Reemployment Services and Eligibility Assessment (RESEA) Evaluation Toolkit

KEY ELEMENTS FOR STATE RESEA PROGRAMS

RESEA

JANUARY 2021

Abt ASSOCIATES

URBAN INSTITUTE

NASWA NATIONAL ASSOCIATION OF STATE WORKFORCE AGENCIES

About This Toolkit

This toolkit provides a basic overview of evaluation elements for program management purposes. It also draws from and provides information about other generally accepted and available evaluation resources that may be useful as an entry point for state RESEA programs that may not make full use of evaluations in program planning or implementation or that may need additional evaluation reference to expand their evaluation activities.

Acknowledgements: Abt Associates, The Urban Institute, and Capital Research Corporation, with support from staff at the U.S. Department of Labor's Chief Evaluation Office, Employment and Training Administration, Office of Policy Development and Research and Office of Unemployment Insurance, developed this toolkit under Contract #1605DC-18-A-0037. Gay Gilbert, Megan Lizik, and Gloria Salas-Kos offered feedback, reviewed, edited, and provided evaluation reference resources and materials. This toolkit was adapted from the *WIOA Evaluation Toolkit* developed by Coffey Consulting and *The Evaluation Toolkit for Prospective Workforce Innovation Fund Grantees, Parts I and II* developed by Abt Associates.

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, or 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 DOL/Office of Unemployment Insurance requirements.

Suggested Citation: Mills De La Rosa, S., Souvanna, P., Clarkwest, A., Kappil, T., Epstein, Z., Rothschild, L., Kuehn, D., Wall, A., Klerman, J., and Nightingale, D. (2021) *Reemployment Services and Eligibility Assessment (RESEA) Evaluation Toolkit: Key Elements for State RESEA Programs*. US Department of Labor.

Contents

1. Evaluating RESEA 1
- 1.1 What Is Evaluation? 1
- 1.2 Why Evaluate? 3
- 1.3 How This Toolkit Can Help States 3

2. Preparing for an RESEA Evaluation 6
- 2.1 Understanding High-Quality Evaluation Standards 6
  - DOL's Evaluation Policy Statement 6
  - CLEAR's Study Review Guidelines and Ratings 7
  - CLEAR's Causal Evidence Guidelines and Ratings 8
  - CLEAR's Descriptive Study Review Guidelines 10
  - CLEAR's Implementation Study Review Guidelines 10
- 2.2 Assessing Internal Evaluation Capacity and Making Enhancements 10
- 2.3 Forming a Preliminary Evaluation Plan 12
  - Develop a Logic Model 13
  - Create RESEA Intervention Learning Agenda with Your Team 14
  - Craft Broad Research Questions 17
  - Develop an Evaluation Design 17

3. Identifying the Evaluation's Purpose, Research Questions, and Designs 18
- 3.1 Determine the Evaluation's Purpose and Define What to Test 18
- 3.2 Develop the Key Research Questions 19
- 3.3 Select From Three Major Types of Evaluations 21

4. Selecting an Evaluation Design: Impact Studies 23
- 4.1 Understanding Impact and the Counterfactual 23
- 4.2 Detecting Impacts – An Important Note about Sample Size 25
- 4.3 Experimental Studies 26
  - Logistics of Random Assignment 27
  - Point of Random Assignment 27
  - Unit of Assignment 28
  - Random Assignment Ratios 28
  - Monitoring Random Assignment and Service Provision 28
  - Sample Size 29
- 4.4 Quasi-Experimental Design Studies 31
  - Matching and Other Regression Methods 33
  - Regression Discontinuity 35
  - Interrupted Time Series 36
- 4.5 How Does a State Agency Decide Which Impact Evaluation Design Is Right for Its Intervention? 38
- 4.6 Combining Impact Studies with Other Study Types 39

5. Selecting an Evaluation Design: Outcome and Implementation Studies 41
- 5.1 Outcomes Studies 41
- 5.2 Implementation or Process Studies 43

6. Data Needs for RESEA Evaluations 47
6.1 Understanding Data Concepts 47
Types of Data 47
Sources of Data 48
Other Data-Related Concepts 50
6.2 How Can Data Be Used for Evaluation? 52
Basic Data Requirements for RESEA Impact Evaluation 53
Identifying Data Elements for an Evaluation 55
6.3 Considerations for Developing a Data Collection Plan 56
Data Availability 56
Data Access Requirements 56
Timing of Data Collection 56
Data Quality 57
Cost of Data Collection and Analysis 58
Data Security 58
6.4 Best Practices and Common Challenges 59
Best Practices 59
Challenges 60
7. Selecting an Evaluator 63
7.1 Defining Evaluator Qualifications and Skills 63
Evaluator Independence 63
What Skills Are Needed to Conduct an Evaluation? 64
7.2 Establishing the Role of State Workforce Agency Staff in Evaluation 65
7.3 Selecting an External Evaluator 66
Creating a Statement of Work (SOW) or Request for Proposal (RFP) 67
Funding the Evaluation 70
Creating an Evaluation Timeline 72
Reviewing Evaluator Experience and Qualifications 73
Soliciting Bids for an Evaluator 75
Partnering with Universities 76
8. Implementing the Evaluation 77
8.1 Fostering Collaborative Relationships with an Evaluator 77
8.2 Integrating the Evaluation into Existing Program Operations 78
Designing the Random Assignment Process 79
Training Staff on Evaluation Procedures 79
Developing Data Extracts from State Administrative Data 80
8.3 Creating an Evaluation Design Report 80
8.4 Protecting the Rights of Human Research Subjects 82
Human Subjects Protections 83
IRB Process Overview 84
8.5 Communicating Evaluation Results 88
Communicating the State Agency's Requirements for Reporting Study Findings 88
Types of Dissemination Activities 89

Appendices **91**
Appendix A. SSA-2018 Authorizing Legislation A-1
Appendix B. Unemployment Insurance Program Letter No. 1-20 ..... B-1
Appendix C. Reemployment Services and Eligibility Assessments (RESEA) Evaluation Resource List C-1
Appendix D. Logic Model Template D-1
Appendix E. Data Availability for Evaluations of RESEA Programs: A Tool to Identify Data Sources E-1
Appendix F. Evaluation Timeline Overview and Worksheet F-1
Appendix G. Designing & Launching a Random Assignment Evaluation of RESEA Programs G-1
Appendix H. Sample Random Assignment Flowchart H-1
Appendix I. Evaluation Design Report Template I-1
Appendix J. Glossary of Terms J-1

List of Exhibits

Exhibit 2-1. U.S. Department of Labor Evaluation Policy 7
Exhibit 2-2. CLEAR's Causal Evidence Ratings for Quality of an Impact Study's Evidence 8
Exhibit 2-3. CLEAR's Causal Evidence Ratings for Effectiveness of an RESEA Intervention 9
Exhibit 2-4. Sample Logic Model of a Generic RESEA Intervention 14
Exhibit 2-5. Important Terms in Identifying and Reviewing Research Evidence 16
Exhibit 3-1. Examples of Broad Research Questions 19
Exhibit 3-2. Example Specific Research Questions for RESEA Evaluations 20
Exhibit 4-1. Factors Influencing Sample Size Needs 25
Exhibit 4-2. Random Assignment Methods – Advantages and Challenges by Type 27
Exhibit 4-3. Sample Size Estimates for Tests Using Experimental Design, by Intervention and Outcome 30
Exhibit 4-4. Comparison of Retrospective and Prospective Studies 32
Exhibit 4-5. Questions to Consider When Selecting an Impact Design 32
Exhibit 4-6. Summary of Potential Quasi-Experimental Design Options for RESEA Evaluations 33
Exhibit 4-7. Generalized Approach for Matching 33
Exhibit 4-8. Visualization of Regression Discontinuity 35
Exhibit 4-9. Implementing Interrupted Times Series in Practice 37
Exhibit 4-10. Important Considerations for Selecting an Impact Study Design 38
Exhibit 4-11. Considerations for Assessing Quasi-Experimental Designs 39
Exhibit 4-12. Highest Possible CLEAR Causal Evidence Rating, by Study Design 39
Exhibit 5-1. Types of Program Evaluation – Outcomes Studies 45
Exhibit 5-2. Types of Program Evaluation – Implementation or Process Studies 46
Exhibit 6-1. Pros and Cons of Data Used for Evaluation, by Source 48
Exhibit 6-2. Common Data Elements Needed for RESEA Evaluations 55
Exhibit 7-1. Evaluator Skills Needed, by Evaluation Design Type 64
Exhibit 7-2. Evaluation Activities and Division of Responsibility 65
Exhibit 7-3. Factors Affecting an Evaluation Budget 68
Exhibit 7-4. Key Design Elements for Evaluations, by Type 69
Exhibit 7-5. Pros and Cons of Funding the Evaluation Incrementally 71
Exhibit 7-6. Pros and Cons of Funding the Evaluation with Optional Tasks 71
Exhibit 7-7. Sample Evaluation Activities and Timing 72
Exhibit 7-8. Illustrative Example of Impact Study Timeline 73
Exhibit 7-9. Evaluator Selection Questions and Evidence of Experience 74
Exhibit 8-1. Elements Included in the Evaluation Design Report 81
Exhibit 8-2. Typical IRB Responsibilities and Study Criteria for Obtaining IRB Approval 84
Exhibit D-1. A Sample RESEA Logic Model D-1
Exhibit D-2. Self-Assessment Logic Model Questions D-4
Exhibit F-1. Evaluation Phases, Tasks, and Activities F-3
Exhibit I-1. EDR Sections, by Study Design Type I-2

List of Boxes

What Is Impact? 2
What Is CLEAR? 7
Explore the RESEA Section of CLEAR's Website 10
Moving From Learning Goals to the Evaluation 15
What Is Impact? 21
CLEAR Study Review Guidelines 22
What is Impact? 23
Impact Study Groups 24
What Is the Counterfactual? 24
Key Terms 25
Options for Creating Sample Size Needed to Detect Impacts 26
A Checklist for Launching an Experimental Study 26
CLEAR's Criteria for Rating Random Assignment Studies 31
CLEAR's Criteria for Rating Matched Comparison Group Designs 35
CLEAR's Criteria for Rating Interrupted Time Series Designs 37
CLEAR's Descriptive Study Review Guidelines 42
CLEAR Implementation Study Review Guidelines 43
Basic Data Elements Required for RESEA Experimental Evaluations 54
Third-Party Evaluator Types and Sources 66
Practical Steps to Maintain Communication with Evaluators 78
Key Terms 82
Ethical Principles 83
Typical IRB Review Levels 85
Understand the Required Elements of Consent 86

List of Acronyms

AJC	American Job Center
CEO	Chief Evaluation Office
CLEAR	Clearinghouse for Labor Evaluation and Research
DOL	U.S. Department of Labor
EDR	Evaluation Design Report
ETA	Employment and Training Administration
EvalTA	evaluation technical assistance
FTP	file transfer protocol
IRB	Institutional Review Board
IT	information technology
LMI	labor market information
NASWA	National Association of State Workforce Agencies
OUI	Office of Unemployment Insurance
PL	Public Law
PII	personally identifiable information
QED	quasi-experimental design
RCT	randomized controlled trial
REA	Reemployment and Eligibility Assessment
RESEA	Reemployment Services and Eligibility Assessment
RFP	request for proposal
SOW	statement of work
SSA	Social Security Act (PL 115-123)
SSN	Social Security number
UI	Unemployment Insurance
WDB	Workforce Development Board
WIOA	Workforce Innovation and Opportunity Act (PL 113-128)

1. Evaluating RESEA

Federal and state governments have increasingly emphasized the use of rigorous evidence to inform policymaking and program design decisions to better serve participants. For example, the Foundations for Evidence-Based Policymaking Act of 2018 (hereafter "Evidence Act") requires all federal agencies to develop evidence-building agendas and evaluation plans. In this spirit, the Unemployment Insurance (UI) system is enhancing its focus on conducting evaluations to steadily build evidence that will ultimately give insight into what activities and policies improve UI claimants' economic outcomes. Most recently, this emphasis is seen with the introduction of evidence and evaluation requirements as part of the 2018 amendments to the Social Security Act which permanently authorized the Reemployment Services and Eligibility Assessment (RESEA) program.[1] State agencies that operate the RESEA program play a major role in developing and disseminating such evidence.

This toolkit is designed to help build state agencies' capacity to plan and conduct high-quality evaluations of RESEA programs. It provides tips to help states from the earliest stages of developing evidence-building goals through planning and carrying out various types of evaluations. State agencies vary in how much experience they have with evaluation, both of UI-related programs and more broadly. The aim of this toolkit is to support this range of expertise by focusing primarily on evaluations of RESEA, a program that has its own specific evidence-building considerations and opportunities.

1.1 What Is Evaluation?

Evaluation is a key method for building evidence about interventions. The interventions may be a program as a whole or some component of that program. Although not all state agencies regularly conduct evaluations, state agencies regularly generate and report program results as part of performance reporting. Performance measurement is a management function that involves ongoing reporting of program activities and progress toward planned goals. That is one type of evidence-building. Evaluation goes beyond regular reporting of outcomes to generate stronger evidence regarding how a program or policy is functioning and what difference it is making.

As the Evidence Act[2] defines it:

> *The term 'evaluation' means an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency.*

The particular type of assessment conducted for evaluations focus on questions of "how" and "why." For example, as the U.S. Office of Management and Budget describes:

> *Evaluations may address questions related to the implementation of a program, policy, or organization; the effectiveness of specific strategies related to or used by a program, policy, or organization; and/or factors that relate to variability in the effectiveness of a program, policy, or organization or strategies of these. Evaluations can also examine questions related to*

---

[1] Social Security Act, 42 U.S.C. § 306 (2018). https://www.govinfo.gov/content/pkg/PLAW-115publ123/html/PLAW-115publ123.htm.

