While SWAs may want to develop both a graphical logic model and narrative description, the graphical template will enable them to identify the essential elements. Once they develop an initial logic model, they may want to refine and assess its comprehensiveness or modify it to meet program evaluation needs. The self-assessment questions in Exhibit 4.1.2 explain steps to further refine or modify components or elements of the logic model to address the gaps identified in the responses to these questions.

Exhibit 4.1.2: Self-Assessment Logic Model Questions

Self-Assessment of Program Logic Models: Discussion Questions

1. Does the model include critical inputs required for the implementation of the service activities? (e.g., accessible technology and other resources, program partner services, staffing, coaches, case managers, recruiting and training trainers, and partners with training services, credentialing, and work-based learning opportunities)
2. Are there system-building activities that are part of the necessary foundation for the program? Are these system-building activities part of the logic model?
3. Does the model include all current activities provided to participants? Is there an existing or expected sequence of participant activities that follows a logical path or pattern?
4. Does the model include all “first-level outputs” of the program? (e.g., measurable milestones that are 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 hypothesized immediate changes and/or outcomes expected for participants, across all relevant domains?
6. Are these immediate changes and/or outcomes an assumed result of specific services?
7. Does the underlying theory of the program design identify expected participant outcomes for particular services?
8. Does the logic model suggest links between intermediate- and longer-term outcomes?
9. Are the longer-term participant outcomes likely to be measurable in the life of the evaluation?
10. As a complete visual or narrative text, does the logic model tell a clear and complete story about the unique program, service strategy, or intervention in the study?
11. If the logic model assumes a theory of change, how does the hypothesis lead to moderate and long-term outcomes?
12. If using a visual representation, does supplementary narrative text provide a clear and complete story?
13. Are there assumptions about external conditions or other external factors that could affect the successful implementation of the program?
14. Are these identified external conditions or other external factors shown in the model?

Questions taken and modified from the WIF Toolkit developed by Abt Associates.

4.2 Determine the Evaluation Purpose and Scope

A critical element of all preliminary evaluation plans is a clearly articulated **purpose and scope**. Stakeholders, partners, and, where informed consent is used, customers need to understand **what is being researched or studied and why**. Identifying an initial set of research questions leads to the purpose and scope of the evaluation. Some general research questions include:

- What are the program processes used for the services provided?
- Does the menu of services provided meet the intended program objectives or goals (for example, employment and increased wages for customers)?
- Do particular services result in better outcomes than other services (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)?
- Is a statewide career pathways system helping or hindering institutions in making a difference for students and workers?

A clearly articulated evaluation purpose and scope helps with stakeholder support as the evaluation activities begin to take place. Stakeholders, such as government executives, legislators, and workforce boards, in turn, can share the evaluation’s purpose and its potential benefits with other involved or interested community members or agency partners. Early stakeholder engagement in the process encourages and adds different perspectives that can fine-tune the purpose of the evaluation. In addition to further fine-tuning the evaluation purpose, use of the program’s logic model as a reference ensures that all aspects of the program and intended outcomes or impacts are considered.

In addition to the purpose of the evaluation, the **scope of the evaluation—which program component(s) to evaluate and how**—is called a statement of work and sets parameters for the study. The evaluation scope allows evaluation planners to identify preliminary quantitative and qualitative data collection methods. For example, existing administrative data sets that document certain program components can support quantitative analyses. On the other hand, other program or service delivery strategies can support qualitative data collection methods, such as interviews, document reviews, and focus groups.

For example, the effects of a program’s occupational training component on participant employment, rigorously assessed in an impact study, rely on data gathered to measure the amount and types of services provided, characteristics of participants receiving the services, and follow-up data on employment through established surveys or administrative records. On the other hand, the effects of a program’s soft skills component may rely on survey data because the intermediate outcomes (such as self-awareness or interpersonal relationships) and longer-term outcomes (such as community involvement or return on investment) are measured through long-term and intensive follow-up surveys. Nevertheless, less rigorous evidence from focus groups with leadership, class teachers, and participants may also add exploratory value to the evaluation.

The scope of an evaluation study also involves other specific considerations that are dependent on the type of evaluation the SWA plans to conduct.

- **Impact and Outcome Studies**
 - **Sites or areas:** determine specific program sites and/or geographic areas to include/use in the study;

Target population: the population served in a program, such as adults or youth; evaluations in which the individuals actually served are a subset of the target population and the sample studied also is a subset of the target population;

Outcomes addressed: the outcomes measured to demonstrate changes or successes;

Observation period: the observation (or follow-up) period extends from the beginning to the end of longitudinal data collection on sample members; the length of this period is usually measured in years, quarters, or months of follow-up available on all cohorts in the sample.

- **Implementation and Cost Studies**

Sites or areas: determine specific program sites and/or geographic areas to include/use in the study;

Elements addressed: the program elements examined to determine fidelity to the model;

Timing and length of study: the effects of program implementation or costs assessed during program planning, development, early operation, or maturity.


While some of the following considerations listed are included as part of studies above, the decision on the type(s) of studies to conduct are dependent on the key research questions identified for the evaluation plan.

4.3 Develop the Key Research Questions

The **key research questions that will guide** an evaluation plan require input from stakeholders. Research questions identify distinct workforce system or program areas to assess in a systemic and credible way. Key research questions share the following characteristics:

- **Specific and Measurable.** The questions identify the specific elements or outcomes to examine and learn about those elements. For example, a specific research question may ask: “Are participants who complete the program in its entirety more likely to be placed in full-time unsubsidized jobs than those who do not complete the program within three months after program exit?” The trends in employment data available may support an outcome study using participants’ post-program data from employers to answer this question. However, UI records available may support an impact study using the unemployment data for both participants and a comparison group at certain points in time. Thus, the research question introduces other possible data collection opportunities.
- **Answerable.** Research questions must be answerable. Some research questions may not be answerable because data may not exist to address the question or the outcome of interest may need further definition. For example, workforce program managers may have an interest in impacts of services on participant self-sufficiency. However, self-sufficiency does not have a standard unit of measurement and may mean different things to different people. To answer this question, evaluation planners, with stakeholder input may want to define the term self-sufficiency and identify observable measurement units. Research questions with outcomes not clearly measurable may also require additional consultation with the selected evaluator.

- **Discreet, High-Level, and Limited in Number.** In general, key research questions should be discreet, meaning that they do not overlap one another. Typically, key research questions are written at a relatively high level and are few in number. A limited number of research questions help all involved stay focused on the “what” and the “why” of a state agency-sponsored evaluation and help clearly articulate the scope of the evaluation to stakeholders, customers, and other interested partners. The selected evaluator will examine the key research questions, explore their relevance to the study, and develop a more discreet set of questions tied to methodology.



Key Research Questions Are

- Specific and measurable;
- Answerable;
- Discreet, high-level, and limited in number (focused on the “what” and the “why”); and
- Rooted in firm knowledge, based on past efforts, and set with realistic expectations.

- **Rooted in Firm Program Knowledge and Realistic Expectations.** Strong research questions are rooted in firm program knowledge, based on understanding past similar efforts with demonstrated program results, and set with realistic expectations for conducting a study that addresses the research questions and explains how the evaluation will be achieved.

4.4 Review the Existing Evidence Base

A preliminary evaluation plan should follow the completion of the logic model(s), the purpose and scope, and the key research questions. A key component of the preliminary evaluation plan is a literature scan or review of the existing research-based evidence related to the subject of the evaluation. The identified research-based evidence provides a foundation for the evaluation plan and design because it provides useful, timely information and justifies how the study will build upon the current knowledge base. The existing evidence will help determine the following aspects of the plan:

- Refine the evaluation purpose, scope, and key research questions by building off of and improving upon the existing evaluation work that has been done;
- Determine what aspects of the program to evaluate using a relevant evaluation design, data sources, and methodology corresponding to how components of other similar programs, systems, strategies, services, activities, or interventions were evaluated;
- Identify appropriate outcomes and how best to measure or otherwise assess them;
- Ensure that the evaluation builds upon the existing evidence and contributes additional information to the current base of evidence (i.e., the evaluation goes beyond what has already been done and sheds new light on the issues/questions);
- Consider how to best integrate evaluation-related activities into program operations; and
- Look ahead to how the SWA may want to disseminate and inform others of eventual evaluation results.

The **evidence review, also called a literature review or scan**, includes references to scholarly studies of programs, systems, strategies, services, activities, or interventions similar to the proposed evaluation of a workforce system program. In particular, evaluations of other job training programs, work-based learning or statewide career pathway systems may be organized and summarized according to how those findings from each study relate to the proposed evaluation plans. In addition, the literature review or scan includes the:

- Studies' methods;
- Overall design and level of rigor;
- Types of data collected, data collection and analysis methods used;
- Implementation processes observed; and
- Research findings and recommendations of interest.

When this level of information is not available in a study's public report, the literature review or scan can also document the missing elements of the evidence gathered. The evidence base collected for the subject of the evaluation may not be limited to exact replicas of the program or its elements. Evaluation planners may want to research subjects or topics related to an area of study that apply to the proposed evaluation plan. For example, research on programs that serve different populations, and with some similarity or variation in design or services, may be useful for the evidence base.

Scholarly evidence sources and publications are available on Federal research and state agency websites, in peer-reviewed journals, and on the research and evaluation websites of universities, non-profit foundations, or professional evaluation firms. Although there is no centralized database of state-conducted evaluations, the NASWA report (NASWA, February 2017) identifies recent research areas pursued by states. Another form of evidence review that is of particular use is a meta-analysis. A meta-analysis is a quantitative statistical analysis of several separate but similar experiments or studies in order to test the pooled data for statistical and practical significance. For workforce development evaluations, there are several primary sources of studies listed in the table in Exhibit 4.4.

Exhibit 4.4: Sources for Finding Workforce Evaluation Research Studies

Sources	Descriptions and Hyperlinks
DOL, Chief Evaluation Office (CEO)	<p>The CEO coordinates, manages, and implements the DOL evaluation program, working closely with all offices and agencies throughout the Department. While offices/agencies conduct their own studies, CEO often commissions or conducts studies on their behalf, including many involving the workforce system on behalf of ETA. CEO maintains the following facilities for researching studies.</p> <p>Clearinghouse for Labor Evaluation and Research (CLEAR): This is a searchable database of identified and summarized studies assessed for quality by the CEO. Database includes all types of studies, including those not conducted by DOL. https://clear.dol.gov/</p> <p>CEO Completed or Current Studies: The CEO maintains these two lists of studies they have commissioned/conducted that provide links to study reports and other information. https://www.dol.gov/asp/evaluation/CurrentStudies.htm</p>
DOL, Employment & Training Administration (ETA)	<p>Research Publication Database: This database provides access to a collection of research and evaluation reports commissioned by ETA to help guide the workforce system in administering effective programs. The searchable database includes studies back to 1995. ETA also provides an annotated bibliography that summarizes selected research publications. http://wdr.doleta.gov/research/eta_default.cfm</p> <p>Workforce System Strategies (WSS): WSS provides a collection of resource profiles supported in ETA's TA platform, WorkforceGPS. The profiles summarize reports, various evaluation studies (beyond DOL/ETA studies), TA tools, and guides (evaluation and other types of guides and tools). While not targeted to only evaluation studies, it is key word and topic searchable. https://strategies.workforcegps.org/</p>
Department of Education, Institute of Education Sciences (IES)	<p>IES What Works Clearinghouse (WWC): This entity reviews existing research on different programs, products, practices, and policies in education to provide educators with the information they need to make evidence-based decisions. The "Review of Individual Studies" tab on the home page provides a searchable database of reviewed studies. https://ies.ed.gov/ncee/wwc/</p>
Department of Health & Human Services, Administration for Children & Families (ACF)	<p>Office of Policy, Planning, Research, and Evaluation (OPRE) Research & Evaluation Clearinghouses: In addition to conducting studies on ACF programs and populations served, OPRE conducts systematic reviews of the evidence and maintains several research clearinghouses (searchable databases) organized by major topic areas. The topical clearinghouses include self-sufficiency, employment strategies for low-income, strengthening families, home visiting, and childcare and early education. https://www.acf.hhs.gov/opre/research-and-evaluation-clearinghouses</p>
U.S. Government Accountability Office (GAO)	<p>GAO Study Reports Database: GAO is an independent, nonpartisan agency that works for Congress, and in general, investigates how the federal government spends taxpayer dollars. Studies generally of most interest to workforce evaluators include those that involve auditing agency operations to determine whether federal funds are spent efficiently and effectively and reporting on how well government programs and policies meet their objectives. To research its database, use the "Reports & Testimonies" tab and either browse by category or use the keyword/advanced search capability found at the top of the webpage. https://www.gao.gov/</p>
Key Journals for Workforce Research	<p><i>Journal of Labor Economics, Social Science Review, and the Journal of Public Policy and Analysis</i></p>