[2] Foundations for Evidence-Based Policymaking Act of 2018, 5 U.S.C. §§ 301-315 (2018). https://www.govinfo.gov/content/pkg/PLAW-115publ435/html/PLAW-115publ435.htm.

understanding the contextual factors surrounding a program, as well as how to effectively target specific populations or groups for a particular intervention.[3]

Evaluations also differ from performance measurement reporting. Performance reporting is typically performed by administrators and staff in the relevant government agency on a regular basis. Evaluations are conducted more periodically and by an independent evaluator—perhaps in a different state agency or outside state government. Independent evaluators can provide a crucial external perspective. They may also have specialized skills in the methods required for the particular type of evaluation.

Different types of evaluations address different types of goals. Common types of evaluation include:

- **Impact study**, which assesses the impact of outcomes relative to a "counterfactual" situation (i.e., what those outcomes would have been in the absence of the intervention.) An impact study uses either experimental (i.e., RCTs) or quasi-experimental (e.g., regression discontinuity, interrupted time series, matched comparison group) design. An impact study answers the question, "how much of a difference has the intervention made?"

What Is Impact?

An intervention's impact is its effect on an outcome of interest. Impact is therefore the change in an outcome that is directly and solely attributable to the intervention rather than to other factors. That is, impact captures how program participants' outcomes differ from those that would have occurred in the absence of the intervention, with the assumption that everything else in their lives except for their participation remained as it was (aka "all else equal").

- **Outcomes study**, which measures the extent to which an intervention has achieved its intended outcome(s), focusing on outputs and outcomes relevant to effectiveness. Unlike an impact study, an outcomes study cannot show causal impacts. An outcomes study can help answer questions such as "How do program participants' outcomes match what was intended?" or "How do outcomes vary among different types of participants?"
- **Implementation or Process study** which assesses how an intervention is delivered relative to its intended theory of change; it often includes information on content, quantity, quality, and structure of services provided. An implementation study does not usually measure outcomes. This toolkit uses the terms "implementation study" and "process study" interchangeably. An implementation study includes not only studying the implementation of an intervention but also aspects related to implementation, such as a study about customer flows. An implementation study can be conducted on its own but is often conducted along with impact and/or outcomes studies. An implementation study can help answer questions such as "was the intervention implemented as intended?" or "how is the intervention operating in practice?"[4]

Impact evaluations produce the type of evidence required of RESEA interventions by the 2018 SSA amendments. Outcome and implementation evaluations can be useful to help prepare for or to otherwise inform impact evaluations. An implementation study can also be a valuable complement to an impact study. An implementation study can help evaluators understand and interpret impact and outcomes study findings by providing descriptive information about how a program operates as well as about the context in which the program operates. For example, the

[3] Office of Management and Budget. *Circular No. A-11: Preparation, Submission, and Execution of the Budget*, December 2019, p. 200-16. https://www.whitehouse.gov/wp-content/uploads/2018/06/a11.pdf.

[4] IBID

statistical analysis from an impact study might suggest that the RESEA intervention has no net effect on claimants' employment outcomes (compared to what would likely have happened without the intervention). In analyzing staff interviews and observations from the implementation study, the evaluator might learn that there were significant differences in how program staff implemented critical program activities or components. The evaluator would then be able to inform some of the findings or patterns that emerge from the impact analysis. Program and service delivery details documented in implementation analysis can also provide more operational context for the impact analysis.

1.2 Why Evaluate?

Evaluation is a key part of a continuous quality improvement process that produces benefits for state agencies and the claimants they serve. Evaluations—like other forms of research and data analytics—can shed valuable light on questions of interest to policymakers as well as program designers and administrators about how to best serve RESEA-eligible claimants. For example, in their responses to a 2018 NASWA "information exchange"[5] questionnaire about RESEA program intervention design, requirements, and evidence, states expressed interest in evidence on:

- Claimants' employment outcomes, to show that services that claimants say are beneficial to them truly are having a positive impact.
- Which claimants to target for RESEA selection to enhance program effectiveness.
- Which elements of an RESEA program are working or not.

Similarly, many states responded to a 2017 NASWA questionnaire asking them to name the "most pressing questions" to understand named program outcomes or impacts.[6] NASWA's summary of responses notes that many state agencies focus their priorities on "understanding customers and their barriers" or "improving program operations and administration." Evaluation findings can help address each of these kinds of learning goals. Gathering this type of insight benefits the individual state agency's program and contributes to a larger body of evidence-based practices that directly supports the broader employment security community. It was with these benefits in mind that Congress and DOL established expectations for state agencies to conduct evaluations of their RESEA programs.

1.3 How This Toolkit Can Help States

State agencies vary in how much evaluation experience and capacity they have, and they will likely vary in how that experience and capacity evolves over time. This toolkit aims to help these state agencies, whatever their current level of expertise. It covers everything from the most basic level of considering what to evaluate, to more technical considerations of evaluation designs to choose from, through practical tips for evaluation planning and implementation. Though evaluations are typically conducted by professional, independent evaluators, this toolkit provides useful information about key aspects of evaluation to help state agencies as they expand their use of and partnerships with evaluators.

---

[5] Yvette Chocolaad, Cynthia Forland, and Julie Squire. 2018 October. "Reemployment Initiatives: Balancing Federal Requirements and State Ingenuity." Internal Report. National Association of State Workforce Agencies.

[6] Yvette Chocolaad and Stephen Wandner. 2017. *Evidence-Building Capacity in State Workforce Agencies*. Washington, DC: National Association of State Workforce Agencies. https://wdr.doleta.gov/research/FullText_Documents/ETAOP-2017-13_Evidence_Building_Capacity_in_State_Workforce_Agencies_Report.pdf.

DOL already offers a broad array of evaluation findings and resources through its Evaluation and Research hub on WorkforceGPS.[7] These offerings are valuable resources. However, each state agency's evaluation of RESEA will involve its own particular set of considerations that result from the intervention's design and its evidence requirements. Those considerations influence the types of evaluations that state agencies are likely to conduct and which particular study designs might or might not be feasible. For example, the RESEA evidence requirements place particular emphasis on impact evaluation. The unique manner in which RESEA programs select claimants makes certain that impact evaluation designs are more readily feasible than others. This toolkit takes into account these types of considerations and provides evaluation tips and resources tailored to evaluations of RESEA.

The toolkit is organized as follows:

Chapter 2 helps state agencies lay the foundation for high-quality evaluation. The chapter provides sources to better understand the features of a high-quality evaluation, how state agencies can assess and build their readiness for leading an RESEA evaluation, and how they can begin to develop pieces of an RESEA research agenda.

Chapter 3 provides suggestions to develop specific research questions to address a state agency's RESEA evidence-building goals. The chapter also explains which types of evaluation are appropriate to answer which types of research questions.

Chapter 4 focuses on impact studies. Those are the types of evaluations that can produce evidence that demonstrates that an intervention is effective or not. The chapter explains what "impact" means and how measuring impact differs from measuring outcomes. It also describes how sample size influences whether or not an impact study will be able to detect impacts. Finally, the chapter describes different types of impact study designs, including their strengths and limitations.

Chapter 5 focuses on outcome studies and process studies, describing what each involves and how state agencies can use them to build evidence. Outcomes and process studies cannot demonstrate an intervention's impact, but as the chapter shows, they might help state agencies prepare for an impact study or can complement an impact study.

Chapter 6 explains which types of data are likely to be needed for different kinds of RESEA evaluations, as well as where state agencies and their evaluators may be able to find the data that they need. The chapter also provides tips for obtaining, assessing, and protecting data.

Chapter 7 describes the types of experience and qualifications that are required of an independent evaluator, depending on the specific type of evaluation to be conducted. It also provides tips to help state agencies through each step of obtaining an evaluator, including through a procurement process. Those tips aim to help state agencies select appropriate evaluators and write a clear statement of work that will generate research outputs that meet the state agency's needs.

Chapter 8 will help state agencies anticipate and prepare for the logistical requirements and challenges involved in conducting a high-quality evaluation. It addresses specific logical issues, ranging from developing a collaborative working relationship with an evaluator to weaving study procedures into program operations to ensuring ethical standards are met to protect research participants.

---

[7] https://evalhub.workforcegps.org/

Throughout the toolkit, the key, light bulb, and bar graph icons are used to highlight important information:

- The key icon appears in call out boxes that contain definitions or explanations of important evaluation terms and concepts.
- The light bulb icon appears in call out boxes containing "tips" on things state agencies could do to facilitate their evaluations.
- The bar graph icon appears in call out boxes containing important information related to ratings from DOL's Clearinghouse for Labor Evaluation and Research (CLEAR).

2. Preparing for an RESEA Evaluation

State agencies' evaluations of their RESEA interventions (i.e., the whole program or aspects of it) are integral to expanding the evidence base of effective service delivery strategies. As such, it is important that state agencies' evaluations of their RESEA intervention meet standards for high-quality and rigor, regardless of the evaluation type (implementation, outcome, impact). All evaluations, when well-conducted, can produce information that is valuable for learning about RESEA intervention operations and inform RESEA intervention improvements.

This chapter discusses concrete steps to help state agencies organize their evaluation planning and lay a strong foundation for a high-quality evaluation:

- ✓ Become familiar with the **standards for high-quality evaluations**. These include U.S. DOL's Evaluation Policy Statement and the study guidelines (including the causal evidence ratings criteria) established by the Clearinghouse for Labor Evaluation and Research (CLEAR).
- ✓ Assess state agency's **internal evaluation capacity** and make enhancements where possible.
- ✓ Form a **preliminary evaluation plan**, which includes developing a logic model, creating RESEA intervention-specific learning goals, and crafting research questions.

By taking these steps, state agencies will be in a better position to develop a well-designed evaluation that meets their needs.

2.1 Understanding High-Quality Evaluation Standards

Conducting an evaluation that meets standards of rigor and quality is important for producing credible evidence regarding RESEA services and their effects. But, what constitutes a high-quality evaluation? Two important resources shed light on this key question and will help state agencies understand the characteristics of high-quality evaluations: (1) DOL's Evaluation Policy Statement and (2) CLEAR's study review guidelines, including those that describe how CLEAR assigns causal evidence ratings. These resources are reviewed below.

DOL's Evaluation Policy Statement

DOL's Evaluation Policy Statement contains principles for designing and conducting evaluations. Adhering to those core principles will help state agencies conduct RESEA evaluations that include valid and reliable data, can be made public, and protect the privacy of individuals who are subjects of the research (in this case, UI claimants). Developed by U.S. DOL's Chief Evaluation Office (DOL/CEO), the policy identifies the following principles: rigor, relevance, transparency, independence, and ethics (in human subjects protections), all of which are relevant to state agency-conducted evaluations. Exhibit 2-1 describes each principle.[8] They are similar to the principles of other federal agencies as well as common standards promulgated by such organizations as the American Evaluation Association.

[8] The policy can be found in its entirety at https://www.dol.gov/agencies/oasp/evaluation/EvaluationPolicy.

Exhibit 2-1. U.S. Department of Labor Evaluation Policy

Principle	Excerpt
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. 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 protections)."
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."

Source: https://www.dol.gov/asp/evaluation/EvaluationPolicy.htm

CLEAR's Study Review Guidelines and Ratings

CLEAR is a central source of research and information on labor-related topics. CLEAR is designed to make labor-related research accessible to a range of audiences, including practitioners, policymakers, researchers, and the public, in order to promote research-informed labor policies and programs. To do this, CLEAR provides systematic evidence reviews on labor topics and then summarizes study methodologies, findings, and policy or program implications. The reviews are guided by a protocol designed to capture all research papers and reports that examine the topic area's research questions and by guidelines appropriate to the study type.[9] This process is transparent, consistent, and of high quality across each evidence review or topic area. CLEAR has review guidelines for three categories of studies: causal, descriptive, and implementation. For state agencies, CLEAR's standards for high-quality labor research can serve as a model for evaluations of RESEA interventions. CLEAR's review guidelines for each category of studies are described below.

What Is CLEAR?

DOL's Clearinghouse for Labor Evaluation and Research is an online repository of high-quality research studies on labor topics. CLEAR identifies and summarizes many types of research. For impact studies it characterizes the strength of the causal evidence produced from such research.

CLEAR's standards for high-quality labor research can serve as a guidepost for states looking to embark on their own high-quality impact evaluations. For more information, see CLEAR's website: https://clear.dol.gov/.

[9] More information about CLEAR's review process can be found at: https://clear.dol.gov/reference-documents/policies-and-procedures-version-31.

CLEAR's Causal Evidence Guidelines and Ratings

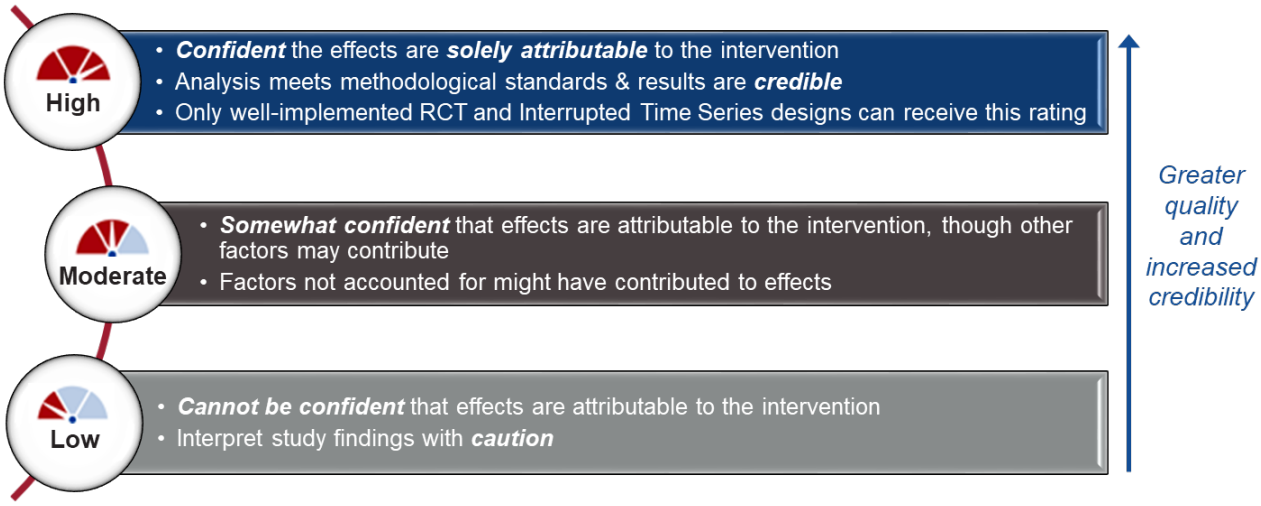
CLEAR uses the term "casual studies" to refer to experimental and non-experimental studies designed to measure the effectiveness (or "impact") of an intervention.[10] Specific study designs that fall within this category include randomized control trials (RCTs) and quasi-experimental designs (QED). CLEAR uses the causal evidence guideline to review and assess a causal study's quality and to assign a rating, which it refers to as a **causal evidence rating**. CLEAR then uses the causal evidence ratings of studies to determine a rating for the effectiveness of specific interventions, which it refers to as **intervention effectiveness ratings**.

CLEAR's Ratings of the Quality of a Study's Causal Evidence

CLEAR's causal evidence rating for a causal study indicates the credibility of that study's impact findings. CLEAR assigns the rating based on a systematic review by independent experts that examines the methods were used in the study and how well those methods were implemented. Those factors affect the credibility of the study's impact findings. In other words, CLEAR's causal evidence rating system indicates how confident one can be that a causal study's impact findings reflect the true effect of the intervention, rather than the effects of something else (or even random chance).

After reviewing a study, CLEAR assigns the study one of three casual evidence ratings: High, Moderate, or Low. CLEAR uses gas gauge icons to depict the causal evidence ratings. Exhibit 2-2 provides a summary of those ratings and the icons used. As shown, the higher the causal evidence rating, the more confident one can be that the study is technically sound—or of greater quality—and that its findings are credible. Conversely, the lower the rating, the more caution is needed when interpreting the study findings. Receiving a "Moderate" or "Low" causal evidence rating, however, does not mean that the study is not useful. For example, state agencies may find the study useful for helping them shape or modify their RESEA intervention strategy, implementation, or evaluation design, even if it is not possible to confidently conclude that the study's observed effects were solely caused by the intervention.

Exhibit 2-2. CLEAR's Causal Evidence Ratings for Quality of an Impact Study's Evidence



Source: Icons from https://clear.dol.gov/about

[10] More information about CLEAR's causal evidence guidelines can be found at: https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21.

CLEAR's causal evidence ratings also take into consideration the study's design. Only RCTs or interrupted time series (a type of QED) studies qualify to receive a "High" causal evidence rating. Quasi-experimental studies that use matched comparison, difference-in-difference, fixed effects, instrumental variable, or other regression methods can receive a "Moderate" causal evidence rating at best. Studies that use a pre-post, correlational, or other descriptive methods only qualify for "Low" causal evidence ratings.

It is crucial to note that CLEAR's causal evidence ratings for individual studies pertain to the credibility of a specific study's findings, not to the effectiveness of the intervention it studied (as discussed in the next section). That is, CLEAR's causal evidence ratings indicate how confident one can be about a given study's impact findings but do not indicate the direction or magnitude of those findings.

CLEAR's Ratings of the Effectiveness of an RESEA Intervention

In the RESEA section of the CLEAR website, CLEAR rates the effectiveness of specific RESEA interventions. These effectiveness ratings are based on the impact findings from all causal studies of those particular RESEA interventions that received a "High" or "Moderate" causal evidence rating.[11] CLEAR's causal evidence rating system on the effectiveness of RESEA interventions has four levels to describe the causal evidence: High, Moderate, Potentially Promising, and No Rating. On the CLEAR website, these effectiveness ratings are depicted with thermometer icons. Exhibit 2-3 provides a summary of those ratings and the icons that CLEAR uses to indicate them. A study is considered credible evidence for the purposes of rating an RESEA intervention's effectiveness if it earned a "High" or "Moderate" casual evidence rating from CLEAR.[12]

Exhibit 2-3. CLEAR's Causal Evidence Ratings for Effectiveness of an RESEA Intervention

High	At least two credible studies found favorable impacts on both employment and UI duration, with statistical significance at the 5% level.*
Moderate	At least one credible study found favorable impacts on employment and one credible study found favorable impacts on UI duration, with statistical significance at the 10% level.* The favorable impacts may come from a single study or two different studies.
Potentially Promising	At least one credible study found favorable impacts on employment or UI duration.
No Rating	No credible studies of the intervention exist, or no credible study has found favorable impacts on employment or UI duration

Source: Icons from https://clear.dol.gov/reemployment-services-and-eligibility-assessments-resea
* That is, less than a 5% or 10% chance, respectively, that findings are due to chance rather than to the intervention.

[11] CLEAR's RESEA page is found here: https://clear.dol.gov/reemployment-services-and-eligibility-assessments-resea.

[12] More details on the criteria for the intervention evidence rating can be found in Unemployment Insurance Program Letter No. 1-20: United States Department of Labor, Employment & Training Administration (October 2019). *Expectations for States Implementing the Reemployment Service and Eligibility Assessment (RESEA) Program Requirements for Conducting Evaluations and Building Program Evidence* (UIPL No. 01-20). Retrieved from https://wdr.doleta.gov/directives/attach/UIPL/UIPL_1-20_acc.pdf.

Explore the RESEA Section of CLEAR's Website

The wealth of labor-related research and evaluation information available in CLEAR may feel a little overwhelming for visitors using the website for the first time. A good place to start with finding RESEA-relevant evaluation evidence is the RESEA topic area tab.This section contains key resources, such as the Reemployment Synthesis report and links to CLEAR's profiles of causal studies of RESEA-relevant interventions that show impacts on employment or UI duration.

CLEAR's Descriptive Study Review Guidelines

CLEAR uses the term "descriptive study" to refer to "studies that use statistical techniques and other quantitative approaches but do not attempt to assess the causal impact of a program or policy."[13] Descriptive studies include outcomes studies which analyze observed characteristics of participants following receipt of an intervention and assesses those characteristics against program goals, across intervention implementations or locations, or over time. (See Chapter 5 for more details about outcomes studies.)

For descriptive studies, CLEAR reviews the appropriateness and reliability of the study design, data quality, data collection, study sample, and analytic methods for addressing the study's research questions. CLEAR examines whether the study's findings are adequately supported by the data and any conclusions derived based on the findings are reasonable. Unlike CLEAR's review of causal studies, CLEAR's review of descriptive studies do not culminate in a rating but are valuable for understanding the characteristics that contribute to a high-quality descriptive study.

CLEAR's Implementation Study Review Guidelines[14]

CLEAR's implementation study guidelines apply to studies that examine the development and operation of a program, policy, or intervention. An implementation study may focus on documenting how services are being delivered on the ground and/or assessing whether a program or intervention is being implemented as planned. (See Chapter 5 for more details about implementation studies.) Similar to descriptive studies, CLEAR's review of implementation studies does not result in a rating of study quality. Rather, CLEAR assesses the implementation study's technical qualities to determine whether the findings reported are appropriate for the study design and to identify the study's strengths and limitations. Specifically, CLEAR's guidelines for implementation studies assess the appropriateness of the study design, sampling strategy, data sources, data collection, and analysis for addressing the research questions. CLEAR also reviews whether the implementation study findings are aligned with the research questions and appropriate in relation to the study design and data.

2.2 Assessing Internal Evaluation Capacity and Making Enhancements

While an independent evaluator[15] will be responsible for designing and conducting the evaluation, state agency staff will play an important role in the evaluation's overall success by helping the

---

[13] Clearinghouse for Labor Evaluation and Research (July 2014). *Operational Guidelines for Quantitative Descriptive Studies.* Retrieved from https://clear.dol.gov/reference-documents/quantitative-descriptive-guidelines

[14] Clearinghouse for Labor Evaluation and Research (July 2014). *Operational Guidelines for Reviewing Implementation Studies.* Retrieved from https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0

[15] Evaluator independence is most readily achieved by selecting an evaluator from outside the state agency, though it may be feasible to achieve this with qualified staff or a research unit in the state agency. If a state agency decides to use

evaluator to understand RESEA intervention operations, access administrative data, and coordinate communication about the evaluation to key stakeholders. State agency staff also will review and provide constructive feedback on evaluation reports and other written deliverables produced by the evaluator to make sure they are well written and accurately describe the RESEA intervention and the evaluation. State agency staff, however, cannot ask the evaluator to make changes to reports and other deliverables in ways that would skew reporting about the evaluation (e.g., asking the evaluator to edit the report to make evaluation results seem more positive). Both the Social Security Act and DOL's Evaluation Policy require that the evaluator be independent, and specifically that its "evaluation functions [be insulated] from undue influence."[16] (See Chapter 7 for more information on the delineation of responsibilities between the state agency staff and the evaluator.)

State agency staff can effectively organize and oversee the evaluator's efforts by understanding the basics of evaluation design and research methods. Below are factors to consider in assessing state agency evaluation capacity to organize and oversee the evaluation:

- **Experience planning and successfully conducting evaluations.** Do state agency staff have experience with the designs the state agency is interested in pursuing? What technical experience and knowledge is relevant will depend on the evaluation type. For example, if a state agency plans to conduct an evaluation that compares differences in outcomes across different groups, does the state agency's staff have prior experience overseeing or conducting an evaluation using that design?
- **Experience collecting and/or analyzing quantitative or qualitative data.** Do any state agency staff have experience collecting and/or analyzing the type of data that the evaluation will require?
- **Educational background.** Does the state agency have staff with degrees in public policy, economics, program evaluation, or other relevant social science fields? Staff with these specialized degrees will typically have some academic training in research and evaluation methods. They may be comfortable leading or overseeing evaluation planning efforts.
- **Time and availability.** Do state agency staff with the right skills and expertise have time to oversee the evaluation, especially given their other duties?
- **Data quality procedures.** Does the state agency have established processes for ensuring consistent data entry, reviewing data regularly, and identifying improvements to data quality, from which staff can implicitly learn best practices for generating and using data?

If a state agency identifies gaps in its evaluation capacity, the state agency can take the following steps:

- **Enhance staff knowledge about evaluations** by encouraging participation in training opportunities and using tools provided by DOL's RESEA Evaluation Technical Assistance (EvalTA) team.[17]

---

an in-house evaluator, the state agency will want to ensure that the in-house evaluator is separate from the RESEA intervention and has not been involved in its development, management, or operations. More importantly, the in-house evaluator must be insulated in such a way that they are not subject to any perceived or real pressure to alter or misrepresent study findings. See Chapter 7 for more information about evaluator independence and selection.

[16] United States Department of Labor. (November 2013). *U.S. Department of Labor Evaluation Policy.* https://www.dol.gov/asp/evaluation/EvaluationPolicy.htm

[17] For the most up-to-date list of available RESEA EvalTA resources and tools, please refer to the "RESEA Evaluation and Evidence Resources" page on WorkforceGPS: https://rc.workforcegps.org/resources/2019/07/30/17/32/RESEA_Evaluation_Evidence_Resources

- **Identify staff with evaluation expertise elsewhere within the state agency.** Though immediate state agency staff in charge of the RESEA evaluation may not have research and evaluation experience or training, there might be staff elsewhere in the state agency with such expertise. These individuals can commonly be found in a state's labor market information, labor statistics, or research department. The state agency's human resources department may be able to help identify staff with the proper skillset.
- **Promote and prioritize staff use of data at all levels.** State agencies might consider implementing the following actions to enhance staff members' understanding of data and how it relates to their day-to-day work:
  - **Develop standardized data tools and procedures for data entry as well as expectations for their implementation.** Tools, such as a data codebook or manual, will help staff to better understand how concepts and measures are being defined and how to enter data accurately. Standardized procedures, such as entering data within 24 hours of meeting a claimant or entering data in the correct fields, can help set system-wide expectations about data entry. These tools can help staff with modest levels of data savvy be more effective contributors to research and strengthen their capabilities.
  - **Establish steps for checking data quality** to promote complete, accurate, and consistent data entry across staff and/or program sites. Creating site-level data completeness reports and dashboards that can be shared with site leadership for follow-up are two ways to help identify data quality problems. This ongoing use of data will help all involved staff become increasingly data savvy over time.
  - **Create opportunities for staff to understand the importance of data for their work, and if relevant, how the data they input can be used in their day-to-day work**. Frontline staff members are more likely to enter data accurately and completely if they are able to use the data as a part of their job and if they understand how the data will be used by others. State agency staff can work with American Job Center (AJC) leadership and frontline RESEA staff to train them and, if possible, learn more about what kind of data products could be created for staff to enhance their day-to-day work. By promoting motivation to use data, staff will involve themselves more and gradually build their skills.
  - **Strengthen data literacy by offering staff training on best practices.** Staff may not be aware of best practices in data collection and data access. Additionally, some staff may not be fully aware of how their data entry practices can affect an evaluation's ability to learn about the intervention and produce meaningful findings. Data literacy training will help staff better understand what high-quality data are, why high-quality data is important, and how to use data to inform their work. For state agencies that already have an annual data training, it may be possible to incorporate information that addresses these issues into their existing training.