4.5 Determine the Research Design and Data Collection Approach

Once the SWA decides on the evaluation purpose and scope, reviews and documents an understanding of the evidence base, and refines the key research questions accordingly, the next task for the preliminary evaluation plan is to determine the research design (i.e., methods) and approach to data collection. A number of factors apply to the decision-making processes for research design and data collection approaches, such as the following:

Methods to Most Accurately Answer Key Research Questions: Some questions, such as who is participating in a program and the characteristics of their participation, may be best answered with an implementation or descriptive outcomes study, whereas other questions about the effectiveness of the program are likely to be best answered with a pre-post outcomes study, RCT, or quasi-experimental study. In some cases, the SWA may want to conduct a study that includes several types of evaluation. For example, often an outcome or impact study will also include an implementation and/or cost study component. The key guiding factor in making the final choice of study design is *what* the agency wants to learn and why, and how sure it wants to be about the findings.

Organizational Capacity to Participate in the Evaluation: Consider how the evaluation activities will blend into the implementation activities of the program, system, strategy, service, activity, or intervention included in the study. Discuss the feasibility and options to carry out and participate in the selected evaluation with organization or agency managers or operators that implement the subject or topic of the study. Include other key stakeholders and partners to identify their organizational capacity to participate. For example, RCT evaluations of service delivery interventions integrate a random assignment process into the participant intake and enrollment processes that may span multiple partners or service providers. Each partner engages in discussions and negotiates agreements to participate in the evaluation.



Research Design & Data Collection Approach Factors

- Methods to most accurately answer the key research questions;
- Organizational capacity to participate in the evaluation;
- Organizational capacity to conduct the evaluation; and
- Data availability for the desired type of evaluation or capacity to collect it.

Organizational Capacity to Conduct the Evaluation: The selection process for an evaluator (see Section 5) also depends on the state's capacity to conduct the evaluation. While funding may be a driver to building evaluation capacity, investments in evaluation management development, staff training, strategic and long-range planning, budgeting, and technical assistance are key elements. Each of these elements supports organizational evaluation capacity, whether the evaluators conduct in-house studies, partnerships are formed with research universities to manage administrative data, or third-party evaluators are procured to conduct independent evaluations. Section 3 includes the cost and rigor considerations for thirteen different types of implementation, outcome, impact, and cost studies. Given the numerous research and evaluation approaches available for state workforce agencies use, organizational capacity and costs for evaluations can also vary greatly.

Data Availability for the Desired Type of Evaluation or Capacity to Collect It: In addition to a final determination of evaluation type(s) for the evaluation plan, the evaluation planner and the selected evaluator will finalize the details to conduct and carry the research: **research design (i.e., methods) and data collection approach**. The type of evaluation and the key research questions will present different methods, data sources, and data collection options. **Data availability or the agency's capacity to collect data are critical factors** in deciding the type of evaluation to conduct. The selected evaluator will refine the study methodology and data collection approach; however, as part of the state evaluation planning process, data availability and capacity to house, transmit, and secure the data must be addressed to put the evaluation on the right track. How the data needs are identified and resolved within the context of the evaluation methodology allow the evaluation planner to work most effectively with the selected evaluator— in-house unit, partner university/organization, or third-party entity. The table in Exhibit 4.5 provides a broad snapshot of data requirements and sources for the four major types of evaluation, followed by additional considerations.

Exhibit 4.5: Data Requirements and Sources for Different Evaluations

Impact Studies	Outcome Studies	Implementation Studies	Cost Studies
Data Requirements			
An impact study needs baseline data on the characteristics of sample members and on pertinent outcomes before random assignment or program enrollment. It also requires follow-up data on pertinent outcomes after random assignment or program enrollment.	All outcome studies require individual-level outcome data on the population or sample of interest. Longitudinal outcome studies need baseline data on the characteristics of sample members and on pertinent outcomes before and after program enrollment.	An implementation study requires a defined set of qualitative and quantitative program and participant data related to the topics covered in the study.	All cost analyses require program cost and outcome information. Cost-effectiveness and benefit-cost studies also require results from impact or outcome studies and data from published sources.
Sources			
Program records on sample members, activities, and outcomes; automated records data on outcomes; and survey data on sample members, activities, and outcomes.		Observation and interviews at program locations.	Program financial records, published cost data, and impact and outcome studies.

- All Impact Studies and Most Longitudinal Outcome Studies.** Impact and outcome studies require individual-level data on sample members' characteristics and on their outcomes before, during, and after program participation. The samples are collected from cohorts in the target population and, if the study is to be generalized to that population, may ideally be randomly drawn to make evaluative comparisons. The data collected for the treatment and control or comparison groups in impact studies from the same source increases fidelity, validity, and reliability. The source may be longitudinal surveys, program records, and/or automated records, such as earnings records from the Unemployment Insurance (UI) system and postsecondary education and training information from the National Student Clearinghouse.

- **Implementation Studies.** Implementation evaluations require data collected in the field about the program being evaluated. These data obtained through observation, field interviews with program staff and other actors, and program documents describe processes, challenges, successes, and lessons learned. They also frequently use outcome data, although these are usually aggregate data rather than the individual-level data used for outcome and impact studies.
- **Cost Studies.** These use financial data to determine cost-effectiveness and benefit-cost from program services. Cost studies supplement selected results from outcome, impact, and implementation studies. Cost comparison may include outcome data for several aspects of the studies, such as unit-cost and return-on-investment estimates.

4.6 Prepare the Preliminary Evaluation Plan

The table in Exhibit 4.6 describes key design elements to consider in the evaluation approach and to prepare a preliminary evaluation plan. The plan should include a discussion of the following elements:

- The logic model;
- Purpose, scope, and key research questions;
- Evidence base (this may be a separate document or a summary in the plan with details provided separately);
- Evaluation type(s) demonstrating an understanding of that kind of evaluation, the feasibility of carrying it out, and why it is appropriate for the proposed study;
- High-level determinations concerning research design (i.e., methods) and data sources/data collection, recognizing that the selected evaluator will address the specific methodological details; and
- If the study needs to address protecting participants' rights (depends on type of evaluation; see Section 3.6 for a discussion of this aspect of evaluation planning).

The following additional considerations may also factor into the plan, as necessary:

- **Budget Needs.** Upon going through the planning process, an evaluation design plan may ideally identify all research needs. However, pricing the plan is incumbent on the relationship between funds available and the length and depth of the study. The evaluation plan, in terms of the type of evaluation and research design (i.e., methods and data collection) identified, may need to be modified or incrementally funded to meet the proposed research needs.
- **Timeline.** Timelines are critical to determining the feasibility of the planned evaluation project. Evaluators with little or no experience may underestimate the amount of time needed for the various phases of an evaluation. If planners are new to evaluation research, they may want to review samples of other research and evaluation timelines to help map out a tentative schedule. Section 3 provides information to help estimate a calendar schedule and timing for various evaluation activities
- **Protecting Participant Rights.** The primary purpose of the Institutional Review Board (IRB) is to protect the welfare of human subjects used in research. If the type of evaluation proposed requires IRB approval, factor appropriate time into the schedule for review, approval, and continuing review. IRB reviews can take considerable time and may have cost implications to both the schedule and the budget. Section 6 provides information on the IRB process that can help with plan adjustments, accordingly.

Exhibit 4.6: Key Design Elements for Evaluation Plans

Key Elements	Impact Studies	Outcome Studies	Implementation Studies	Cost Studies
General Purpose	An impact analysis estimates the difference in individual outcomes attributable to a specific program or policy.	An outcome study compares individual outcomes against goals, across programs or locations, or over time.	An implementation study documents program operation or compares it against goals, across locations, or over time.	A cost analysis estimates program costs, makes cost comparisons, or weighs costs against outcomes or impacts.
Study Purpose	Specific purpose statement(s) aligned with overall purpose of the type of study—impact, outcome, implementation, and/or cost.			
Study Scope	<ul style="list-style-type: none"> • Sites or areas • Target population • Outcomes addressed • Observation period 	<ul style="list-style-type: none"> • Sites or areas • Target population • Outcomes addressed • Observation period 	<ul style="list-style-type: none"> • Sites or areas • Elements addressed • Timing and length of study 	<ul style="list-style-type: none"> • Sites or areas • Elements addressed • Timing and length of study
Types of Key Research Questions	Hypotheses based on a review of evidence base	Questions to assess whether the program is meeting its objectives	Questions to assess program delivery process, such as: <ul style="list-style-type: none"> • How is it operating? • Is it operating as planned? • Level of participation? 	Questions to assess program costs such as: <ul style="list-style-type: none"> • What is the cost? • Is program cost-effective? • Is it a good investment?
Analysis Approach (see Section 6.2 for more on data analysis plan)	<ul style="list-style-type: none"> • Research design <ul style="list-style-type: none"> ○ Experimental ○ Quasi-Experimental ○ Rapid Cycle ○ Theory-Based • Sampling methods and sampling frame • Statistical methods (tests by which impacts are determined to be statistically significant) 	<ul style="list-style-type: none"> • Research design <ul style="list-style-type: none"> ○ Longitudinal ○ Cross-Sectional • Sampling methods and sampling frame • Statistical methods 	<ul style="list-style-type: none"> • Research design <ul style="list-style-type: none"> ○ Case Study ○ Fidelity Study ○ Performance Study ○ Site Comparison Study • Qualitative research methods such as interviews, focus groups, document reviews, observations 	<ul style="list-style-type: none"> • Research design <ul style="list-style-type: none"> ○ Cost Analysis ○ Cost-Effectiveness ○ Benefit-Cost • Cost research methods • Cost items and considerations such as staff time, budget aspects expected to be changed, total costs, and cost breakdowns of interest
Data Collection/Data Sources	<ul style="list-style-type: none"> • Survey data collection • Records • Other data collection 	<ul style="list-style-type: none"> • Survey data collection • Records • Other data collection 	<ul style="list-style-type: none"> • Field data collection such as interviews or focus groups, program observations • Record reviews such as program documents, data on participation and completion • Other data collection 	<ul style="list-style-type: none"> • Cost data collection • Records such as internal financial records or databases, program records and budgets, other administrative data • Other data collection

5. Select an Evaluator

A high-quality evaluation sponsored by a SWA is dependent on a well-conceived preliminary evaluation plan and careful selection of the evaluator who will conduct the study. To obtain a high-quality evaluation, an SWA needs to select an evaluator with the expertise and capacity to conduct the type of evaluation it is pursuing. In general, a state agency uses three approaches to secure an evaluator:

- In-house unit;
- Partnership with a university or other organization; or
- Third-party firm or organization selected via an RFP process.

This section provides guidance on the qualities to look for in any potential evaluator, and processes and considerations for each of the three approaches.

5.1 What Approaches to Consider

A first step is to determine which of the three approaches to use—in-house, university/other partnership, or third party—for the evaluation. Two key factors in choosing between an in-house unit, a partnership approach, or a third-party approach involve **in-house staffing levels** and **evaluator independence**. For many types of evaluation studies, an SWA may have the in-house expertise to conduct the study, especially among the larger states. However, as noted by the NASWA report, many states are hampered by inadequate staffing levels in their workforce research shops and/or state pay-level requirements that do not allow for employing the highly experienced, senior-level evaluators who would typically lead a major study (NASWA, February 2017).

In addition, the complexity of conducting impact studies (e.g., RCT or quasi-experimental) means the qualifications, experience, and dedicated time commitment of a potential evaluator are critical to success in conducting the evaluation. Furthermore, impact studies generally demand evaluator independence to ensure that the study is conducted objectively and the results can be “trusted.” Consequently, and in general, it is advisable for states to use an independent third-party evaluator for impact studies.

The third factor involves identifying the **levels of and types of experience and qualifications** for the evaluator(s) selected to conduct the evaluation. The table in Exhibit 5.1 outlines some questions to ask and examples of evidence to examine that may help determine which qualities are most important, which can then, in turn, inform the decision about which of the three approaches to take. “Potential evaluator” in the context of this table means in-house unit, university/other partner, or third-party firm/organization. Note that in the case of a third-party evaluator, the SWA will solicit bids and specifically seek a firm or organization with the qualifications identified. The table can also help with thinking through RFP specifications concerning experience and qualifications.

Exhibit 5.1: Evaluator Selection Questions and Evidence of Experience

Topic	Questions	Examples of Evidence
Evaluator Competencies and Experience	<p>Does the potential evaluator demonstrate:</p> <ul style="list-style-type: none"> • Competency through experience in conducting the type of evaluation the SWA wants conducted? How much experience? • Knowledge of and experience with subject matter of the proposed evaluation or a closely related topic (e.g., transitional job programs, employment services, youth job training programs)? 	<ul style="list-style-type: none"> • Summary of previous studies using proposed evaluation design or in a similar content field • Publicly available links to previous publications or other deliverables from these studies • Summary of the vendor's mission, history, and experience of potential evaluator • Demonstrated record of previous clients with similar needs
Staff Qualifications	<p>Does the potential evaluator have:</p> <ul style="list-style-type: none"> • A team that includes a mix of seniority levels—senior-level, mid-level, and junior-level staff? (a mix helps with cost-effective conduct of an evaluation) • The staff qualifications with related education and experience, for the Project Director (PD) and Principal Investigator (PI) (or team members in similar lead roles)? • Other required staff qualifications/skills needed to conduct the evaluation as envisioned (e.g., proficiency in data collection and analysis, observational techniques)? 	<ul style="list-style-type: none"> • Resumes or C.V.s • Publications written by senior staff demonstrating familiarity or skills in a particular evaluation type or context • Identification of staff by tasks described in the evaluation requirements
Capacity and Resources	<p>Does the potential evaluator have:</p> <ul style="list-style-type: none"> • Sufficient capacity to carry out the tasks that are likely associated with the proposed type of evaluation? For example, an evaluation team conducting a multi-site RCT study would likely require more staff than an evaluation team conducting an outcomes study in a single site. • Demonstrated capacity and resources to: (1) collect data, (2) analyze data (e.g., statistical analysis programs), (3) interpret the results, (4) produce graphs and tables, (5) write reports, and (6) provide technical assistance on any evaluation-related issues? • Administrative, IT, and publication support needed for the evaluation and subsequent deliverables? 	<ul style="list-style-type: none"> • Details and examples of similar evaluation capacity by team members • Demonstrated administrative and IT resources to meet the evaluation requirements (such as management and review processes, interview and site visit protocols, and secure data and transfer sites) • Demonstrated uses of data collection software and/or analysis tools created and/or used and examples within previous studies • Sample products produced for previous clients
Specialized Knowledge	<p>Does the potential evaluator have:</p> <ul style="list-style-type: none"> • Specialized knowledge to meet the evaluation requirements? (e.g., familiarity with relevant geographic, cultural, or other contextual elements) 	<ul style="list-style-type: none"> • Demonstrated credentials assure that they have appropriate skills, cultural competence, knowledge, and professional training to conduct the study in accordance with the standards and principles of the evaluation profession.

5.2 In-House Unit/Staff Approach

State workforce administrators and evaluation managers will likely readily know if the agency has the in-house capability to conduct the planned study. That expertise may reside within a Labor Market Information (LMI) unit or elsewhere in the agency. If the evaluation involves using education data or may be of interest to the state department of education, especially the unit involved in career technical education, both agencies may be able to establish an in-house evaluation team consisting of qualified evaluators drawn from both the workforce and education agencies. Key questions to consider include the following:

- Does the agency have staff with the requisite knowledge, skills, and experience in the kind of evaluation planned? Does the agency have someone who can lead the study and function as the project director and/or principal investigator? For example, the agency may have qualified staff to conduct implementation and certain types of outcome studies, but maybe not to conduct an RCT or QE study.
- Does that staff have the calendar time and available work hours to conduct the study expeditiously in relationship to their other duties? Can they be allocated full time or nearly full time to conduct the study?
- Can the agency field an evaluation team that is relatively cohesive and can be committed to working on the study (in other words, splitting up pieces of an evaluation study among too many team members is generally not advisable)?
- Would it likely be more or less expensive to conduct the study in-house?
- To what extent is evaluator independence critical to state and stakeholder trust of the findings? Whether the agency has the qualified staff and in-house resources to do the study, perceptions of the quality and trustworthiness of study findings may dictate the use of a third-party evaluator.

5.3 Partnership Approach

As noted in the NASWA report on state capacity building, a number of states form partnerships with universities (and other research organizations) to expand their capacity to conduct program evaluations. There are some advantages to establishing formal partnerships with universities, such as the ability to:

- Supplement and augment agency research and evaluation staff with highly qualified evaluators, especially regarding quasi-experimental and RCT evaluation approaches and possibly at less cost than hiring a third-party private research firm;
- Minimize costs of conducting major evaluations by employing graduate students; they need real projects to work on and can usually get college credit for doing so;
- Host and otherwise support the use of “big data” by providing secure computing facilities (social science research universities/centers);

- Obtain IRB approvals for human subjects' research since many major universities have established, federally registered IRBs;
- Minimize costs of conducting major evaluations by providing in-kind resources to support studies on common interests and of mutual benefit to the state (e.g., state universities);
- Bring other university, foundation, and research organization partners to the table—major social science research universities often have existing networks and partnerships of their own; and
- State or local socio-demographic and socio-economic context research is often included in state-level evaluations—and it is likely the institution has done a lot of work in this area already.

Partnerships, such as those mentioned above, are options to conduct a specific study (a one-time effort) or to form an ongoing relationship to make state administrative data available for external research. An ongoing relationship with a research entity at a university is much more involved and affords many more options for external evaluations, such as in Ohio, as documented in the NASWA report.

The key to forming partnerships is a mutually beneficial partnership agreement or Memorandum of Understanding (MOU) to document the partner roles and responsibilities, specific in-kind resources from the university, and in-kind state resources. Allocation or payment of “hard costs,” among other items, also undoubtedly requires a form of agreement. In this regard, it is different from hiring a third-party entity through an RFP process. Hiring in-house expertise means the agency controls and manages the types of evaluation and research produced; partnering means there is a mutual compact about the evaluation and research work performed and studies produced.

5.4 Third-Party Approach

In a typical approach to acquire the services of a third-party evaluator, states may use an RFP bidding process. In general, in the sponsoring state agency, the evaluation planner will need to work with a procurement office to prepare the Statement/Scope of Work (e.g., tasks, timeline, deliverables), articulate bidder qualifications (corporate and staff), and provide a general budget estimate as the basis for development of the actual RFP. In addition, publicizing the RFP is critical to attracting as many qualified bidders as possible.

Statement/Scope of Work: The evaluation preliminary plan and high-level timeline and budget will help in developing the Statement/Scope of Work for use by the procurement office. In general, the scope of work provides background information and context for the evaluation, purpose and scope of the evaluation, a list of the key research questions, details for specific tasks to perform, and a list of deliverables and deliverable timing requirements.

Evaluator Qualifications: Selection of an evaluator requires an understanding of evaluator competencies and experience, staffing qualifications, capacity and resources, and specialized knowledge. Examples of some of the questions to consider when making the selection are in the table in Exhibit 5.1.

At a minimum, it is reasonable to require a qualified evaluator to have:

- (1) Experience implementing an evaluation of the proposed evaluation type or in the proposed content field;
- (2) Senior staff with post-graduate education and five or more years of experience that can demonstrate the technical skills necessary to implement the study;
- (3) Capacity and resources to negotiate agreements, facilitate stakeholder and partner meetings, and complete data collection, data analysis, and report writing; and
- (4) Other specialized knowledge that may be critical to understanding the subject of the evaluation and/or the methodologies to be employed.

Overall, evaluator competencies and staff qualifications are often the important factors to ensure a quality evaluation. These criteria, as well as specialized knowledge of the evaluation subject or topic, are described in RFP selection criteria.

Assessment of Bidder Proposals: Ideally, the agency will receive several proposals in response to an RFP. The state procurement office likely has a process in place to review and assess proposals to select the winner. Most often, the selection criteria are weighted for the technical proposal and staff qualifications. For example, while evaluator and staff experience are typical factors, the agency may also value content knowledge or experience in the geographic area in the evaluation. Assigning weight to each factor can help identify the evaluator that best meets the needs for the study. To assess proposals, the agency may want to consider appropriate trade-offs between quality and cost competitiveness. While there will be some cost variation in potential bids, most bids are responsive to the key needs outlined in the RFP, and budgets should reflect the associated level of effort.

Getting the Word Out—Publicizing the RFP: Once approved, the RFP is posted in the public domain. The state agency, no doubt, has a process for formally announcing the RFP and requesting responses similar to what the federal government does for its evaluations and other procurement actions. However, to ensure the broadest outreach to a wide range of potential evaluators, additional outreach may be useful to increase the number of responses. All additional “advertisements” should comply with state procurement process. The following advertising avenues may be appropriate to consider, and when doing so, be sure to include information on where to find the official state announcement and copy of the RFP:

- Send letters or emails that announce the RFP to a likely group of evaluators;
- Post a notice about the RFP on evaluation-focused websites;
- Post a notice about the RFP on the state agency’s website;
- Announce the RFP on any social media sites associated with the agency, including, but not limited to, LinkedIn, Twitter, and Facebook; and
- Announce the RFP in local, state, or national evaluation newsletters or publications.

6. Implement the Evaluation

The key elements of preliminary planning described in Section 4 lead to implementation of the evaluation. Evaluation implementation activities are the responsibility of the selected evaluator, whether it is in-house staff, third-party evaluator, university or other organizational partner, or the state agency in charge of the evaluation. As such, the state workforce agency (SWA) will need to understand key implementation elements to oversee the process and to make overall determinations concerning the study's timeline, resources needed (staff and funding), and other factors that could affect scope and quality. This section describes key evaluation implementation activities that state workforce administrators and evaluation managers need to be successful:

- Creating the evaluation design report to guide each phase of the evaluation;
- Developing a data analysis plan for inclusion in the evaluation design report;
- Addressing the protection of participants' rights (for certain types of studies);
- Linking or coordinating evaluation and program activities (for certain types of studies); and
- Reporting by the evaluator on interim and final study findings.

6.1 Create the Evaluation Design Report

The first task of the selected evaluator will be to develop a plan that follows the technical proposal to implement the evaluation. Typically called an Evaluation Design Report (EDR), this document serves as a guide for the evaluator and state agency to unfolding various evaluation activities. Depending on the type of evaluation conducted, it also serves as a guide to program staff on how various evaluation activities link to or coordinate with program operations. The EDR builds upon the preliminary evaluation plan (see Section 3) and includes a data analysis plan developed by the selected evaluator. The selected evaluator expands and refines each element in the state's preliminary evaluation plan to create a detailed and feasible evaluation implementation plan. In addition to expanding upon the elements of the preliminary evaluation plan and depending on the type of evaluation conducted, the evaluator may propose or otherwise suggest variations and additions concerning the following items:

- Appropriate and reliable outcomes that can be measured through available resources;
- Evaluation method(s), including data collection processes and sources;
- Data analysis plan/approach, including suitable controls for mitigating any threats or risks to successful interpretation of findings, and overcoming any limitations to the maximum extent possible;
- Timeline and milestones for evaluation activities; and
- Reporting details to convey evaluation progress, results, and findings.