2.3 Forming a Preliminary Evaluation Plan

Creating a preliminary evaluation plan is a key first step in designing and conducting an evaluation, regardless of the type of evaluation to be conducted. The preliminary evaluation plan will form the basis for the work that the state agency's independent evaluator will do. Once onboard, the independent evaluator will work with the state agency to finalize the evaluation details outlined in the preliminary evaluation plan. Most importantly, the preliminary plan will help foster an evaluation that will fully address the state agency's needs and interests.

Elements of the preliminary evaluation plan may include:

- A **logic model** that describes the RESEA intervention or aspects of the intervention.
- An RESEA-specific **learning agenda** that articulates what the state agency wants to learn about its RESEA intervention.
- Broad, high-level **research questions**.
- Possible **evaluation designs**, with consideration for sample size issues.

The preliminary evaluation plan will help develop a successful evaluation and identify what is required to design and implement it. For example, establishing clearly articulated research questions will help the evaluator determine what methods, data collection, and data analysis techniques to use to answer those questions.

Develop a Logic Model

In order to build a preliminary evaluation plan, state agencies must know *which* intervention will be evaluated—the whole program or specific components of it? *Why* that particular intervention? *How* the intervention potentially affects claimants' outcomes? And *what* resources and activities are required for the intervention to be implemented? A good starting point for answering these questions is creating a logic model.

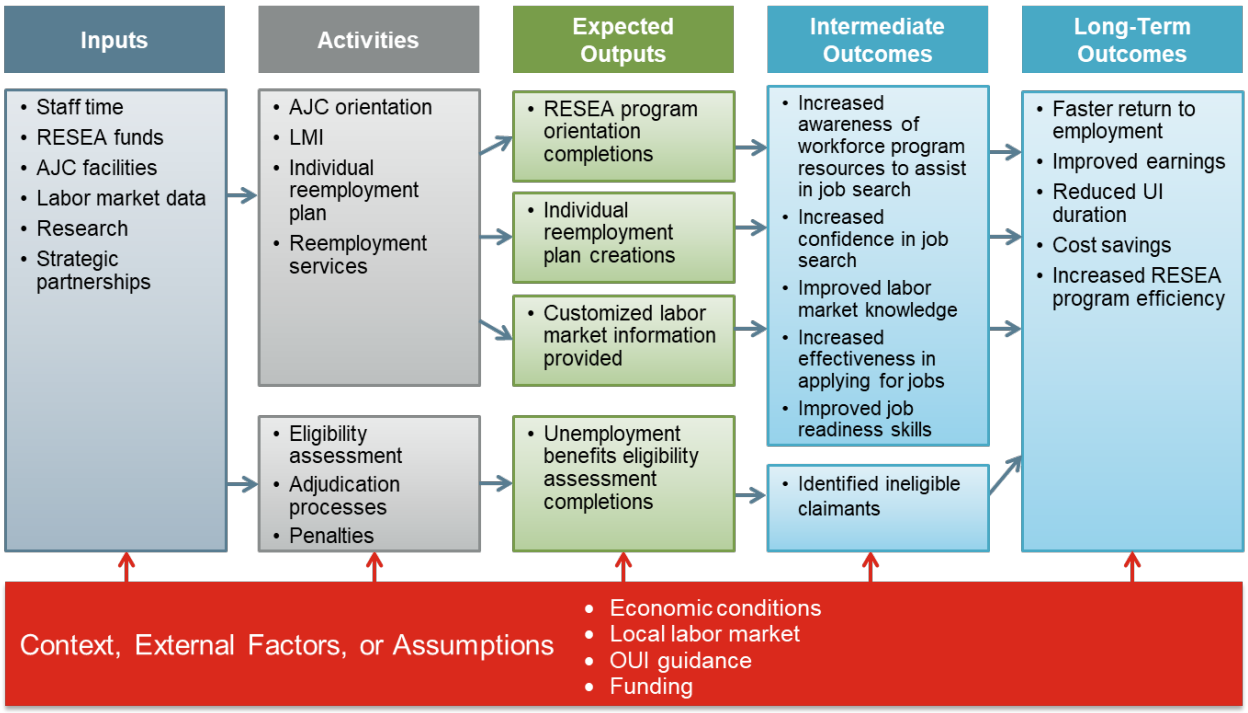
A logic model is a visual representation of the intervention's key elements, showing how the intervention elements are related to one another and work together to produce the desired results. A logic model is useful for describing program management operations; identifying inputs, activities, outputs, and outcomes used to measure success; and providing a blueprint for the implementation and evaluation of the intervention.

Typical elements in a logic model are:

- **Program inputs,** resources the state agency puts into operating the intervention, for example, time from UI and AJC workforce staff, funds by source, etc.
- **Program activities** or services delivered to claimants as part of the intervention, for example, eligibility assessments, RESEA meetings, etc.
- **Outputs** are products developed, deliverables completed, or milestones accomplished when claimants engage with RESEA intervention activities. For example, claimant attends the RESEA meeting and receives an orientation, claimant develops an individual reemployment plan, claimant receives notice that benefits are suspended due to noncompliance, etc.
- **Outcomes** are changes that result from the outputs, including changes in claimants' behavior, attitudes, aptitudes, skills and knowledge, or UI benefits:
  - **Intermediate (short-term) outcomes** are those that can lead to long-term outcomes; for example, claimant is more aware of job search resources, claimant is more confident about job search, or claimant is found to be ineligible for benefits.
  - **Long-term outcomes** are the ultimate objectives of the intervention; for example, claimant more quickly finds and keeps a job, UI benefits decrease in duration, or individual and family well-being increases.
- **Context, External Factors, Assumptions** are factors that may influence the intermediate- and long-term outcomes, for example, economic conditions, local labor market, Office of Unemployment Insurance guidance, and/or funding.

Exhibit 2-4 provides an illustration of what a logic model may look like for a generic RESEA intervention. Appendix D describes how to create a logic model and provides a template.

Exhibit 2-4. Sample Logic Model of a Generic RESEA Intervention



Create RESEA Intervention Learning Agenda with Your Team

Before moving forward with an evaluation, the state agency will want to determine the key information it aims to learn about its RESEA intervention. Development of a RESEA "learning" agenda can lay out specific evaluation activities that support the state agency's priorities. The process of developing the learning agenda involves understanding and building consensus around broad learning goals, as well as reviewing the existing evidence base.

Establish Broad Learning Goals

Learning goals specify the areas in which the state agency is interested in generating knowledge and program insight. To identify these learning goals, it is useful to address questions that include:

- What do we know about how well our RESEA intervention helps to reduce UI claim duration and improve employment outcomes? What aspects of the RESEA intervention do we think are important in driving outcomes and in what ways? What innovations might we be interested in testing?
- How well do we understand the RESEA intervention model and how it is being implemented across the state? When did we last make updates to the intervention model? How long has it been in a steady state? Where do we want to make improvements?
- What do we still need to learn about our RESEA intervention? How could we begin to gather that information?
- What specific elements of our intervention do we want to learn more about in the next one to two years? In the next three to five years?

In developing these learning goals, state agency staff may find it helpful to engage key stakeholders who work with RESEA, such as:

- **Frontline RESEA staff and AJC leadership** who can help highlight issues that are most pressing to those who deliver services to claimants.
- **State agency leadership and other staff** who can help align RESEA research priorities with other learning efforts across the state.
- **Researchers** who can help determine what has already been learned through previous research and what new questions might be interesting for the state agency and the field.

Seeking input from these sources can help identify a wider range of research questions and help build a richer set of learning goals.

Moving From Learning Goals to the Evaluation

The time spent up front on creating learning goals will pay off when trying to determine the scope and direction of the evaluation. For example, with increased flexibility in how claimants are selected for RESEA services, a state agency might set the following as its learning goal: understanding whether the state agency's selection methods result in identification of claimants who can be best helped by the RESEA intervention. There are many ways the state agency could gather information to learn more about this example learning goal. For example, the state agency could hold discussions with local AJC staff about how different types of claimants seem to respond to different services. Also, the state agency might want to conduct an evaluation that examines how the impacts of its RESEA intervention differ across groups of claimants.

Review Existing Evidence Base

Reviewing the existing evidence base can help state agencies to further develop a well-informed learning agenda. The evidence base includes the outputs from prior evaluations that document the intervention's implementation, how that intervention has been systematically evaluated, the findings related to claimants' outcomes, and how the intervention affected those outcomes. Reviewing existing research will help provide useful information about gaps in the current evidence base that new evaluations could help fill. For instance, the state agency's RESEA intervention may include a service strategy that has not been evaluated. For that reason, learning more about that particular strategy might be worth adding to the state agency's RESEA learning agenda.

CLEAR is a valuable resource for learning about the body of research that already exists about RESEA-related interventions, what that research says about the impacts of particular interventions, and how confident state agencies can be in the research findings. CLEAR can help state agencies identify high-quality evidence for interventions that their RESEA intervention currently uses or may want to adopt. In addition to CLEAR, there are numerous government websites, such as DOL/CEO's Completed Reports webpage[18] and DOL's Employment and Training Administration's Publication Database,[19] that have valuable information on existing evidence.

While CLEAR contains a comprehensive review of current research evidence on reemployment interventions, in some instances relevant information may exist that is not included in CLEAR. This may be the case for accounts of intervention designs that are new to the field of labor research, for

[18] https://www.dol.gov/agencies/oasp/evaluation/completedstudies

[19] http://wdr.doleta.gov/research/eta_default.cfm

example, a study described in a newspaper article or a report on an organization's website. Without CLEAR's prior review, it is important to assess the quality of that information. This can be done by:

- Identifying the source of the article or report. In general, studies published in a peer-reviewed journal (e.g., *Journal of Labor Economics, Social Science Review, Journal of Policy Analysis and Management*) or reports available on a federal or state agency website (e.g., DOL/CEO's Completed Reports or Current Studies),[20] a university research center, or professional evaluation firm are preferable.
- Scanning the article or report to see whether the authors use terms commonly found in scholarly literature (see Exhibit 2-5).

Exhibit 2-5. Important Terms in Identifying and Reviewing Research Evidence

Evaluation Type	Study Participants & Methods	Data Collection & Measurement	Analysis	Results
• cost-benefit • implementation study • interrupted time series • pre-post • matched comparison • quasi-experimental design (QED) • random assignment • randomized controlled trial (RCT) • regression discontinuity	• comparison group • contamination • control group • counterfactual • crossover • population • sample • intervention group	• baseline • follow-up • outcome measure • scale • variable	• analysis of variance • chi-square test • Cronbach's alpha • regression • regression model • significance testing • t-test	• alpha • bias • confound • generalizability • p-value • reliability • significance testing • statistically significant • validity

Source: *Evaluation Toolkit for Prospective Workforce Innovation Fund Grantees*, https://www.doleta.gov/workforce_innovation/pdf/grantees/FINAL_WIF_EvaluationToolkit_5-12-2014.pdf

In reviewing past evidence, it is important to differentiate between anecdotal accounts and scholarly evidence. Anecdotes from program staff about how their intervention worked or how they believe it helped their claimants may be useful for generating ideas or hypotheses but are not considered rigorous. Scholarly evidence related to the effectiveness of a program or policy are based on impact studies, discuss statistical significance, and provide numerical estimates of the size of any effects for study claimants who received the intervention compared to study participants who did not receive the intervention. The following examples illustrate differences between them.

Anecdotal evidence
"Our program was very successful. We served 630 participants. The program really helped people find jobs and improve their resumes."

Research-based evidence
"Over a 12-month period, 75.7 percent of program participants were employed compared with 64.8 percent of otherwise similar individuals not enrolled in the program. The program increased employment by 10 percentage points, which is statistically significant at the 1 percent level."

[20] DOL/CEO maintains two databases of studies it has conducted/commissioned that provide links to study reports and other information: Completed Reports (https://www.dol.gov/asp/evaluation/CompletedStudies.htm) and Current Studies (https://www.dol.gov/asp/evaluation/CurrentStudies.htm).

Craft Broad Research Questions

An essential component of the preliminary evaluation plan is a set of broad research questions that reflect the state agency's learning goals. Research questions will vary according to the type of evaluation conducted. Impact questions will help with understanding whether the intervention resulted in any changes in claimant outcomes. Outcome questions will help with understanding whether the intervention is meeting its goals. Process questions will help with understanding how the intervention is working on the ground.

Chapter 3 describes how to develop an initial set of broad research questions and how to work with an independent evaluator to refine these into a list of specific, well-defined research questions for inclusion in the state agency's evaluation.

Develop an Evaluation Design

Once the logic model and key research questions are developed, the next step is to determine which evaluation design(s) are appropriate for answering those research questions. The state agency's independent evaluator will help to develop and refine these decisions, but it is also important for the state agency to consider the research design beforehand as it will allow the state agency to work more effectively with its evaluator as the evaluation progresses.

To determine the most appropriate design(s), state agencies will want to consider the following:

- **Whether the state agency's preferred design(s) can answer the research questions of interest.** Some questions, such as how the RESEA intervention is operating and who is participating in it, may be best answered with an evaluation design that includes examining the intervention's implementation (an implementation study). Alternatively, questions about the effectiveness of the RESEA intervention are best answered with an evaluation design that compares the outcomes of claimants who received the intervention with those who did not (an impact study).
- **How large a sample the preferred design needs**. Some evaluation designs, particularly those that answer questions about intervention impact or effectiveness, require information on a large number of individuals in order to detect the impacts of the intervention. Even with designs that can answer questions about intervention effectiveness, some studies will have larger sample size requirements than others. When selecting your preferred design type, consider how many claimants will be selected for RESEA and included in the study. Studies that include a smaller number of claimants than is indicated by an evaluator's "power analysis"[21] run the risk of failing to detect impacts when impacts might actually exist.