The common EDR elements are included in the table in Exhibit 6.1. Most of these elements are described or mentioned in Section 4: Develop an Evaluation Design Plan.

Exhibit 6.1: Elements Included in the Evaluation Design Report, as Appropriate

Element	Summary Description
Program Purpose, Scope and Logic Model	Describe the purpose and scope of the program, service activity, or intervention. Details for each of the organizational and service delivery components are identified (if more than one) to explain the theory of change or logic model. Specify the expected outcomes or impacts from the program and possible effects of other activities on the variables of interest beyond those in the evaluation.
Evidence Base (Literature Review)	Review, synthesize, and summarize in the EDR to understand gaps in research or replicate a comparable study. Include relevant past interventions and evaluations, such as findings, and discuss research designs used, if applicable. Describe how the evaluation will enhance the state's workforce system, the broader workforce system, and/or contribute to the workforce evaluation literature.
Study Purpose and Scope	Articulate the specific purpose of the study and describe the components of the program, sites, and geographic areas that are to be evaluated and included in the study scope and other scope parameters.
Key and Detailed Research Questions	Identify the detailed research questions linked to the specific program, services, or intervention, and align the questions within a logic model or theory of change. These questions may be descriptive, or process-, general outcomes-, and/or impact-related (hypotheses about expected outcome changes due to the intervention), depending on the types of evaluation being conducted.
Evaluation Type	Describe the type of evaluation and methods to use and provide the justification for the approach selected. Describe the overarching evaluation goals or objectives and explain their appropriateness to the evaluation design and program or intervention to be evaluated.
Participants, Samples, and Units of Analysis	Describe the program sites and/or geographic areas selected for implementation studies OR the unit or units of analysis for outcome and impact studies. Define the eligibility or exclusion criteria for program participants (i.e., the target population, if applicable); the overall population targeted or from which generalizations will be made; whether the evaluation will be conducted on the entire population vs. a sample; and, if it is a sample, whether it is representative.
Data Collection Plan	Propose the primary process and/or participant outcomes; valid, appropriate, and reliable measures; and any qualitative information collected. Describe data sources and collection methods for each measure, and for each process or qualitative line of inquiry, include new primary or existing administrative (secondary such as from an MIS) data collection sources, along with details about unique sources to the program and evaluation. Delineate timeline-specific processes and schedules for collecting the data from each source and describe proposed or drafts of data collection protocols/instruments to address the research questions.
Data Analysis Plan (see Section 6.2)	Develop a data analysis plan based on the research questions and type of study selected. For example, impact studies may include sampling plans that describe the purpose, method of sampling, and anticipated sample sizes. For RCT impact designs, include the power calculations to use, describe the process for random assignment of participants to treatment and control groups, and demonstrate any impact formulas and other analytical assumptions. Describe the analytical software or other tools that are appropriate to the evaluation design. Discuss validity/threats and mitigation strategies—whether issues of internal and external validity, threats to validity and their implications, and strategies to mitigate selection bias, if needed.
Supplemental Studies	Supplemental studies, such as implementation or cost studies often become part of a larger evaluation for outcome or impact studies. The EDR may cover plans for such studies, including research questions, data sources and collection, and analysis methods.
Reporting	Provisions for reporting identify presentations, brief reports, as well as interim and final reports; dates for deliverables, ongoing progress, and handling data sets; and a de-identified data set(s) at the conclusion of the study (as appropriate).

6.2 Develop the Data Analysis Plan

The selected evaluator is responsible for developing data analysis plan based on the preliminary plan the SWA provides. The evaluator will use the preliminary plan as the foundation, suggest expansion or modifications, and connect the dots via the data analysis plan—how to get from the high-level purpose, scope, key research questions, and research design to the analysis and determination of findings. A data analysis plan outlines the key steps and processes used to analyze the data collected prior to actually collecting data. The data analysis plan is a roadmap that connects the research questions to the data, describes how the data are analyzed, separates the key research questions into “testable” hypotheses, and aligns each with the data and analytical methods used. A data analysis plan identifies the metrics for outcomes measured—both process outcomes that describe the program implementation activities, and outcome measures that define the intended results, along with the variables examined. The table in Exhibit 6.2 provides a list of items to be covered using an RCT study example. Depending on the type of study conducted, some of these items would not apply. A data analysis plan is an integral part of the EDR; hence the overlap between tables.

Exhibit 6.2: Items to Cover in a Data Analysis Plan (RCT Example)

Items	Brief Description of What to Include
Translation of Research Questions	Translate questions and possibly logic model elements into testable hypotheses. Specify key outcomes and specific associated measures, steps along the causal chain to be measured, and whether any subgroup analyses will be done. Describe validity, reliability, and use history of each measure to be used.
Data Sources	Describe exactly which sources will be used, such as program records, administrative data, or specific data collection instruments designed for the study. Link each outcome to be measured to how it will be measured (method to be used such as a survey, test, or interview, or drawn from an existing data set). Include data collection protocols and instruments (e.g., surveys, interview guides).
Study Population, Sample, and Sampling Plan	Describe the overall study population and sub-groups to be studied. Describe purpose and method of sampling, including how the sample will be drawn to be representative of the study population, and anticipated sample sizes. Describe random assignment procedures for both treatment and control groups. Include power calculations, any impact formulas, and other analytical assumptions.
Planned Analyses and Analytical or Statistical Methods	Discuss statistical analysis methods to be used, how hypotheses will be tested, and how potential confounding and bias will be assessed and addressed. Specify how variables will be constructed (e.g., what levels used, handling of missing variables, procedures for dealing with outliers). Specify how the treatment effect equation will be estimated (e.g., difference-in differences), regression controls to be included, calculation of standard errors. Specify methods for dealing with multiple outcomes, survey attrition (if a survey is done), and outcomes with limited variation.
Tracking Study Participants	Describe how the study will track participants and their data, including mechanisms for monitoring, tracking, and troubleshooting issues, and data security procedures. Specify how missing data will be handled (e.g., attrition rate calculations).
Validity/Threats	Discuss issues of internal and external validity, threats to validity, and implications. Describe strategies to mitigate selection bias, if needed.
Analysis Tools	Describe the analytical software or other tools appropriate to the study.

6.3 Protect Participant Rights

A key evaluation implementation activity involves protecting the rights of the individuals who participate in the study. This aspect generally pertains to impact (RCTs) and pre-post outcome studies. In these types of studies, the SWA will be collecting and storing detailed information about individuals who agree to participate. For studies involving individual-level data, the information study participants provide will allow the SWA to measure the effectiveness of the program. As such, the study participants, or human subjects, are the heart of an evaluation.

Although part of the evaluation implementation conducted by the selected evaluator, the preliminary plan will need to factor in whether the evaluation will require adhering to federal rules concerning the protection of participant rights via the Institutional Review Board (IRB) application and review process. It is important to note that human subject considerations have very practical implications for an evaluation timeline and budget, and they may also affect other major and minor features of the study, such as the following:

- Evaluator chosen to conduct the study (e.g., experience with the IRB review process and data security capacity);
- How the SWA and the evaluator will structure data collection (data types and methods);
- How much time to allow for IRB application and review;
- Additional resources needed to conduct informed consent; and
- Technology (encryption software, secure file transfer protocol [FTP]) needed to securely transfer and store participant information.

IRB Process Overview: To ensure appropriate protections are in place and maintained, 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 continual review by monitoring active research projects, reviewing and addressing 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 rules to protect research participants from harm. The table in Exhibit 6.3 summarizes the main roles and responsibilities of IRBs. IRBs are registered and regulated by the Department of Health and Human Services (HHS). In general, IRBs are at medical centers and universities; although, some research firms also host their own IRBs.

The selected evaluator will likely have experience with and knowledge about which IRB to use and how to contact the IRB, and the overall process involved to receive approval for the study (if required) before engaging human subjects. The IRB will let the evaluator know whether the study is “exempt” from review or is “non-exempt” and needs to undergo a formal board review.



Key Terms & Definitions

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 (Code of Federal Regulations 45 CFR 46.102(f)).

Private information is Personally Identifiable Information (PII) that can be used to trace a person's identity, that is not publicly disclosed nor publicly associated to the service or intervention received in a program.

Identifiable means that the individual agrees to participate in a study and may be associated with the information or disclosed to the evaluator during the course of the study.

PII is information that associates an individual's identity within an evaluation context; PII includes names, social security numbers, birthdates, addresses, and other related contact information.

The decision to include a protocol for IRB review—and the level of review, either expedited or full board— depends on a variety of the following factors, the:

- Level of risk to participants overall and relative to the potential social benefit of the research;
- Vulnerability of the population under study; and
- Steps the evaluator takes to minimize risks and safeguard participants.

Generally, studies that may pose minimal or greater than minimal risk to participants needs review by the IRB. Evaluation requirements that constitute exempt research, minimal risk, or greater than minimal risk are decided by an IRB, not the evaluator.

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.

Four 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;
- 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. The state agency sponsor for the study will want to be aware of and may be involved in executing these aspects of the study protocol.

Exhibit 6.3: 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 <p>Additional safeguards to protect rights and welfare of vulnerable populations (e.g., children, pregnant women, “persons who are economically or educationally disadvantaged”)</p>

Informed Consent: To collect information from and about individuals participating in a research study, the evaluator needs to obtain their legally effective informed consent (or that of a parent or guardian if the participant is a minor under age 18).

Obtaining informed consent includes giving prospective study participants sufficient opportunity to consider participation and minimizing undue influence and coercion. 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 participation. 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.



Informed Consent Tips

Undue Influence: using suggestive language, offering excessive reward or creating a sense of obligation to participate may deter an individual's consent.

Use of "Lay Language": articulating the purpose, procedures, alternatives, risks, and benefit in everyday terms that enables an individual to voluntarily decide to participate as a research subject.

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 **privacy** means collecting data in such a way that participants' information or personally identifiable information (PII) is not seen or overheard by others. For example, they take surveys in a private space where others cannot view their answers. Protecting **confidentiality** of participant information means:

- (1) Not sharing information about participants except with those authorized to have it; and
- (2) Complying with a study-wide plan for secure collection, transfer, storage, and use of participant information.

Data Security: The state sponsoring agency and the evaluator chosen to conduct the study have a collective responsibility to protect participant information and assure its **data security**. All parties will need to work together to create a climate of accountability and responsibility for the data collected, shared, and analyzed. Part of the accountability and responsibility includes establishing a data security plan. A data security plan describes the requirements to protect PII and identifies potential penalties for losing data and for failing to destroy data once they are no longer needed. It lays out steps to inform individuals or entities that their information was lost, stolen, or otherwise compromised. Robust data security can help safeguard against the accidental loss or disclosure of participants' information. The SWA and the evaluator will need to work out the specifics as they apply to the 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 concerning the protection of participant rights include the following:

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

The state sponsoring agency will want to be sure that the selected evaluator is prepared for such potential problems and may want to have specific procedures in place for the evaluator and the program operator to work together to resolve such issues should they arise.

6.4 Coordinate Evaluation and Program Activities

This section is only relevant when the SWA is evaluating a specific program, service, or intervention that requires field research and direct interaction with the program implementation staff and participants (e.g., grantees). Such an evaluation means that clear procedures are identified to establish regular connections during each phase of the study. The guidance provided here primarily pertains to interactions between the evaluator chosen to conduct the study and the program implementation staff. However, the state entity sponsoring the study has a responsibility to help set the stage for the interactions with the program being evaluated.