Chapter 3 describes three common evaluation design types (i.e., impact, outcome, and implementation/process) that are often used to answer questions about reemployment interventions. Chapter 4 and Chapter 5 describe these designs in more detail. Chapter 4 also includes a discussion of sample size implications.

[21] Power analysis is a statistical method conducted to estimate the sample size needed to detect an effect.

3. Identifying the Evaluation's Purpose, Research Questions, and Designs

Defining the evaluation's purpose and what it will test is a key first step to planning a strong RESEA evaluation. Doing so will allow for the creation of a set of precise questions that the evaluation will aim to answer. These research questions will drive the evaluation's design, analysis, and findings. As such, it is important to develop questions that are specific, measurable, and appropriate for the RESEA intervention being tested. These questions may vary depending on the state agency's interests and needs. For the purposes of satisfying the statutory evidence requirements in the 2018 amendments to the SSA, the questions will involve demonstrating the impact of the RESEA intervention (i.e., the whole program and/or components of the program) on employment, earnings, and UI duration.

The evaluation design and methods selected will depend on what is required to answer the research questions. For example, the particular research questions will help determine whether the evaluator will need to collect *qualitative* data (e.g., derived from interviews, focus groups, document review, or on-site observations), *quantitative* data (e.g., numeric information obtained from program data, wage records), or both. (See Chapter 6 for more information on data.)

The following section discusses the importance of and process for:

- ✓ Establishing the evaluation purpose;
- ✓ Defining what to test; and
- ✓ Developing specific and measurable research questions.

The end of this chapter provides a brief introduction to three types of evaluation designs often used to answer questions about reemployment interventions. Chapter 4 and Chapter 5 describe these designs in more detail.

3.1 Determine the Evaluation's Purpose and Define What to Test

All evaluations require a clearly articulated purpose. Purpose refers to what the state agency wants to learn about its RESEA intervention. For example, a state agency may be interested in learning about its program processes (i.e., How do claimants flow through the steps of our RESEA intervention?) or about its impacts (i.e., Does our RESEA intervention improve claimants' employment and earning outcomes?). Identifying the purpose of the study will help a state agency better define the specific research questions that the evaluation will focus on answering with their evaluator. The state agency will also need to determine which intervention(s) will be studied.

Defining the evaluation's purpose and which interventions to test provides clarity on what the evaluation will (and will not) be able to do. For many state agencies, the evaluation's purpose may be associated with an initial set of broad research questions that stakeholders (i.e., agency leadership, AJC and frontline staff, staff from other agencies) can discuss and prioritize. Incorporating feedback from these groups can ensure that the evaluation is addressing their most pressing questions and that efforts align with the state agency's larger learning agenda. Broad research questions may concern understanding how an RESEA intervention is currently implemented (and any variation in that implementation across the state), how frequently claimants

receive particular services, claimants' employment outcomes, or the impact of the intervention on claimants' outcomes. Exhibit 3-1 offers examples of broad RESEA-relevant research questions.

Exhibit 3-1. Examples of Broad Research Questions

Question Type	Questions
Impact	• Does selection for RESEA improve claimants' employment outcomes compared to what their outcomes are likely to be otherwise? • Does more intensive case management by AJC staff improve claimants' employment and earnings outcomes?
Outcome	• How soon did claimants become reemployed after being selected for RESEA? • How do failure-to-report rates vary across AJCs?
Process	• What reemployment services and activities do claimants participate in? • What activities take place during meetings between claimants and case managers?

The following considerations can also be used to develop research questions:

- **Intervention of interest.** What RESEA intervention(s) does the state agency want to examine? To what extent do they want to determine the intervention's implementation fidelity (i.e., whether activities and services as delivered adhere to a certain intervention design or model)?
- **Outcomes addressed.** What outcomes will be measured to assess changes or successes?
- **Sites and service areas.** Which specific program sites (or offices) and/or geographic areas will be included?
- **Target population.** What population of claimants served by the intervention will be included in the study? For example, will all RESEA-eligible claimants be included or only those with interest in a certain industry? If the state agency does not plan to study the entire population, will enough claimants receive the intervention in order to detect the impacts of the intervention being studied?
- **Timing and length of study.** How long does the state agency plan to run the evaluation, and how will that time frame overlap with intervention development and operation? For example, will the proposed evaluation timeline allow the intervention to be observed during planning, in development, in early operation, or at maturity? Will intervention implementation stay consistent during the study period? Ensuring intervention implementation consistency is important for accuracy interpreting results.
- **Observation period.** Over what time period will outcomes be measured and when will tracking begin? In the RESEA setting, tracking typically begins at the outset of the UI claim period (UI claim application date) as one of the key measures of interest is duration of UI receipt, in addition to employment success. The length of the observation period can vary considerably and can be measured in years, quarters, or months.

3.2 Develop the Key Research Questions

Working with its independent evaluator, the state agency will refine previously developed, broad research questions and learning goals into specific evaluation research questions. These key research questions must be defined before beginning an evaluation. Research questions identify the

distinct RESEA intervention(s) to assess in a systematic 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. For example, a specific research question may ask, "Does offering a second one-on-one RESEA meeting help claimants return to work and leave UI more quickly than holding a single meeting with no option for a second?" In addition to asking about a specific intervention (a second one-on-one RESEA meeting) and outcomes (returning to work, leaving UI), this research question is measurable. That is, the question can be answered using administrative data to determine whether claimants were employed in a particular quarter. The question will likely require an impact study using administrative data to compare employment for claimants who were scheduled for a second meeting and for those who were not.
- **Answerable.** Key research questions must be answerable. Some research questions may not be answerable because data may not exist that readily capture the outcome of interest. For example, RESEA program managers may have an interest in impacts of services on claimant confidence in conducting a job search. However, confidence does not have a standard unit of measurement and may mean different things to different people. Research questions with outcomes not clearly measurable may also require additional consultation with the selected evaluator and the creation of new data collection instruments.
- **Discrete and limited in number.** In general, key research questions should be discrete, meaning that they do not overlap with one another. Typically, key research questions are few in number to help retain focus on the "what" and the "why" of the state-agency-sponsored evaluation and to articulate the parameters of the evaluation to stakeholders. The independent evaluator will further refine the questions and tie them to study methodology.
- **Rooted in program knowledge and realistic expectations.** Strong research questions are rooted in firm program knowledge—that is, an understanding of how the program operates and its intended theory of change. Research questions should also be based on an understanding of past similar efforts with demonstrated program results, and set with realistic expectations.

Exhibit 3-2 provides some examples of research questions that state agencies may use to learn more about their RESEA interventions. Once key research questions have been formulated, the state agency and its independent evaluator will need to determine the type of evaluation necessary to answer these questions. There are three major types of program evaluation, discussed next.

Exhibit 3-2. Example Specific Research Questions for RESEA Evaluations

Question Type	Example
Impact (Answers *does the program work?*)	• What is the impact of specific types of assistance (e.g., AJC orientation, self-directed use of online tools, general LMI) versus the impact of more intensive, individualized career services on claimants' employment outcomes six months after their UI claim? • What is the effect of case management or one-on-one assistance from staff on claimants' duration of UI benefit receipt? Is more intensive case management (e.g., more regular contact) more effective? • What is the impact of using technology to provide individualized LMI or other assistance to claimants on their employment, wages, and UI duration? • What is the effect of review of continued eligibility for benefits on claimants' average total UI benefit and benefit duration?

Question Type	Example
Outcomes (Answers *did the program meet its targets?*)	• What are the employment and earnings outcomes of RESEA intervention claimants six months after their initial RESEA session? • How soon do claimants who receive RESEA services find reemployment after filing their initial claim? • How long do RESEA claimants receive UI benefits?
Process or Implementation (Answers *how is the program designed and implemented?*)	• Who are RESEA claimants? What are barriers to participating in RESEA activities? • What activities do RESEA claimants engage most and least commonly? • What reemployment services do claimants find helpful? • What activities take place during meetings between claimants and case managers? Do activities vary by site?

3.3 Select From Three Major Types of Evaluations

Once the research questions are selected, state agencies will need to select an evaluation design that can appropriately answer them. Here and throughout the toolkit, three major types of evaluations are discussed—impact, outcome, and implementation/process studies—with various subtypes within each. The evaluation type ultimately selected will depend on (1) the research questions to be addressed (e.g., only impact studies can provide an estimate of the intervention's effect on claimant outcomes) and (2) the level of confidence with which the state agency wants to be able to interpret results. As noted earlier, an intervention may be a program as a whole or some component of that program. The three types of evaluation are:

- **Impact study,** which assesses the impact of an intervention on outcomes relative to the counterfactual—that is, some estimate of what those outcomes would have been in the absence of the intervention. An impact study uses either experimental (i.e., RCT) or quasi-experimental designs. An impact study answers the question, "did the intervention work better than a specific alternative (e.g., no intervention)?"
- **Outcomes study**, which measures the extent to which an intervention has achieved its intended outcome(s), focuses on outputs and outcomes relevant to effectiveness. Unlike an impact study, an outcomes study cannot show causal impacts. An outcomes study can help answer the question, "Did the intervention, policy, or organization achieve its pre-established targets or success measures?"
- **Implementation/Process study**, which assesses how the intervention is delivered relative to its intended theory of change and often include information on content, quantity, quality, and structure of services provided. An implementation study does not usually measure outcomes. An implementation study can be conducted on its own but is often paired with an impact and/or outcome evaluations. Implementation studies can help answer questions such as "Was the intervention implemented as intended?" or "How is the intervention operating in practice?"

What Is Impact?

An intervention's impact is its effect on an outcome of interest. Impact is therefore the change in an outcome that is directly and solely attributable to the intervention rather than to other factors. That is, impact captures how program participants' outcomes differ from those that would have occurred in the absence of the intervention, with the assumption that everything else in their lives except for their participation remained as it was (aka "all else equal").

Evaluation designs are on a continuum of study with low to high levels of focus, richness, and causal evidence. At one end of the continuum are implementation studies, designed to explore program

experiences and structures more broadly (though without a statistically rigorous assessment of the effect of the program). Implementation studies are often exploratory or motivated by broader policy concerns and questions. At the other end of the continuum are rigorous impact studies, designed to identify whether the intervention caused impacts on specific claimant outcomes as well as measure the magnitude of those impacts using the most statistically rigorous methods possible. In between, studies of varying methods and evidence levels can provide supporting evidence for program decisions, strategies, and practice.

Evaluators often sequence or combine multiple evaluation design types to maximize learning and provide a more complete assessment of the program being studied. For example, evaluators may choose to conduct an implementation study before or alongside an impact study.[22] An implementation study can produce a detailed understanding of how a program is implemented or is operating, which can provide important context for interpreting impact findings.

Though each type of evaluation can produce valuable information, only impact studies can provide the "high or moderate causal evidence" that the intervention will "reduce benefit duration as a result of improved employment outcomes" required by the SSA.[23] When a state agency conducts an impact study, consideration should be given to whether the evaluation design will earn a "High" or "Moderate" causal evidence rating from CLEAR if the evaluation is executed well. The next two chapters discuss each evaluation design type in more detail, including the likelihood of earning "High" or "Moderate" causal evidence ratings from CLEAR.

CLEAR Study Review Guidelines

CLEAR conducts systematic evidence reviews on labor topics and then summarizes study methodologies, findings, and policy or program implications. The reviews are guided by a protocol—designed to capture all research papers and reports that examine the topic area's research questions—and by guidelines appropriate to the study type. CLEAR has three different review guidelines:

- **Causal Evidence Review Guidelines** apply causal impact studies (i.e., experimental and nonexperimental studies that attempt to measure the effectiveness of an intervention) and assign them a rating indicating the strength of their causal evidence. https://clear.dol.gov/reference-documents/causal-evidence-guidelines-version-21
- **Descriptive Study Review Guidelines** are used to review quantitative studies that do not attempt to measure the causal effects of a program or policy. https://clear.dol.gov/reference-documents/quantitative-descriptive-guidelines
- **Implementation Studies Review Guidelines** apply to studies that examine the development and operation of a program, policy, or intervention. https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0

[22] According to UIPL 01-20, "DOL considers activities leading up to an impact evaluation that has the capability of producing a high or moderate causal rating to be interventions designated as "under evaluation." https://wdr.doleta.gov/directives/attach/UIPL/UIPL_1-20_acc.pdf.

[23] Social Security Act, 42 U.S.C. § 306 (2018)

4. Selecting an Evaluation Design: Impact Studies

State agency staff and other stakeholders are often interested in the effects of an intervention. In the case of RESEA, this means examining if and to what extent an RESEA intervention changes claimants' outcomes from what they would have been in absence of the intervention. In particular, state administrators may seek to understand how the RESEA intervention—disentangled from claimant characteristics or other factors—affects claimant outcomes such as UI duration, employment, and earnings. This information can help states inform program development and demonstrate the intervention's value to internal and external stakeholders. To answer questions about the effect of the intervention on claimant outcomes, an impact study is needed. An impact study design is the only type of evaluation design that can estimate not just what claimants' outcomes were, but *how* claimants' outcomes *with* the intervention differed from what would have occurred *without* the intervention. That is, properly constructed impact studies can determine *whether* the intervention changed claimants' outcomes, and, if so, *how much* change the intervention caused.

What is Impact?

An intervention's impact is its effect on an outcome of interest. Impact is therefore the change in an outcome that is directly and solely attributable to the intervention rather than to other factors. That is, impact captures how program participants' outcomes differ from those that would have occurred in the absence of the intervention, with the assumption that everything else in their lives except for their participation remained as it was (aka "all else equal").

This chapter discusses:

- ✓ The meaning of "impact" and "counterfactual"
- ✓ Sample sizes required to detect impacts
- ✓ Experimental studies, also called "randomized controlled trials" (RCTs) or random assignment studies
- ✓ Quasi-experimental design (QED) studies, including the three QED subtypes appropriate for evaluating RESEA interventions
- ✓ Key factors to consider when deciding on an impact study design.