It is the responsibility of both the state sponsoring entity and the evaluator to ensure that the study team 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 can be operationalized through regular project management activities and communication mechanisms, such as conference calls or meetings, and ensuring the evaluator is informed of program design decisions or changes. Training program staff to understand the significance of the evaluation and to provide information as required by the evaluator supports the timely completion of a study. Program personnel who regularly communicate with program participants must be able to articulate the requirements of the evaluation. For example, they should be able to explain the random assignment process if an RCT design is used. Training provided by the evaluator informs program staff about the evaluation and ensures that program staff understand and work with the evaluation processes and requirements.



Benefits of Coordination

- Reduced burden on staff and participants;
- Clear information for participants;
- Minimal effect of the evaluation on the program; and

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;
- **Implementation of random assignment design** occurs if conducting an RCT; and
- **Additional data collection** (such as follow-up surveys) is conducted.

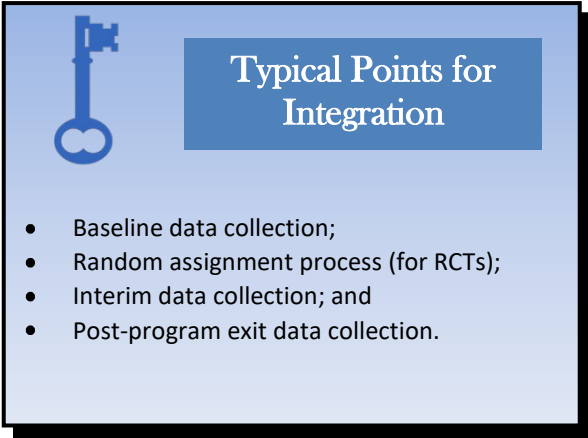
At each of these points, the evaluator and program staff need to work closely together to find the best way to integrate the study activities into program activities. Close and sensible integration of evaluation activities into program activities has many benefits, and includes:

- Reduction of burden on staff;
- Informed program applicants and study participants;
- Minimal effect of the evaluation on the program (ideally, no effect on the program); and
- Buy-in to the evaluation among both staff and study 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. Staff involved in conducting this data collection need training by the evaluator. The evaluator and program staff should discuss when baseline data are collected and how this data collection is integrated into the typical enrollment processes to reduce burden on both staff and study participants. For example, if the program already collects some data on participants at enrollment, it may be appropriate to consider

combining the evaluation data collection and standard program data collection into one document (i.e., one data collection form), or administer both forms simultaneously to ensure that the program and the evaluator receive all needed information.

Random Assignment Process for RCT Evaluations: Assignment of individuals to the treatment condition or the control condition (i.e., those who will receive status quo services) is conducted at the time of enrollment. The evaluator generally trains program personnel involved in this process. The program staff and the evaluator jointly decide upon the point of “randomization” within the program enrollment or intake flow. To the extent feasible, it may be useful to integrate the random assignment and other processes required for an RCT evaluation into the standard intake procedures. Such integration lessens the burden on program staff and provides for the informed consent requirements with applicants during the program enrollment process.



Typical Points for Integration

- Baseline data collection;
- Random assignment process (for RCTs);
- Interim data collection; and
- Post-program exit data collection.

For example, the random assignment step added to enrollment as they begin the process facilitates informed consent. When participants come to an information session, they complete eligibility screening forms. Once the selection process is completed, those individuals deemed eligible will progress forward and participate in an interview, at which time their consent is obtained and the baseline data collection forms are completed. Then, a staff member uses this information to conduct random assignment using a system designed by the evaluator. While there are many ways to integrate random assignment into the enrollment process, the key is to do it as smoothly as possible to not lose applicants’ interest, not burden program staff, and maintain continuity through the randomization period. Usually, evaluators develop an RCT procedures manual to document how the randomization process is handled.

Other Data Collection Activities (Interim or Post-Program Exit): The evaluator determines an appropriate timeframe for data collection and coordinates timing with the program staff well in advance of when collection will occur. If data collection involves surveying participants during the program or after program exit, program staff remind study participants about data collection activities when they are on the horizon. The evaluator may ask program staff to endorse any data collection effort whenever they make contact with participants, or the evaluator may ask to send a letter on behalf of the program. While program staff should not assume a large amount of burden for the evaluation data collection, it benefits both the program and the evaluation if program personnel do what they can to encourage participants to respond.

6.5 Evaluation Report Requirements

Reports regarding a project are deliverables that the evaluator will produce; these documents are important because they represent the culmination of all evaluation activities. They convey the evaluation’s findings to relevant stakeholders, including the sponsoring state agency, state stakeholders, DOL, the larger workforce community, and other potential funders. 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/or implementation. Interim reports typically describe program activities, assess fidelity to the model, and share activities and findings at given points

of a study period. Final reports serve as the official record of the evaluation and present all findings appropriate to the type of study conducted. Progress reports document evaluation activities along the way. It is up to the state agency sponsoring the evaluation to determine what types of reports need to be produced by an evaluator. Below is a description of different report types, followed by guidance on how to communicate reporting expectations to an evaluator.

Interim Reports: These reports convey findings about the program evaluation as it is implemented. Interim reports will allow all those involved to learn more about short- and medium-term participant or other outcomes during implementation. For multi-year evaluations, interim reports are submitted at the evaluation's mid-point. At a minimum, evaluators submit annual or semi-annual reports that include a description of program implementation, participant characteristics, and short-term outcomes.

Progress Reports: These reports are written reports provided on a regular basis (e.g., monthly, quarterly, or yearly) that provide updates on evaluation activities during the given time period. These reports can help: (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.

Final Reports: This report serves as the capstone to the study. Final reports are published on the state website, DOL's website, and/or elsewhere, depending on study sponsors and funders. Final reports are helpful to all those involved and ultimately state policymakers and the larger workforce community. Key topics typically covered by the final report, and to a lesser extent by the interim report, include the following:

- Program being evaluated or other subject of the evaluation;
- Economic, geographic, and/or political context that may have contributed to program implementation or evaluation results;
- Information on the research questions and evaluation methodologies used;
- Information on the sources of both quantitative and qualitative data;
- Program operations as planned and implemented (process/operational aspects);
- Analysis of quantitative and qualitative data;
- Interpretation of results and presentation of findings (objective presentation);
- Identification of lessons learned, or promising and best practices; and
- If appropriate to the type of study conducted, how other programs may use information in the report to replicate or scale-up programs like the one being studied.

Depending on the type of study conducted, the evaluator prepares and submits the public-use data set as part of final reporting deliverables.

Communicate Reporting Requirements to the Evaluator: The SWA will need to clearly communicate all reporting expectations in an RFP, in a joint agreement, or in guidance to in-house staff (see Section 5 for evaluator selection guidance). To ensure that everyone is clear, and in particular, to ensure the SWA receives appropriately budgeted proposals from prospective third-party evaluators via an RFP, the SWA should explicitly state how many reports the state sponsoring agency expects and when it expects them. Note that increasing reporting requirements increases evaluation costs, no matter who conducts the evaluation. The SWA should carefully balance the value of the information desired with cost considerations. Key things to communicate in writing include the following:

- What type of reports, how often, and approximate due dates and timeframes
- For progress reports, how often and what they need to cover with regard to evaluation activity reporting (e.g., monthly or quarterly)
- For interim and final reports, how many interim reports and what they should cover, when both interim and finals should be delivered for review as drafts, and when the final versions are due after incorporating comments (progress reports are typically not done in draft and final forms)

When specifying due dates for reports, keep in mind the following, depending on the type of evaluation being conducted and research design and data collection approach being used:

- **Final Reports.** Preparation and submission of these reports may need additional time to allow follow-up periods to track participant outcomes. For example, if evaluators plan to track participant outcomes 12 months after program enrollment, the SWA 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 also need time for analysis, writing, and revisions. Writing final reports process may take a period of two to five months, depending on the sample size, complexity of analysis, and number of revisions.
- **Interim Reports.** Interim reports take into account data collection periods and summarize evaluation activities as a given point, such as the analysis from site visits, interviews, focus groups, and other preliminary results. It may be reasonable to expect some results or details about program processes to measure the evaluation's progress. For example, if the program requires six months of training with a research interest in participant outcomes, an interim report may summarize of the implementation study examples share above and include preliminary input and output data about the first group of training completers.

Conclusion

The American Evaluation Association (AEA) succinctly summarizes the value and benefit of evaluation as “evidence to find out what is and is not working.” Furthermore, it helps governments, businesses, not-for-profit agencies, philanthropic foundations, and international organizations around the world “to make decisions about how best to allocate scarce resources, develop staff, choose quality products they need, and more effectively meet people’s needs.” Independent evaluations can increase public confidence about receipt of credible information about how funds are being spent, what is being accomplished, and what is not being accomplished.

The benefits of evaluation extend beyond a particular project when the findings are used to expand the project to a larger number of communities. For example, positive findings from a pilot program can be used to support dissemination and expansion of the program, as when a pilot parent education program becomes a national model based on evaluation of its effectiveness.

This toolkit contains references and links to many resources, like those from AEA, that are available in the public domain. As an agency works to broaden its research and evaluation capacity, consider the information in this guide as part of the principles, steps, and standards to build upon an evaluation framework. Each of the six sections presents options to help frame the following questions:

- What is the best way to evaluate workforce system programs, services or strategies?
- What are we learning when we evaluate?
- How can we use the learning from these efforts to improve our programs and services?

Appendix F may be of particular interest for SWA that want to determine the level of evaluation capacity that exists within their organizations. It includes two WIOA Quick Start Action Planners (QSAPs) to help further an SWA’s ability or expand capacity to improve a state or local workforce agency’s readiness to conduct rigorous evaluations. The Evaluation Readiness Assessment QSAP focuses on evaluation culture and awareness; funding strategies; data management; staff skills, capacity and knowledge; and strategic planning. The Evaluation Design and Implementation Assessment QSAP is organized into five short sections that focus on evaluation design and research questions, data collection and analysis plan, evaluator selection, participant rights, and reporting. The instructions provide an overview of, uses for, and steps to implement for the assessment.

Using this toolkit and the evaluation readiness assessment may help evaluation planners think about effective program evaluation as a systematic way to improve services, inform customers and stakeholders, and create a foundation for evidence-based practices.

Appendices

- Appendix A: Performance, Research and Evaluation Framework
- Appendix B: WIOA Evaluation Regulations
- Appendix C: Evaluation Resources and References
 1. National and Federal Policy, Planning and Guidance Resources
 2. Evaluation Resources and Online Toolkits
 3. Behavioral Insight Studies
 4. Cost Studies
 5. Data Analytics
 6. Implementation Studies
 7. Interrupted Time Series
 8. Logic Models
 9. Outcome Studies
 10. Power Analysis
 11. Quasi-Experimental Studies
 12. Randomized Controlled Trial (RCT) Studies
 13. Theory of Change
- Appendix D: Example of the Learning Agenda Cycle
- Appendix E: Logic Model Template
- Appendix F: Evaluation Readiness Assessment – Overview and Instructions
 1. Evaluation Readiness Assessment
 2. Evaluation Design and Implementation Tool
- Appendix G: Glossary of Terms

APPENDIX A: Performance, Research and Evaluation Framework

Performance	Monitoring	Research	Evaluation
The What			
Performance management processes or systems that use goals, measurement, analysis, and data-driven reviews to improve program results and the effectiveness and efficiency of agency operations. Performance information promotes data-driven decision-making. Program performance data tracked also feeds into evaluations.	Systematic, routine, and ongoing collection and review of project or program implementation in relationship to the stated plan or requirements (e.g., are participants receiving services consistent with program intent; is the program in compliance with the “rules”). Results of monitoring may be used for program management or evaluations.	Empirical process that uses workforce data to develop descriptions, measurements, comparisons, and tests of hypothesized relationships. Results usually fed into and used by evaluation.	Empirical analysis that uses program and other data to describe the operation of a program, measure the program impacts on outcomes of policy and program interest, and/or determine cost effectiveness of the program to identify improvements, best practices, and/or what works and does not work.
The How			
<p>Key components:</p> <ul style="list-style-type: none"> • Goals and priorities—strategic, cross-agency, performance • Strategic plans—mission, goals, strategies/programs, measurement • Performance measures—data that gauge how a program or organization is performing to inform strategic planning, accountability, decision making, improvements • Progress reviews—data driven reviews of goal progress and strategy/program performance • Reporting—performance transparency and accountability 	<p>Types of monitoring examples:</p> <ul style="list-style-type: none"> • Results monitoring • Process/activity monitoring • Compliance monitoring • Participant monitoring • Financial monitoring • Organizational monitoring 	<p>Types of research examples:</p> <ul style="list-style-type: none"> • Descriptive statistics • Descriptive analyses • LMI research • Research to support policy development • Data analytics 	<p>Types of evaluation examples:</p> <ul style="list-style-type: none"> • Implementation studies • Outcome studies • Impact studies (including behavioral insight studies) • Cost studies <p>(Section 2 provides more details).</p>
Adapted from Multiple Sources: NASWA, February 2017; Performance Improvement Council, no date; Council for International Development, June 2014; National Academy for State Health Policy, August 2015			

APPENDIX B: WIOA Evaluation Regulations

§ 682.220 What are States' responsibilities in regard to evaluations?

- (a) As required by § 682.200(d), States must use funds reserved by the Governor for statewide activities to conduct evaluations of activities under the WIOA title I core programs in order to promote continuous improvement, research and test innovative services and strategies, and achieve high levels of performance and outcomes.
- (b) Evaluations conducted under paragraph (a) of this section must:
 - (1) Be coordinated with and designed in conjunction with State and Local WDBs and with State agencies responsible for the administration of all core programs;
 - (2) When appropriate, include analysis of customer feedback and outcome and process measures in the statewide workforce development system;
 - (3) Use designs that employ the most rigorous analytical and statistical methods that are reasonably feasible, such as the use of control groups; and
 - (4) To the extent feasible, be coordinated with the evaluations provided for by the Secretary of Labor and the Secretary of Education under WIOA sec. 169 (regarding title I programs and other employment-related programs), WIOA sec. 242(c)(2)(D) (regarding adult education), sec. 12(a)(5), 14, and 107 of the Rehabilitation Act of 1973 (29 U.S.C. 709(a)(5), 711, 727) (applied with respect to programs carried out under title I of that Act (29 U.S.C. 720 et seq.)), and the investigations provided by the Secretary of Labor under sec. 10(b) of the Wagner-Peyser Act (29 U.S.C. 49i(b)).
- (c) States must annually prepare, submit to the State WDB and Local WDBs in the State, and make available to the public (including by electronic means) reports containing the results, as available, of the evaluations described in paragraph (a) of this section.
- (d) States must cooperate, to the extent practicable, in evaluations and related research projects conducted by the Secretaries of Labor and Education under the laws cited in paragraph (b)(4) of this section. Such cooperation must, at a minimum, meet the following requirements:
 - (1) The timely provision of:
 - (i) Data, in accordance with appropriate privacy protections established by the Secretary of Labor;
 - (ii) Responses to surveys;
 - (iii) Site visits; and
 - (iv) Data and survey responses from local subgrantees and State and Local WDBs, and assuring that subgrantees and WDBs allow timely site visits;
 - (2) Encouraging other one-stop partners at local level to cooperate in timely provision of data, survey responses and site visits as listed in paragraphs (d)(1)(i) through (iv) of this section; and
 - (3) If a State determines that timely cooperation in data provision as described in paragraph (d)(1) of this section is not practicable, the Governor must inform the Secretary in writing and explain the reasons why it is not practicable. In such circumstances, the State must

cooperate with the Department in developing a plan or strategy to mitigate or overcome the problems preventing timely provision of data, survey responses, and site visits.

(e) In fulfilling the requirements under paragraphs (a) through (c) of this section, States are permitted, but not required, to:

- (1) Conduct evaluations that jointly examine title I core program activities and activities under other core programs in WIOA titles II–IV, as determined through the processes associated with paragraph (b)(1) of this section;
- (2) Conduct any type of evaluation similar to those authorized for, or conducted by, the Department of Labor or the Department of Education under the laws cited in paragraph (b)(4) of this section, including process and outcome studies, pilot and demonstration projects that have an evaluative component, analyses of administrative and programmatic data, impact and benefit-cost analyses, and use of rigorous designs to test the efficacy of various interventions; and
- (3) Conduct evaluations over multiple program years, involving multiple phases and such tasks and activities as necessary for an evaluation, such as a literature or evidence review, feasibility study, planning, research, coordination, design, data collection, analysis, and report preparation, clearance, and dissemination.

(f) In funding evaluations conducted under paragraph (a) of this section, States are permitted, but not required to:

- (1) Use funds from any WIOA title I–IV core program to conduct evaluations, as determined through the processes associated with paragraph (b)(1) of this section; and
- (2) Use or combine funds, consistent with Federal and State law, regulation and guidance, from other public or private sources, to conduct evaluations relating to activities under the WIOA title I–IV core programs. Such projects may include those funded by the Department of Labor and other Federal agencies, among other sources.

APPENDIX C: Evaluation Resources and References

The following items provide references to documents cited in the toolkit and additional resources for more in-depth understanding of evaluation concepts and analysis methods. Hyperlinks are provided, where available; and the resources are organized by broad-base categories. Other than documents specifically produced and used by and for the Department of Labor, it is important to note that DOL does not endorse other cited publications, resources and materials, or their authors.

1. National and Federal Policy, Planning, and Guidance Resources

Laws, Regulations, and Guidance

Workforce Innovation and Opportunity Act (WIOA) of 2014, PL 113-128:

For a copy of the law, see <https://www.congress.gov/113/bills/hr803/BILLS-113hr803enr.pdf>

For the final regulations, see <https://www.doleta.gov/wioa/about/final-rules/>

President's Executive Order on Expanding Apprenticeships in America, June 15, 2017.

Available at: <https://www.whitehouse.gov/presidential-actions/3245/>

U.S. DOL, Draft DOL Strategic Plan for 2018-2022. Available at:

<https://www.dol.gov/sites/dolgov/files/legacy-files/budget/2019/FY2018-2022StrategicPlan.pdf>

Office of Management and Budget (OMB). OMB M-17-28 on Fiscal Year 2019 Budget Guidance, July 7, 2017. Available at:

[whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2017/M-17-28.pdf](https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2017/M-17-28.pdf)

ETA Training and Employment Guidance Letter (TEGL) No. 14-15, March 4, 2016 on Workforce Innovation and Opportunity Act (WIOA) Requirements for Unified and Combined State Plans; and TEGL No. 06-17, January 24, 2018 on Modification Requirements for WIOA Unified and State Combined Plans. Available at: https://wdr.doleta.gov/directives/attach/TEGL/tegl_14-15.pdf
https://wdr.doleta.gov/directives/attach/TEGL/TEGL_6-17.pdf

Office of Management and Budget, Required Elements for Submission of the Unified or Combined State Plan and Plan Modifications under the Workforce Innovation and Opportunity Act, February 18, 2016. OMB Control No. 1205-0522. Available at:

https://www.doleta.gov/wioa/docs/Supporting_Statement_WIOA_State_Plans.pdf

Commission on Evidence-Based Policymaking. (September 2017). "The Promise of Evidence-Based Policymaking." Bi-partisan US Congressional Commission on Evidence-Based Policymaking, Washington, DC. Available at: <https://www.cep.gov/cep-final-report.html>

Bipartisan Policy Center. (May 2017). "Congress and Evidence-Based Policymaking: Creating a 21st Century Legislature." Bipartisan Policy Center, Washington, DC. Available at: <https://bipartisanpolicy.org/library/congress-and-evidence-based-policymaking/>

Employment and Training Research, Assessments, Training, and Guides

Mathematica Policy Research, MPR. (March 2017). "Employment Research in Brief: An Annotated Bibliography of ETA-and CEO-Sponsored Studies 2012-2016." MPR for U.S. Department of Labor, Employment and Training Administration, Washington, DC. This document along with previous year bibliographies and other related documents are available at <https://www.doleta.gov/research/>

National Association of State Workforce Agencies, NASWA (February 2017). "Evidence-Building Capacity in State Workforce Agencies: Insights from a National Scan and Two State Site Visits." NASWA for US Department of Labor, Employment and Training Administration, Washington, DC. Available at: https://wdr.doleta.gov/research/FullText_Documents/ETAOP-2017-13_Evidence_Building_Capacity_in_State_Workforce_Agencies_Report.pdf

Employment and Training Administration. (February 2015 & December 2016). "Listening Session on States' Role in Evaluation and Research under WIOA, February 24, 2015, and Listening Session on Evaluation and Research: Building Capacity under WIOA." December 7, 2016." U.S. Department of Labor, Washington, DC. Available at: <https://www.workforcegps.org/events/2016/11/17/12/14/WIOA-Wednesday-Research-Evaluation-and-More> https://lmi.workforcegps.org/resources/2015/06/18/11/23/Research_and_Evaluation_for_Continuous_Improvement

Pew-MacArthur Results First Initiative. (November 2014). "Evidence-Based Policymaking: A Guide for Effective Government." The Pew Charitable Trusts, Washington, DC and the MacArthur Foundation, Chicago, IL. Available at: <http://www.pewtrusts.org/en/projects/pew-macarthur-results-first-initiative>

Pew-MacArthur Results First Initiative. (January 2017). "How States Engage in Evidence-Based Policymaking: A National Assessment." The Pew Charitable Trusts, Washington, DC and the MacArthur Foundation, Chicago, IL. Available at: <http://www.pewtrusts.org/en/projects/pew-macarthur-results-first-initiative>

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

2. Evaluation Resources and Online Toolkits

Guides, Handbooks, and Related References

Corporation for National and Community Service, Social Innovation Fund. (October 2016).

“Evaluation Reporting Guidance: Feasibility, Implementation and Impact Study Reports.”

Corporation for National and Community Service, Washington, DC. Available at:

https://www.nationalservice.gov/sites/default/files/resource/GP_SIF_Evaluation_Reporting_Guidance_0.pdf

National Academy for State Health Policy. (August 2015). “Determining the Impact of State Demonstrations: Considerations for State and Federal Policymakers.” National Academy for State Health Policy, Washington, DC. Available at:

<https://nashp.org/determining-the-impact-of-state-demonstrations-considerations-for-state-and-federal-policymakers/>

Coalition for Evidence-Based Policy. (October 2014). “Which Study Designs are Capable of Providing Valid Evidence About a Program’s Effectiveness: A Brief Overview.” Coalition for Evidence-Based Policy, Washington, DC. Available at:

<http://coalition4evidence.org/>

What Works Clearinghouse, Procedures and Standards Handbook (Version 2.1). Available at:

https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_1_standards_handbook.pdf

Corporation for National and Community Service, Social Innovation Fund. (2013). “Evaluation Plan Guidance: A Step-by-Step Guide to Designing a Rigorous Evaluation.” Corporation for National and Community Service, Washington, D.C. Available at:

<https://www.nationalservice.gov/sites/default/files/documents/SIF%20Evaluation%20guidance%208%205%202014.pdf>

U.S. Government Accountability Office, GAO (June 2013). “Program Evaluation: Strategies to Facilitate Agencies’ Use of Evaluation in Program Management and Policy Making, GAO-13-570.

GAO, Washington, DC. Available at: <http://www.gao.gov/assets/660/655518.pdf>

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://documents.worldbank.org/curated/en/650951468335456749/pdf/520990PUB0EPI1101OfficialUseOnly1.pdf>
wds.worldbank.org/external/default/WDSPContentServer/WDSP/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101OfficialUseOnly1.pdf

J.S. Wholey, H.P. Hatry, and K.E. Newcomer, eds. (2010). *Handbook of Practical Program Evaluation*. Jossey-Bass, San Francisco, CA.