4.1 Understanding Impact and the Counterfactual

An intervention's net impact is the change in an outcome of interest that is attributable to that intervention and measures what would have happened without the intervention. Measuring claimants' progress toward outcomes (e.g., faster reemployment or higher earnings) will help state agencies determine whether an RESEA intervention is meeting its goals, but measuring impact provides an estimate of how much an RESEA intervention may have contributed to those outcomes.

Ideally, to measure impact, claimants' outcomes would be measured in two parallel worlds: one in which they participated in the intervention and another in which they did not participate but everything else in their life remained the same. Of course, it is impossible to observe the exact same claimants over the exact same period of time in two parallel worlds. To substitute for that impossible set of conditions, an impact study creates a "comparison" group of claimants who are as

similar as possible to the claimants selected for the intervention (called the "intervention group")—except that the comparison group does not receive the intervention. The impact study can estimate the intervention's impact by comparing the outcomes of the intervention group to those achieved by the comparison group. If the study is properly designed and implemented (e.g., services are provided to the respective groups as planned, holding all else equal), then the experience of the comparison group represents what the intervention group would have experienced absent the intervention. This is called the "counterfactual."

Impact Study Groups

- **Comparison group**: made up of claimants not selected for the intervention, but otherwise identical to claimants who were selected. Also known as the "control" group in random assignment studies, because the group is created through a controlled experimental process.
- **Intervention group**: made up of claimants selected for the intervention.

As evaluators compare the outcomes of study participants across intervention and comparison groups to determine the intervention's impact of, the claimants who make up the two groups must be similar across both measureable and unmeasurable characteristics. If the intervention and comparison groups differ on those kinds of characteristics, then any difference in outcomes might be because of those differences in claimants' characteristics, not because of the intervention.

What Is the Counterfactual?

The counterfactual represents the condition or outcomes claimants will experience in the absence of the intervention. It is impossible to naturally observe the counterfactual for intervention group members. The most rigorous way for evaluators to construct a counterfactual is by random assignment. For a valid counterfactual, there must be no systematic differences between the intervention group and the comparison group. Random assignment creates two groups with no systematic differences, other than their random assignment status. It is difficult, often impossible, for other impact evaluation designs to achieve the same.

Evaluators can construct the counterfactual condition—as best as possible—using either of two broad impact study designs: experimental or quasi-experimental.

- **Experimental studies,** also known as randomized controlled trials or random assignment studies, are the strongest form of impact study. Experimental studies are also unique in their ability to maximize similarity between the intervention group and comparison group on both observed characteristics (such as age, gender, and work history) and unobserved characteristics (such as level of confidence and personality) that might influence outcomes. In experimental studies, the counterfactual condition is created by *random assignment* of eligible individuals to either the intervention group or the comparison group. In the case of RESEA, the evaluator would randomly assign claimants eligible for RESEA either to a group that receives the RESEA intervention being tested (intervention group) or to a group that does not receive the RESEA intervention being tested (comparison group). Claimants are assigned to these groups based on the functional equivalent of a coin toss, usually by computer algorithm. Random assignment allows the evaluator to reasonably attribute any difference in outcomes between the two groups to the intervention being tested, rather than to the characteristics of claimants.
- **Quasi-experimental studies** aim to create a counterfactual condition that is as similar as possible to the intervention condition by using methods other than random assignment. Often this involves using administrative data to form a comparison group by identifying claimants not selected for the intervention who have similar characteristics to those claimants selected for the intervention (i.e., intervention group).

- One key drawback to using quasi-experimental methods to create the counterfactual condition is that the characteristics recorded in administrative data are not exhaustive. If claimants in the comparison and intervention group differ on one or more key characteristic not recorded in the administrative data, the two groups will not be made up of truly similar claimants.
- Any situation where a claimant receives the intervention based in part on unmeasured factors (e.g., caseworker judgement or claimants' motivation) will make it impossible to create a comparison group that is truly similar to the intervention group. The two groups might be similar on observed characteristics. But they will differ on important unobserved characteristics that determine receipt of the RESEA intervention (e.g., attendance at meetings or workshops) and are also likely to affect outcomes.

The next section discusses key concepts related to study sample size needed to detect impacts. Then, sections 4.3 and 4.4 describes experimental and quasi-experimental designs, respectively.

4.2 Detecting Impacts – An Important Note about Sample Size

Determining what sample size will be needed is an important factor to consider when selecting an evaluation design. Without an adequate sample size, evaluators will not be able to detect impacts, regardless of the design they choose and how well the study is executed. The term "sample" refers to the individuals, whether in the intervention or comparison group, who are included in the study. The number of individuals included is the study's "sample size." The size of a study's sample has important implications for the confidence with which it is possible to make conclusions about that study's findings, and small sample sizes lead to less confidence in the study's results.

Key Terms

- **Sample**. The individuals in the intervention and comparison groups (i.e., all those included in the study).
- **Sample size**. The number of individuals in the sample.
- **Power**. The probability that the study will be able to statistically detect the intervention's impact if an impact does actually exist.
- **Minimum detectable effect**. Smallest impact of the program that you can reasonably expect to detect given your sample size and study design.

How many study participants ("sample members") an evaluation needs is a critical consideration for detecting impacts (a concept known as a study's "power"). There is no magic sample size number that will guarantee that the evaluation will detect impacts, but evaluations with larger sample sizes are more likely to detect impacts. How large a sample the evaluation will need depends on several factors, listed in Exhibit 4-1.

Exhibit 4-1. Factors Influencing Sample Size Needs

Likely impact of the intervention being tested	• Whole-programs likely have larger impacts than individual program components. • Tests of interventions with larger likely impacts require less sample than do tests of interventions with smaller relative impacts.
Evaluation design type used	• To detect an impact of any given size, quasi-experimental designs require larger sample sizes than do experimental designs. • Randomly assigning at the individual (claimant) level, rather than at the cluster level (e.g., by AJC), reduces sample size requirements.
Outcomes measured	• Measuring impact on proximal outcomes (e.g., meeting attendance) requires smaller sample sizes. • Measuring impact on employment and earnings requires sample sizes in the thousands or tens of thousands, depending on evaluation design and intervention tested.

Exhibit 4-3 in Section 4.3 provides some rough sample size numbers for experimental and quasi-experimental designs. These are estimates, and an independent evaluator can (and should) conduct a "power analysis" to provide appropriate sample sizes.[24] A power analysis that is tailored to the evaluation will generate the best estimate of the sample size needed to detect impacts of the intervention studied. A power analysis can also estimate the study's minimum detectable impact (MDE), which is how small of an impact the study would likely be able to detect, given the expected sample size. If the impact that the intervention is expected to have is smaller than the MDE, then the study will need to generate a larger sample.

Options for Creating Sample Size Needed to Detect Impacts

The sample size needed to detect impacts may appear large—perhaps even huge samples. State agencies that are concerned about sample sizes may consider these options:

- **Evaluate the whole program rather than any one component**. Whole-program tests usually require smaller sample sizes than tests of individual components. A whole RESEA program evaluation will require the smallest sample.
- **Conduct the evaluation over multiple years**. A longer evaluation study period will provide time to accumulate the sample. For example, say a state agency has approximately 4,000 RESEA claimants per year. If the state agency conducts rolling random assignment of claimants for three years, the state agency will end up with a sample of approximately 12,000 total claimants in its evaluation.
- **Pool samples across multiple states**. This approach would require state agencies with similar interventions to work together with a single evaluator to combine their data to test the impacts of the shared intervention. State agencies interested in working with other state agencies to pool sample should work with a qualified evaluator to do so.
- **Revise the research question**. A state agency might consider focusing on a larger intervention or on different outcomes that require smaller sample sizes.

4.3 Experimental Studies

Experimental studies provide the best scientific evidence of whether an intervention is effective (i.e., having the intended impact on outcomes) or not. RCTs are characterized by using random assignment to determine which study participants get access to the intervention. With a large enough number of RESEA intervention-eligible claimants included in the study, the characteristics of claimants in the intervention group and in the comparison group will be equivalent on average at the outset of the study (at "baseline").

Randomly assigning claimants to groups thus holds all else equal—beyond assignment or not

A Checklist for Launching an Experimental Study

- Select an evaluator with adequate experience in experimental design.
- Ask the evaluator to conduct a power analysis to determine a sample size estimate.
- Work with the evaluator to:
  - Select a point of random assignment.
  - Incorporate random assignment into RESEA selection and scheduling processes.
  - Create and train staff in evaluation procedures.
  - Monitor implementation to ensure staff are providing services based on claimants' random assignment status.

[24] A power analysis is a calculation that can estimate—given a likely program impact, a desired statistical significance threshold, and analysis design—the number of claimants required in order for your study to detect impacts. Power analyses typically rely on the findings from previous research to determine a likely program impact that can be included in the calculation.

to the RESEA intervention. Differences in outcomes between the intervention group and the comparison group are therefore a strong estimate of the intervention's impact.

While experimental studies provide high-quality results, they require effort to implement with fidelity. The following issues should be considered when implementing this type of design.

Logistics of Random Assignment

In consultation with their evaluator, state agencies will select a method to randomly assign claimants into the intervention or comparison group as needed. Random assignment is most easily and reliably done using a computer algorithm that either is directly incorporated into the state agency's scheduling system or is an external web-based random assignment system. However, if using a computer algorithm is not feasible for your state agency, there are ways to perform random assignment manually. Exhibit 4-2 summarizes the advantages and challenges for each random assignment method.

Exhibit 4-2. Random Assignment Methods – Advantages and Challenges by Type

Type	Advantages	Challenges
Computer algorithm (web-based or computer software)	• Reduces staff burden, depersonalizes the random assignment process, and can be incorporated into existing processes to reduce burden on frontline staff. • Allows stratification, which ensures intervention and comparison groups are the same with respect to particular characteristics, for example within AJCs, across particular subgroups of claimants, and across time. • Can be inserted directly into your scheduling system or can use external system. • Can be easily monitored to make sure it is working correctly; cannot be easily tampered with.	• If inserted directly into state agency systems, requires several days of IT staff time for programming and testing.
Manual random assignment	• Does not require changes to your IT systems.	• Cannot be easily monitored; difficult to ensure random assignment is working correctly. • Requires a regular commitment from a central staff member to communicate weekly assignments to sites in the study • Evaluator cannot oversee the random assignment process; if assignment is not truly random, findings will not be reliable

While both are viable options, a computerized algorithm is a decidedly stronger approach, if feasible. Computerized random assignment is more reliable, easier to monitor, depersonalizes the entire process, and eliminates any suspicions of human error or manipulation. An algorithm will ensure random assignment ratios are maintained across offices and sub-groups of interest and dramatically reduces frontline staff burden related to the evaluation. If the state agency's chosen assignment mechanism does not truly randomly assign claimants to the intervention, the evaluation will not be able to confidently detect the impacts of the program itself.

Point of Random Assignment

In consultation with their evaluator, state agencies will select a "point of random assignment." This refers to the place in the claim process in which claimants are randomly assigned to the

intervention or comparison group. The point of random assignment typically be as close as possible to when claimants would start receiving the intervention if the claimant is assigned to the intervention group.

- For evaluations of the impact of a **whole program**, random assignment likely occurs as part of the RESEA claimant selection process.
- In evaluations of a **program component**, random assignment may also occur as a part of the RESEA claimant selection process. In that case, selection will define which version of the program the claimant is selected for—one that either includes the component being tested or excludes it. Random assignment can also occur after selection for RESEA, but needs to occur before the claimant is scheduled for the component being tested.

See Appendix H for a sample RESEA random assignment flowchart. The evaluator can help state agencies select a point of random assignment that is appropriate for its program and design the logistics of implementing random assignment.

Unit of Assignment

Random assignment can occur by claimant or by cluster. Assignment by claimant is most typical. Claimant-level assignment means that each claimant could be assigned to either the intervention or comparison group irrespective of where they are from. Cluster-level assignment assigns sets of claimants to the treatment or comparison group. For instance, in a study of components, a study might randomly assign some AJCs to provide a version of the program that includes the component and other AJCs to provide a version of the program that excludes that component. In that example all claimants served by the first set of AJCs would be in the intervention group and all claimants served by the second set of AJCs would be in the in the comparison group. So assignment is determined by the AJCs. That is a "cluster"-level assignment. Cluster-level assignment can have some logistical benefits, but studies that use cluster-level assignment require larger sample sizes than do studies that use individual-level assignment.

Random Assignment Ratios

Random assignment "ratios" refer to the percentage of RESEA-eligible claimants assigned to the intervention group versus the comparison group. For example, a study that assigns an equal number of claimants to each group is said to be using a 1:1 random assignment ratio.

Other random assignment ratios are possible, such that more RESEA-eligible claimants end up in the intervention group than in the comparison group, or vice versa. For instance, suppose a state wants to evaluate its RESEA program as a whole and it is able to select no more than one-third of its non-exempt UI claimants for RESEA. In that case, the state might use a 1:2 random assignment ratio, randomly assigning one claimant to the intervention group for every two claimants randomly assigned to the comparison group. This approach allows the state to maximize its sample size.

All else equal, a 1:1 ratio will produce greater statistical power and require less sample to detect impacts than will uneven ratios.

Monitoring Random Assignment and Service Provision

Procedures for closely monitoring the random assignment process are needed to confirm that it is carried out as intended. This means monitoring claimants to ensure they are:

- Being assigned in accordance with the pre-determined random assignment ratio (whatever that study's ratio was designed to be, 1:1, 2:1, etc.).