P. Rossi, M. Lipsey, and H. Freeman (2003). *Evaluation: A Systematic Approach*. 7th Edition. Sage Publications, Thousand Oaks, CA.

Online Toolkits and Searchable Research and Evaluation Publication Resources and Collections

Community Toolbox. Chapter 36: Introduction to Evaluation. The Toolbox is a public service provided by the University of Kansas, Center for Community Health and Development. The toolbox is an online searchable toolkit of information, resources, tools, and related on multiple community topics. Chapter 36 addresses evaluations. Available at: <https://ctb.ku.edu/en/table-of-contents/evaluate/evaluation>

Better Evaluation. The Rainbow Framework. The evaluation framework is an online searchable toolkit on evaluation that includes information, resources, tools, and related on a full range of evaluation topics developed and maintained by an international collaboration with the core team based in Australia/New Zealand. Available at: <http://www.betterevaluation.org/>

U.S. Department of Labor, Chief Evaluation Office (CEO). Policy and Resources Webpages. The CEO provides a wealth of information, some of which was cited or used in the toolkit. In particular, see the policy page and resources page. Available at: <https://www.dol.gov/asp/evaluation/EvaluationPolicy.htm>
<https://www.dol.gov/asp/evaluation/Resources.htm>

U.S. Department of Labor, Employment and Training Administration (ETA). Research and Evaluation Webpage. ETA provides a range of evaluation resources, some of which were cited or used in the toolkit. See the main webpage for all that is available. Available at <https://www.doleta.gov/research/>

U.S. Department of Labor, Employment and Training Administration (ETA). WorkforceGPS. This technical assistance platform provides evaluation information, tools, guides, resources within primarily two collections (Innovation and Opportunity Network collection and the Workforce System Strategies Collection). The following two webpages provide access to evaluation resources. https://ion.workforcegps.org/resources/2015/12/07/18/37/Research_and_Evaluation_for_Continuous_Improvement and <https://strategies.workforcegps.org/resources>

3. Behavioral Insight Studies

M. Darling, et. al. (April 2017). "Practitioner's Playbook for Applying Behavioral Insights to Labor Programs." Mathematica Policy Research for US DOL. Princeton, NJ.

J. Lefkowitz, et. al. (May 2017). "Using Behavioral Insights to Strengthen Labor Programs Lessons Learned." Mathematica Policy Research for U.S. DOL. Princeton, NJ.

4. Cost Studies

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.

H. M. Lewin, E. Garcia, and J. Morgan. (2012). *Cost-Effectiveness of Accelerated Study in Associate Programs (ASAP) of the City University of New York (CUNY)*. Columbia University, New York.

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.

5. Data Analytics

See the Commission on Evidence-Based Policy report listed above, and other reports and white papers on the CEP website. Available at: www.cep.org

M. Zook, et.al. (March 2017). "Ten Simple Rules for Responsible Big Data Research." *PLOS Computational Biology*, 13(3):e1005399. Available at: <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399>

I. Foster, et al., Editors (2016). *Big Data and Social Science: A Practical Guide to Methods and Tools*. Chapman and Hall/CRC Press.

E. Johnson. (August 2016). "Can Big Data Save Labor Market Information Systems?, a Policy Brief" Publication No. PB-0010RTI Press, North Carolina. Available at: https://www.rti.org/sites/default/files/resources/Can_Big_Data_Save_Labor_Market_Info_Systems.pdf

R. Connelly, et al. (April 2016). "The Role of Administrative Data in the Big Data Revolution in Social Science Research." *Social Science Research*, 59 (2016) 1-2. Available at: <https://www.sciencedirect.com/science/article/pii/S0049089X1630206X>

6. Implementation Studies

The National Implementation Research Network (NIRN) is a center that supports research and prepares resources on measuring implementation. Available at: www.fpg.unc.edu/~nirn.

J. Century, M. Rudnick, and C. Freeman. (2010). "A Framework for Measuring Fidelity of Implementation: A Foundation for Shared Language and Accumulation of Knowledge." *American Journal of Evaluation* 31:2.

A. Werner. (2004). *A Guide to Implementation Research*. The Urban Institute Press, Washington, D.C.

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." *Evaluation Review* 7:5.

7. Interrupted Time Series

D. Ho, K. Imai, G. King, and E.A. Stuart. (2007). "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Policy Analysis*, 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 CaliforniaGAIN Program." *Journal of Labor Economics*, 24:2 (July):521-566.

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.

H.S. Bloom. (1999). "Estimating Program Impacts on Student Achievement Using 'Short' Interrupted Time Series." Manpower Demonstration Research Corporation, New York.

8. Logic Models

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.
- 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.
- Knowlton, L.W. and Phillips, CC. (2009). *The Logic Model Guidebook: Better Strategies for Great Results*. Los Angeles: Sage; 2009.
- 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.
- 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>
- McLaughlin, J., and Jordan, G. (1999). "Logic Models: A Tool for Telling Your Program's Performance Story." *Evaluating and Program Planning*, 22, 65-72.

9. Outcome Studies

- J. McDavid, I. Huse, and L. Hawthorn (2006). *Program Evaluation and Performance Measurement: An Introduction to Practice*. 2nd Edition, Sage Publications, Thousand Oaks, CA.
- 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. King, L. Morris, and C. Fitz-Gibbon. (1987). *How to Assess Program Implementation*. Sage Publications, Thousand Oaks, CA.

10. Power Analysis

S.W. Raudenbush. (2011). "Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]." Available at: www.wtgrantfoundation.org

S.W. Raudenbush, Martinez, and Spybrook (2007). "Strategies for Improving Precision in Group-Randomized Experiments." *Educational Evaluation and Policy Analysis*, 29(5).

Bloom, H. (2006). "The Core Analytics of Randomized Experiments for Social Research." Manpower Demonstration Research Corporation, New York. Available at: http://www.mdrc.org/sites/default/files/full_533.pdf

11. Quasi-Experimental Studies

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.

J.D. Singer and J.B. Willett. (2003) *Applied Longitudinal Data Analysis*. Oxford, U.K.: Oxford University Press.

W. Shadish, T. Cook, and D. Campbell. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 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.

D. Rubin and N. Thomas. (March 1996) "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52:1.

T. Cook and D. Campbell. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin, Boston, MA.

12. Randomized Controlled Trial (RCT) Studies

W. Alferes (2012). *Methods of Randomization in Experimental Design*. Sage Publications, Thousand Oaks, CA.

J.D. Singer and J.B. Willett. (2003) *Applied Longitudinal Data Analysis*. Oxford, U.K.: Oxford University Press.

Shadish, T. Cook, and D. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, Boston, MA.

T. Cook and D. Campbell (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin, Boston, MA.

13. Theory of Change

D. H. Taplin and H. Clark. (2012). “Theory of Change Basics.” ActKnowledge. New York, NY.

J. Reisman and A. Gienapp. (2004). “Theory of Change: A Practical Tool for Action, Results and Learning.” Organizational Research Services, Seattle, WA. Available at <http://www.aecf.org/upload/publicationfiles/cc2977k440.pdf>.

APPENDIX D: Example of the Learning Agenda Cycle

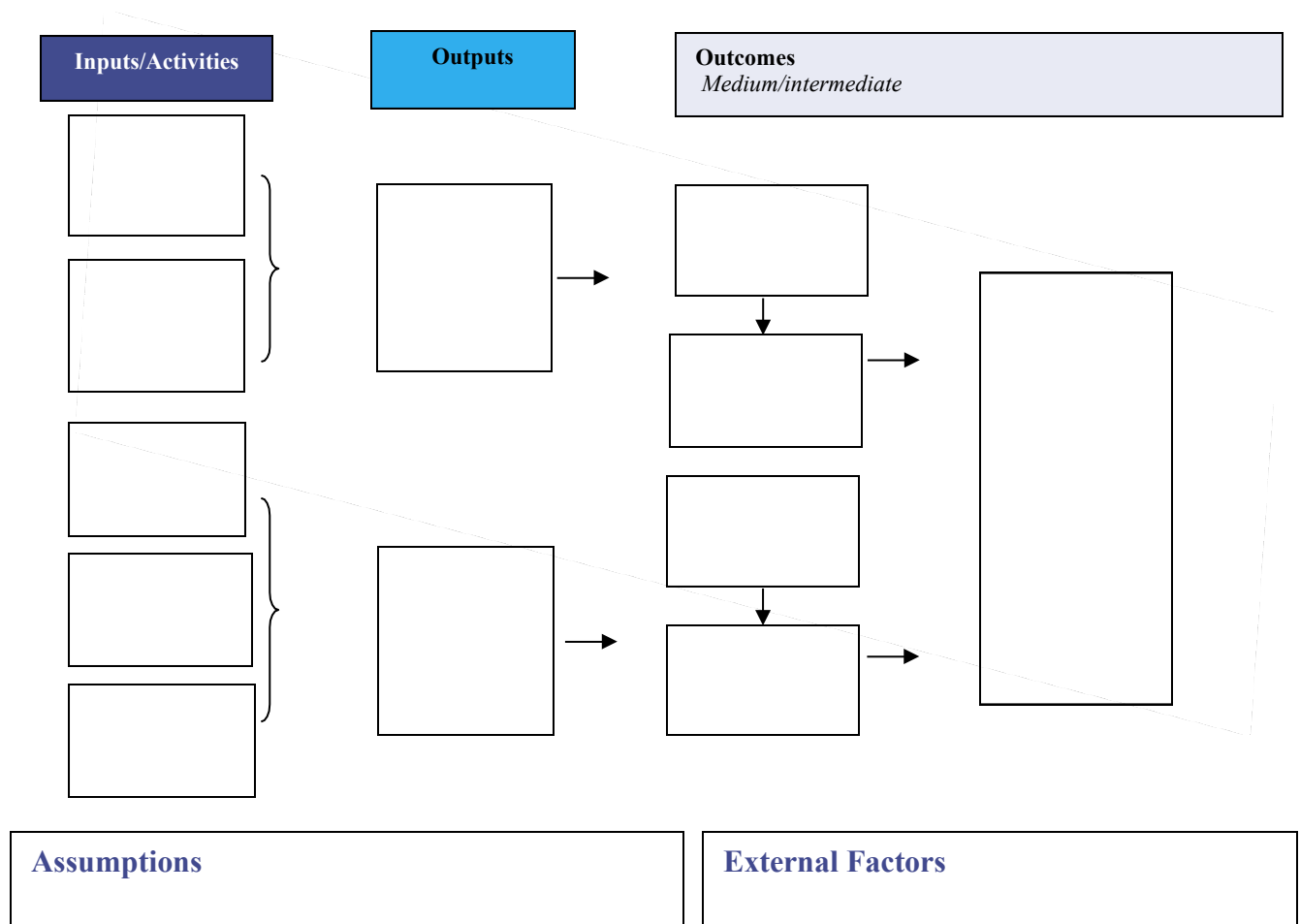


APPENDIX E: Logic Model Template

Program/service/interventions: _____ (name) _____ Logic Model

(Use text boxes to describe how a program, services, or series of interventions within the context of a given situation; add/change boxes and arrows, as needed)

Situation:



APPENDIX F: Evaluation Readiness Assessment—Overview and Instructions

1. Overview of Assessment

The purpose of the Evaluation Readiness Assessment (ERA) is to help State Workforce Agencies (SWA) and other entities gauge their overall readiness to conduct rigorous evaluations. Specifically, agencies can use the ERA to assess their current operational environment against evaluation readiness benchmarks and guidance derived from the DOL/ETA Evaluation Toolkit: Key Elements for State Workforce Agencies.

Agencies can use the assessment results to identify and explore factors inhibiting evaluation capacity and areas where additional resources and/or technical assistance (TA) are needed.

The ERA consists of the following five sections:

1. Evaluation Culture and Awareness;
2. Funding Strategies;
3. Data Management;
4. Staff Skills, Capacity, and Knowledge; and
5. Strategic Planning.

2. Instructions for Completing the Assessment

Each section contains an overarching question followed by a series of statements aligned to the evaluation topic of the section. Agencies should review each statement carefully and assign a rating on a scale of 1–5 to indicate the extent to which the statement is currently addressed by the agency. Each statement includes a reference to the appropriate section of the Evaluation Toolkit where more detailed information and guidance can be found.

After completing the assessment, the agency should implement the following next steps:

- Identify strengths and opportunities for improvement within each evaluation readiness topic area based on the assigned ratings;
- Hold facilitated meetings to review and discuss the assessment results as a team (e.g., state or local workforce board, WIOA implementation workgroup, or partner coalition); and
- Develop and implement an action plan to address any opportunities for improvement identified during the team discussions.

WIOA *Quick Start* Action Planner (QSAP)

Evaluation Readiness Assessment

Section 1		Evaluation Culture and Awareness									
		<i>Do agency staff and partners understand the benefits of evaluation, use evidence-based results to inform decisions, and plan to conduct evaluations to add to the existing evidence base?</i>									
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding						
Statement		Rating (Choose One)					Notes	Evaluation Toolkit References			
		1	2	3	4	5					
1. Agency staff and partners are familiar with available resources for evidence-based research and evaluation and regularly review recent reports to inform decisions.		<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>		Sections 1.1, 1.2, 1.3, 2.1 Pages 1–13			
2. Agency staff work strategically to cultivate cross-agency relationships and support for evaluation from the Governor’s office, State Workforce Boards, agency heads, and State Legislative staff.		<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>		Sections 1.1, 1.2, 1.3, 2.1 Pages 1–13			
3. The agency promotes partnerships with universities, foundations, or other entities that have the capacity to conduct evaluation.		<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>		Sections 1.1, 1.2, 1.3, 2.1 Pages 1–13			
4. When planning to implement new or revamped programs and services, agency staff and partners regularly consider effective evaluation strategies.		<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>		Sections 1.1, 1.2, 1.3, 2.1 Pages 1–13			

Section 2		Funding Strategies					
		<i>Does the state or region actively pursue funding for and invest in evaluations?</i>					
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding		
Statement	Rating (Choose One)					Notes	WIOA Evaluation Toolkit References
	1	2	3	4	5		
1. The state or region uses the Governor’s statewide set-aside funds strategically to conduct evaluations of Title I core programs, as required by WIOA. ¹	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 2.1 Pages 9–13
2. The state or region uses (or has used) discretionary grants from the DOL Workforce Data Quality Initiative (WDQI) or ED State Longitudinal Data System (SLDS) to develop data infrastructure.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 2.1 Pages 9–13
3. The state or region pursues additional funding for evaluation through competitive grant programs administered by DOL and other agencies.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 2.1 Pages 9–13
4. The state or region incorporates evaluation requirements into funding opportunity and competitive procurement requirements.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 2.1 Pages 9–13

¹ As required by § 682.200(d), States must use funds reserved by the Governor for statewide activities to conduct evaluations of activities under the WIOA title I core programs in order to promote continuous improvement, research and test innovative services and strategies, and achieve high levels of performance and outcomes.

Section 3		Data Management					
		<i>Does the agency have adequate operational capacity, IT infrastructure, and policies and procedures for collecting and using data for evaluations?</i>					
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding		
Statement	Rating (Choose One)					Notes	WIOA Evaluation Toolkit References
	1	2	3	4	5		
1. The agency has access to cross-agency longitudinal administrative data that cover a range of public programs, including Unemployment Insurance (UI) wage record data.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1 Pages 9–13
2. There is a centralized entity in the state or region which maintains data across agencies, and there are clear, streamlined procedures for processing data requests.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1 Pages 9–13
3. The state or region has data sharing agreements to facilitate interstate exchange of UI wage record data for both WIOA reporting and evaluation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1 Pages 9–13
4. The agency leverages other federal data collection efforts to support evaluation activities. ²	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1 Pages 9–13

² Examples include the Integrated Post-Secondary Education Data System (IPEDS), Federal Statistical Research Data Centers (FSRDCs), and the Center for Administrative Records Research and Applications (CARRA).

Section 4		Staff Skills, Capacity, and Knowledge					
		Does the agency have sufficient staff with the skills, knowledge, and experience needed to conduct or oversee third-party evaluations?					
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding		
Statement	Rating (Choose One)					Notes	WIOA Evaluation Toolkit References
	1	2	3	4	5		
1. The agency has an in-house research and evaluation unit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1; Pages 9–13 Section 5.1, 5.2 Pages 43–45
2. In the absence of an in-house research and evaluation unit, the agency has designated—and allocated resources to—staff to lead evaluation activities internally or to serve as liaison to a third-party evaluator.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 5.1, 5.2 Pages 43–45
3. Agency staff are familiar with the major evaluation types and understand the relative advantages and disadvantages of each.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 3.1, 3.2 Pages 17–24
4. Agency staff know how to select the most appropriate evaluation design based on the primary research questions and other practical considerations.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 3.1, 3.2 Pages 17–24
5. Agency staff know how to conduct market research to identify third-party evaluators with the necessary qualifications, capabilities, and experience.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 5.1, 5.2, 5.3, 5.4 Pages 43–47
6. Agency staff can identify and clearly articulate key personnel requirements, such as education level, evaluation experience, and subject matter expertise.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 5.1, 5.2, 5.3, 5.4 Pages 43–47
7. Agency staff can develop reliable labor effort estimates that accurately reflect the scope of the evaluation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 5.1, 5.2, 5.3, 5.4 Pages 43–47

Section 5		Strategic Planning						
		<i>Does the state have a comprehensive strategic plan that includes evaluation as an integral part of the plan?</i>						
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding			
Statement		Rating (Choose One)					Notes	WIOA Evaluation Toolkit References
		1	2	3	4	5		
1. The state has a process for involving all key agencies in the development of a plan that includes evaluation goals, study priorities, funding mechanisms, and roles/responsibilities.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.2 Pages 14–16
2. The state has developed an ‘evidence portfolio’ on particular subject areas of interest. This evidence portfolio includes evidence reviews and descriptive research using existing data.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.2 Pages 14–16

WIOA *Quick Start* Action Planner (QSAP)

Evaluation Design and Implementation Assessment

Section 1		Evaluation Design and Research Questions									
		<i>Has the agency been thorough in the selection and development of the evaluation design and research questions?</i>									
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding						
Statement		Rating (Choose One)					Notes	Evaluation Toolkit References			
		1	2	3	4	5					
1. The agency has developed a logic model that clearly illustrates the theory of change—from program inputs to intended long-term outcomes.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 4.1 Pages 28–32			
2. The agency has performed a thorough evidence review, related to the subject of the evaluation, to identify key research questions and how the results will build upon existing evidence.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 4.3, 4.4 Pages 34–37			
3. The agency has selected discrete, specific, measurable, and answerable research questions based on the theory of change.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 4.3 Pages 34–35			
4. The agency has engaged partners and key stakeholders to obtain input on, and endorsement of, the evaluation design.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 2.1, 2.2 Pages 9–16 Section 4.2 Pages 33–34			
5. The agency has selected the most rigorous evaluation design that is feasible for answering the primary research questions.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 4.5 Pages 38–40			
6. The agency has considered implementing a small pilot evaluation and assessing the results before moving forward with the full-scale evaluation.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 3.2 Pages 24			
7. The agency has developed a detailed evaluation timeline that is realistic and accounts for all critical evaluation activities.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 3.3 Pages 24–25			

Section 2		Data Collection and Analysis Plan					
		<i>Has the agency identified the necessary data sources and developed a comprehensive data analysis plan with detailed descriptions of each step in the process?</i>					
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding		
Statement	Rating (Choose One)					Notes	Evaluation Toolkit References
	1	2	3	4	5		
1. The agency has the ability to house, transmit, and secure the data to be collected.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2, 6.3 Pages 48–54
2. The agency has identified the specific data sources and data elements required for calculating evaluation outcome or impact measures.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2, 6.3 Pages 48–54
3. The agency currently has access, or has verified its ability to obtain access, to all required administrative data sources.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2, 6.3 Pages 48–54
4. If applicable, the agency has developed a plan for primary data collection, including procedures for secure storage and transmittal of personally identifiable information (PII).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2, 6.3 Pages 48–54
5. The analysis plan clearly describes how the data sources and individual variables will be used to construct outcome or impact measures.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2 Pages 48–50
6. The analysis plan includes a description of the study population and a statistically sound sampling plan.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.1, 6.2 Pages 48–50

Section 3		Evaluator Selection						
		<i>Has the agency developed a solid plan for identifying an evaluator with the qualifications and experience required to successfully implement the evaluation?</i>						
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding			
Statement		Rating (Choose One)					Notes	Evaluation Toolkit References
		1	2	3	4	5		
1. The agency has carefully considered the pros and cons of using an in-house, university, other partner, or third-party evaluator.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Sections 5.1, 5.2, 5.3, 5.4 Pages 43–47
2. For third-party evaluation, the agency has crafted a clear RFP that outlines the purpose, objectives, and requirements of the evaluation and the criteria for selection.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Sections 5.1, 5.2, 5.3, 5.4 Pages 43–47
3. For third-party evaluation, the agency has established a proposal review committee comprising relevant subject matter experts and other staff, as appropriate.		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Sections 5.1, 5.2, 5.3, 5.4 Pages 43–47

Section 4		Participant Rights						
		<i>Have the agency and evaluator developed careful processes to protect the privacy of the study participants?</i>						
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding			
Statement		Rating (Choose One)					Notes	Evaluation Toolkit References
		1	2	3	4	5		
1. The evaluator has created a specific plan for protecting the privacy of participants, including having a secure IT system to transfer PII data safely.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 6.3; Pages 51–54
2. The evaluator has submitted and received approval from an Institutional Review Board (IRB).		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		Section 6.3; Pages 51–54

Section 5		Reporting					
		<i>Has the agency developed a plan for disseminating evaluation results to different audiences?</i>					
Ratings:	1—Not at all	2—Making progress, but a long way to go	3—Have some of this, sometimes	4—Yes, in place now	5—In place and exceeding		
Statement	Rating (Choose One)					Notes	Evaluation Toolkit References
	1	2	3	4	5		
1. The evaluator and agency have agreed upon deliverables designed to disseminate interim and final evaluation results to different target audiences.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.5 Pages 56–57
2. The agency has scheduled meetings with legislators and other stakeholders to report the evaluation results to ensure they are used to improve programs and services.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.5 Pages 56–57
3. The agency has developed a plan for making the final evaluation report and results publicly available.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.5 Pages 56–57
4. If applicable, the agency has developed a plan for creating and disseminating a public-use data file for the evaluation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		Section 6.5 Pages 56–57

APPENDIX G: Glossary of Terms

The glossary contains definitions for common evaluation terms and concepts. The glossary is not comprehensive, but the terms and concepts included are some of the most frequently used in evaluation design, implementation, and analysis. Terms are in alphabetical order.

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: 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. The analysis includes documentation on program operational costs at the per-participant or per-system level and looks only at the costs of a program. In most cost analyses of employment and training programs, the analysis focuses on unit costs (e.g., per participant, enrollee, or FTE position).

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.

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 (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.

Net Cost: This is the gross cost of the program minus the cost of providing comparable services to the control or comparison group with which the impact estimates are made. Typically, the control or comparison group receives fewer or less intensive services than program participants, but the cost of comparable services is not zero. Thus, the net cost is less than the gross cost.

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. A power calculation is a calculation that estimates, given a specific sample size and analysis design, how likely it is that a program effect will be significant.

Pre-Post Data Analysis: A type of outcome 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 (QE) Study: 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 (RCT) Study: 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 the underlying understanding of the issue the SWA is addressing—it clarifies why evaluators are doing what they 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.