- Assigned such that, on average, the intervention and comparison groups contain similar portions of claimants with particular observable characteristics.
- (If assigning individual claimants), assigned such that, on average the proportion assigned to treatment and comparison groups are similar across all AJCs participating in the study.
- Receiving services based on the group (intervention or comparison) that they were assigned to. More importantly, claimants who are assigned to the comparison group should not receive the services provided through the intervention being evaluated. This is often referred to as "cross-over."

In addition to monitoring random assignment, evaluators and states should also monitor service provision. Typically, when testing an intervention, you will want as many intervention group claimants as possible to actually receive the intervention. Similarly, claimants assigned to the comparison group should not receive services that are unique to the intervention. At the same time, for RESEA, it is likely that some proportion of intervention group claimants will not participate fully (or sometimes at all). For instance, some claimants in the intervention group may not attend the expected number of meetings. Their non-participation may be for completely legitimate reasons. For example, a claimant may become reemployed before the date of the first scheduled meeting and therefore have no reason to attend.

It is possible that some claimants assigned to the comparison group will receive services elsewhere in the community that are similar to the RESEA intervention being tested, and that is appropriate as long as it is not part of the RESEA intervention. For example, comparison group members can pursue AJC-based reemployment services similar to RESEA's on their own. This flexibility and likelihood does not pose a problem for the integrity of the RESEA evaluation. The appropriate counterfactual for RESEA is what the intervention group claimants will have received otherwise—and for some fraction of those claimants, that could have included availing themselves of AJCs' reemployment services.

An important part of a random assignment impact evaluation of RESEA is understanding and documenting how members of the intervention and comparison groups differ in their use of services. Measuring and analyzing such data help demonstrate the extent to which the RESEA program has succeeded in its aim of promoting greater use of reemployment services. Understanding those differences in services received between the groups is, in turn, important for understanding what is behind any observed differences (or lack thereof) in the employment and UI duration outcomes generated by the two groups. An implementation study can help provide that information to complement impact estimates (see Chapter 5).

Sample Size

As discussed in Section 4.2, the minimum sample size needed to detect impacts depends on several considerations. Random assignment studies that assign equal numbers of claimants to the intervention group and comparison group require the smallest sample sizes among the impact designs that can be used to evaluate RESEA. That being said, very large sample sizes are required to detect impacts on many outcomes of interest for RESEA interventions, even for random assignment studies with equal random assignment ratios.

Informed by prior random assignment evaluations of the RESEA program,[25] Exhibit 4-3 provides rough estimates on the number of claimants that state agencies will need to randomly assign to detect impacts on key outcomes. However, these ranges are for illustrative purposes. State agencies should work with their independent evaluator to generate sample size estimates for their particular evaluation and the intervention that is being evaluated. Here an intervention's size ("large," "small") refers to how large of an impact it is expected to have. Roughly speaking, for the exhibit's purposes, a "large" component is one that accounts for a third or more of the program's overall impact. A "small" component is one that accounts for less than one-third of the program's overall impact. For example, a small component might be providing labor market information or using a self-scheduling system. A large component might be intensive case management or a complete reemployment services package.

Exhibit 4-3. Sample Size Estimates for Tests Using Experimental Design, by Intervention and Outcome

Intervention Size	Outcome		
	Proximate Outcome (e.g., meeting attendance)	*UI Duration*	*Employment (2 quarters after start of claim)*
"Small" Component	1,000-3,000	100,000+	Hundreds of thousands
"Large" Component	500-1,000	50,000-100,000	100,000+
Whole Program		5,000-10,000	10,000-25,000

Note: Estimates are for random assignment evaluations of programs that assign equal numbers of claimants to the intervention group and the comparison group (1:1 ratio). Studies that use quasi-experimental designs, use uneven assignment ratios, or that assign by cluster (e.g., by AJC) rather than by claimant will require still larger sample sizes.

The whole RESEA program will have a larger impact than will any of its individual component parts. The larger the impact expected, the smaller the sample size needed to detect it. As the exhibit shows, sample sizes needed to detect impacts of particular components of an RESEA program are much larger than the sample sizes needed to detect impacts of the whole program. For example, state agencies will likely need to randomly assign 100,000 claimants or more to estimate the impact of a particular program component on claimants' employment outcomes, even for a relatively large component. A smaller component requires even larger sample sizes. Sample size requirements will also be larger—often several times larger—for studies that:

- Do not assign an equal number of claimants to the intervention and comparison groups; and/or
- Perform assignment at the "cluster" level (e.g., group status is determined by what AJC or WDB the claimant is served by), rather than at the individual claimant level.

As noted earlier in this chapter, state agencies that serve too few RESEA claimants to support a credible impact study on an outcome they are interested in may be able to work with other states to implement similar interventions and pool their data into a single evaluation that will meet sample size requirements. The independent evaluator should conduct power analyses tailored to the

[25] Benus, J., Poe-Yamagata, E., Wang, Y., & Blass, E. (2008). *Reemployment and Eligibility Assessment (REA) study FY2005 initiative*. Columbia, MD: IMPAQ International; Poe-Yamagata, E., Benus, J., Bill, N., Carrington, H., Michaelides, M., & Shen, T. (2011). *Impact of the Reemployment and Eligibility Assessment (REA) initiative*. Columbia, MD: IMPAQ International. Retrieved from http://wdr.doleta.gov/research/FullText_Documents/ETAOP_2012_08_REA_Nevada_Follow_up_Report.pdf; Michaelides, M., Poe-Yamagata, E., Benus, J., & Tirumalasetti, D. (2012). Impact of the Reemployment and Eligibility Assessment (REA) initiative in Nevada. Columbia, MD: IMPAQ International.

intervention they plan to test, the size of the impact they expect to detect, and methods they will use to detect that impact.

CLEAR has well-defined criteria for random assignment studies, which are outlined in the box below. State agencies should review these criteria with their evaluators while designing and implementing their random assignment evaluation. Random assignment evaluations that can meet CLEAR's criteria can be confident that their findings truly do reflect the interventions' impacts, rather than some other factor.

CLEAR's Criteria for Rating Random Assignment Studies

CLEAR has established criteria for assessing the quality of studies that use random assignment designs. Those criteria focus on whether:

- Assignment to the intervention or comparison group was random.
- Once assigned, claimants' intervention or comparison group status was maintained (no switching of groups after assignment).
- Data needed for analysis were obtained for nearly all members of both groups (i.e., the fraction of participants that are lost from the study, termed "attrition," is low). Note that acceptable levels of attrition depend on the combination of overall attrition (i.e., across both groups) and differential attrition (i.e., difference between the groups). For example, when overall attrition is five percent, CLEAR standards allow levels of differential attrition between 6-10 percent.

Evaluators or program operators can run into problems with any of those criteria. There may be interest in making non—allowable exceptions to which group claimants are and are not assigned. Or the evaluator and program operator may decide that a claimant should be moved to a different group, or removed from the study, after assignment has occurred, as a result of factors like their assessment of claimants' needs. Or data collection may not be thorough enough to reach all study participants, attrition is high, or there may be high rates of missing data on key elements of interest.

When those problems happen, you no longer can be sure that the comparison group is statistically similar to the intervention group. Such a random assignment study is no longer eligible for CLEAR's High rating. To receive CLEAR's next highest rating (Moderate), a study with high attrition must provide additional proof that the intervention and comparison groups are equivalent on key pre-intervention characteristics If it does not, the study receives a "Low" rating.

4.4 Quasi-Experimental Design Studies

QEDs attempt to measure an intervention's impact by using methods other than random assignment to create a comparison group that is as similar as possible to the intervention group. There are many types of QEDs, but they all fall into one of two broad groups:

- **Retrospective designs** use administrative data that have already been collected to create a credible comparison group. Because the data already exist, quasi-experimental studies using retrospective designs can be completed more quickly than evaluations using other types of impact designs. However, a retrospective design can only address questions that can be answered by that existing data.
- **Prospective designs** plan to use data yet to be collected. As a result, these evaluations take longer, as data collection instruments are identified, modified, or developed and then used to collect the data to be analyzed. An advantage of evaluations using prospective designs is that evaluators can tailor data collection to capture information of interest that is not already in existing data sets. Random assignment designs are also prospective designs, but the rest of this section will focus on prospective QEDs.

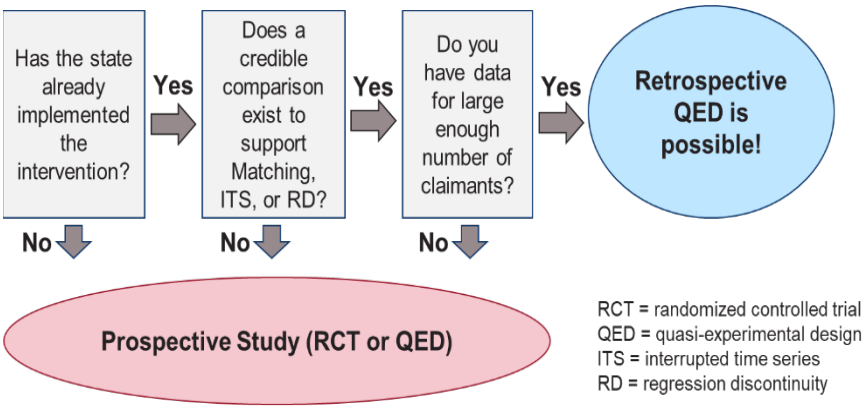
Exhibit 4-4 highlights key differences between retrospective and prospective designs.

Exhibit 4-4. Comparison of Retrospective and Prospective Studies

	Retrospective Study	Prospective Study
Process	• Collect data, then design study. • Use data that have already been collected, typically in the day-to-day administration of the program.	• Design study, then implement intervention, enroll study participants, and collect data.
Advantages	• Much quicker. • Likely cheaper.	• Can answer a broader set of research questions. • If using an RCT, more confidence in findings' internal validity.
Disadvantages	• Analytically complex; requires advanced statistical expertise. • Cannot answer questions for which data do not exist.	• Much slower.
Possible design types	• QED only.	• RCT or QED.

Exhibit 4-5 offers a decision-making tree to determine whether a retrospective QED is an option for a state agency to evaluate its RESEA intervention's impact. As discussed in Section 4.2, evaluations using QEDs typically need larger sample sizes than randomized controlled trials need to detect impacts.

Exhibit 4-5. Questions to Consider When Selecting an Impact Design



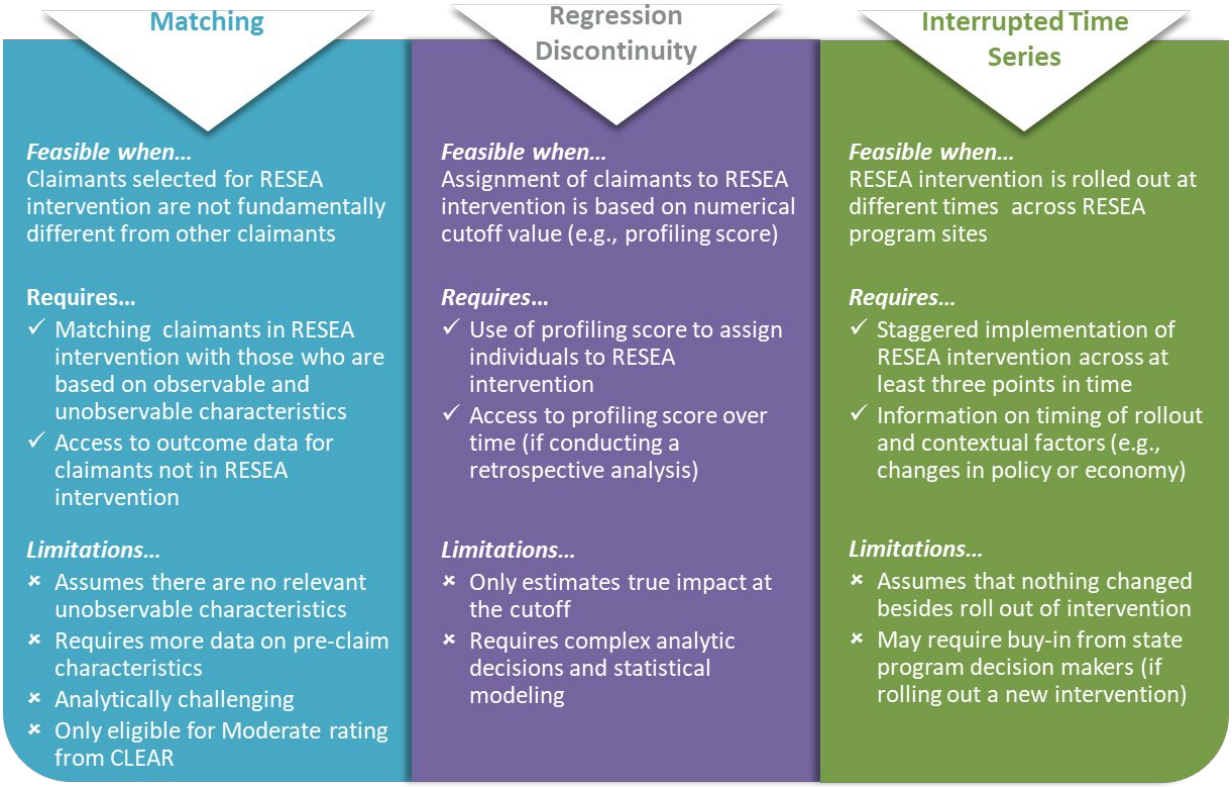
QEDs do not proactively or randomly assign claimants to groups, but QEDs need to create equivalent intervention and comparison groups using a number of methods. For RESEA evaluations, the ability to use retrospective data may be the most prominent advantage of QEDs, because it allows studies to be conducted more quickly and cost-effectively. However, using retrospective data means that state agencies can evaluate only interventions implemented in the past (for which data exist), not a new component, approach, or model.

On the whole, QEDs can be less intrusive on program operations than experimental studies. However, quasi-experimental methods raise more questions about internal validity, i.e., whether the study has truly attributed impacts to the program. Further, to be considered credible, QEDs usually require high-quality data as well as evaluator technical expertise in creating equivalent comparison groups. Depending on the design used, results from QEDs can be challenging to interpret, especially for non-technical readers. At best, a well-executed QED study can achieve a "Moderate" causal evidence rating from CLEAR. It is also important to note that the sample sizes required for some types of QEDs are larger, often several times larger, than for random assignment

studies and so may be feasible only for very large states or by accumulating program data over a long period of time (e.g., multiple years).

The following sections describe three types of QEDs that are either common or particularly relevant to evaluating RESEA interventions. Exhibit 4-6 gives an overview of these QED options.

Exhibit 4-6. Summary of Potential Quasi-Experimental Design Options for RESEA Evaluations



Matching and Other Regression Methods

Matching and other regression methods are QEDs that aim to compare the outcomes of claimants who participated in the RESEA intervention of interest to outcomes of very similar claimants who did not participate, based on information in the claimants' administrative data. These methods make statistical adjustments for any observed differences between the two groups. A common way to do this is through "matching," a process that identifies key pre-intervention characteristics that are likely to influence outcomes and creates a comparison group that is as similar as possible to the intervention group on those characteristics. Exhibit 4-7 illustrates the basic approach for creating matched comparison groups.

Exhibit 4-7. Generalized Approach for Matching

Evaluators start with data on study participants who received the intervention and those who did not receive the intervention. Among those who did not receive the intervention, the evaluator has to identify a subset:

- Who are otherwise similar to the intervention group members in all observed ways that could affect their outcomes; and
- Whom the evaluator has no reason to believe are likely to differ in unobserved ways, such as their motivation or knowledge, which could affect outcomes.

This type of design is typically used retrospectively, rather than prospectively. That is, matching is used in an attempt to mimic random assignment, when working with existing data for past periods.

The RESEA context presents numerous challenges to creating internally valid matching designs. In many states, the structure of RESEA may make matching designs infeasible. For example, if a state agency uses a profiling score as the method to select claimants for a statewide RESEA program, there cannot be two RESEA-eligible claimants with the same profiling score where one was assigned to RESEA and the other was not assigned to RESEA. As a result, RESEA claimants fundamentally differ from non-RESEA claimants in a way that makes it impossible to identify valid matches for a matched comparison group approach. Cases where there are non-RESEA claimants with the same profiling score as RESEA claimants are likely to still violate some other matching rule. For instance, if the state agency uses different profiling score cutoff values in different weeks or in different offices, the evaluator will want to do the comparison only within those weeks or within those offices. As such, there are no valid comparisons among the population of UI claimants who were not selected for RESEA.

One situation where valid comparisons may exist for a matching study is if the state agency implemented the intervention in one part of the state but not others. In that case, there may be claimants in parts of the state where the intervention was not implemented who are reasonably similar to those who received the intervention. Still there may be concerns that the parts of the state where the intervention was implemented are inherently different from the parts where it was not. Another drawback is that because receipt of the intervention in such a case is determined by geography, this is a cluster-assignment design, and so will require much larger sample sizes than a random assignment design or even a matching design where assignment is at the individual level.

Matching strategies are generally not used to estimate the impact of RESEA program components. Take an example where there is an interest in estimating the impact of RESEA meetings by comparing RESEA claimants who attended a meeting with those who did not attend. In most situations, matching is likely not a viable strategy for credibly estimating the impact of attending a meeting, because claimants who attend an RESEA meeting are almost certainly systematically different from those who did not attend. Consider the likely reasons that someone may not attend a meeting: they may have major life barriers, or they might have less motivation. It can be expected that those claimants to have worse average outcomes than claimants who did attend. Conversely, some claimants may choose not to attend because they already have a job lined up. In this case, those claimants can be expected to have better average outcomes. Better or worse, RESEA claimants who did and did not attend a meeting are likely to vary in important ways associated with the outcomes of interest. Thus, non-attenders are not a valid comparison group. These kinds of comparisons might be interesting or valuable to examine for other reasons, but they cannot be used as the basis for demonstrating an intervention's impact. These sorts of complexities highlight the importance of working with a skilled, independent evaluator to understand the most feasible, appropriate, rigorous design to answer the evaluation's research questions.

As with random assignment studies, CLEAR has well-defined criteria for matched comparison group QED studies, which are outlined in the box below.

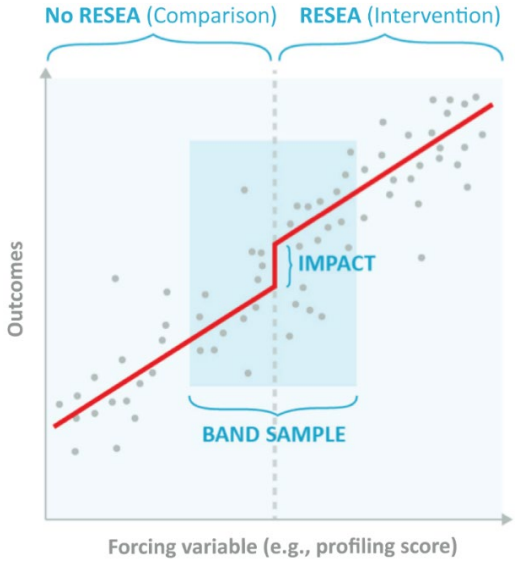
CLEAR's Criteria for Rating Matched Comparison Group Designs

CLEAR's standards for assessing the quality of matched comparison group designs focus on confirming that the intervention and comparison group members truly are similar. This confirmation process includes testing whether averages on key pre-intervention characteristics (such as income, employment history, education, and demographics) are similar for both groups. Until the evaluator getswell into the analysis, it is hard to know whether the design will meet the standards, because much of this work involves creating the analytic sample. The evaluator needs that analytic sample to know how well matched group members are on relevant baseline characteristics. (For a random assignment study, once a state agency decides on random assignment, the design is more likely—though not guaranteed—to meet standards.)

Regression Discontinuity

Regression discontinuity study designs can be used for impact evaluations when assignment to RESEA is based on a strict cutoff score on a numeric measure, such as a profiling score. Regression discontinuity designs take advantage of this fact: claimants who are just below the profiling score selection cutoff should be nearly identical to those just above the profiling score selection cutoff in all ways (even unobservable ones) other than having been selected for the program. Regression discontinuity designs estimate the impact of a program by comparing the outcomes of claimants just above the cutoff (i.e., those selected for RESEA) to the outcomes of claimants just below the cutoff (i.e., those not selected for RESEA). As such, regression discontinuity studies typically limit their analysis samples to claimants with scores near the cutoff. That range of scores are referred to as the "bandwidth" (or "band"). CLEAR does not yet have explicit criteria or guidelines for regression discontinuity designs; however, CLEAR does acknowledge that regression discontinuity designs are strong. Exhibit 4-8 gives a graphical representation of how regression discontinuity designs form their intervention and comparison groups.

Exhibit 4-8. Visualization of Regression Discontinuity



In general, the level of complexity involved in developing the design and analyzing results is much higher for regression discontinuity designs than for random assignment designs. Beyond that complexity, using regression discontinuity to evaluate RESEA presents three major challenges:

- Regression discontinuity design estimates of impact apply only to claimants near the selection cutoff. If RESEA has a very different impact on claimants well above or well below the cutoff, a regression discontinuity design will not be able to show that. In contrast, a random assignment design—with large enough samples—helps create separate estimates for high/medium/low profiling scores. This characteristic of a regression discontinuity design would not be a major problem if the policy question were whether to change the cutoff a little. For example, for a state agency considering a small change to the profiling score cutoff for RESEA, a regression discontinuity design could estimate the impact on those claimants on the margin of inclusion in the program. However the statutory policy question is, does RESEA work for everyone served? For this policy question, regression discontinuity is an imperfect design. It only works if the assumption—without evidence—is that impact is nearly invariant with profiling score. That would be a problematic assumption.
- Evaluations using a regression discontinuity design require samples several times larger than those required for random assignment evaluations. This is partially because the regression discontinuity design evaluation can use only a subset of claimants, those with profiling scores close to the cutoff. Given that most state agencies will have trouble finding enough RESEA-eligible claimants for random assignment from among all RESEA-eligible claimants, much less only those near the cutoff, regression discontinuity is likely to be feasible only for very large states or a large consortium of states.
- A regression discontinuity design will be able to estimate only the impact of the intervention for which selection was determined by the profiling score. This typically means that regression discontinuity designs can be used only to evaluate the RESEA program as a whole. A regression discontinuity design cannot estimate the impact of individual intervention components or activities unless a state agency had different cutoffs for those different levels of services. For example, suppose that claimants with profiling scores above one cutoff level are selected for a base version of RESEA and claimants with scores above some higher cutoff level are selected for an enhanced version that includes some additional component. In that case, regression discontinuity could (given a large enough sample) be used to estimate the impact of that component.

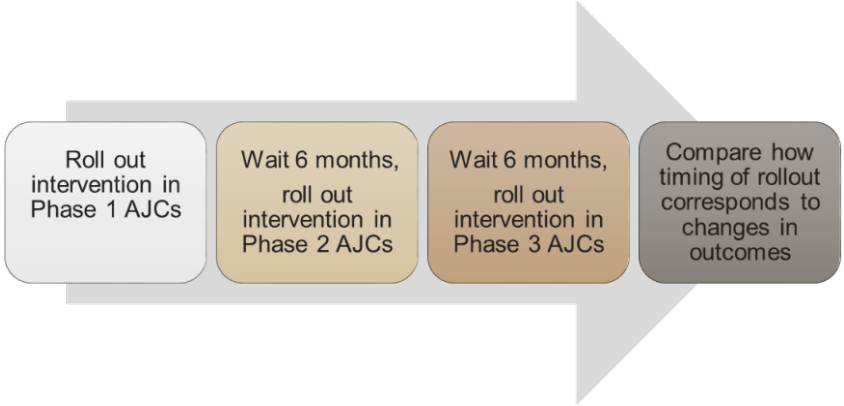
Interrupted Time Series

An interrupted time series study design exploits staggered rollout of a program or intervention. For example, suppose that a state has many, perhaps several dozen offices and that different offices begin to implement a new RESEA intervention at different times.[26] Suppose further that which offices implement this new intervention earlier versus later is not decided based on some characteristics of the local economy, AJC, or population served that could be associated with claimant outcomes. Under such conditions, the impact of the intervention could be estimated by comparing changes in outcomes at sites that rolled out the intervention versus sites that have not yet rolled out the intervention. In analysis, impact is adjusted for the trend in outcome change before and after the onset of the intervention, thus improving the credibility of the impact estimate.

Exhibit 4-9 provides a basic overview of how an interrupted time series design might work. A state agency could split its AJCs into three groups and roll out the intervention sequentially, starting with the Phase 1 AJCs. Phase 2 AJCs would implement the intervention six months later. Six months after that, the Phase 3 AJCs would roll out the intervention.

[26] Interrupted time series (ITS) may also be referred to as *comparative interrupted time series* (CITS).

Exhibit 4-9. Implementing Interrupted Times Series in Practice



This kind of interrupted time series design uses both change within offices and comparisons across offices to help ensure that other statewide changes occurring at the same time as the intervention is being rolled out do not contaminate the impact estimates. Interventions cannot take credit (or penalize) for a rapidly improving (or worsening) local labor market, for example. The staggered rollout addresses this concern. Interrupted time series designs use outcomes for claimants from offices where the intervention was not yet rolled out to control for how outcomes would have changed absent the intervention. An interrupted time series formalizes this insight and shows how to properly compute the precision of the resulting estimates.

Below are some considerations for state agencies interested in using an interrupted time series design to evaluate RESEA:

- The study design requires very large sample sizes because assignment is at the cluster (e.g., office) level, not the claimant level. In practice, this means that there must be a large number of offices (probably several dozen) and (relative to random assignment) there must be many more RESEA-eligible claimants.
- Accumulating large samples may take a long time in the context of the RESEA program. Other than in very large states that involved all their AJCs in the evaluation, an interrupted time series design with a one-year rollout period would not be sufficiently powered to detect intervention impacts. Results would have to be pooled across several years. This would mean that sites in later phases, such as Phase 3 above, would have to agree to delay intervention rollout for several years.

CLEAR has well-defined criteria for interrupted time series QED studies, which are outlined in the box below.

CLEAR's Criteria for Rating Interrupted Time Series Designs

CLEAR's standards for interrupted time series designs require that the order in which sites are chosen to roll out the intervention must be unrelated to their pre-intervention outcomes. For RESEA, that means that the average pre-intervention outcomes of employment and UI duration of claimants served by the sites that rolled out the RESEA intervention earliest must be similar to pre-intervention outcomes at sites where the intervention is rolled out later. Further, there must be at least three different points at which implementation occurs for different sites. Claimants must not be able to choose a site based on knowledge of whether the intervention has yet been (or will be) implemented there. Finally, to meet the CLEAR requirement that timing of rollout cannot be based on any factors that might be related to outcomes, states and their evaluators cannot select which sites deliver the intervention being tested based on staff capacity, experience, or past performance.

4.5 How Does a State Agency Decide Which Impact Evaluation Design Is Right for Its Intervention?

The previous two sections summarized the design and implementation considerations for experimental and quasi-experimental evaluations, respectively. As discussed, there are inherent strengths and limitations for each design type. Additionally, the appropriateness of each design type will vary based on the research questions, logistical considerations, and likely sample in each state agency. Determining which impact evaluation design is best suited for evaluating its RESEA intervention is one of the most important decisions a state agency will make about its evaluation.

When making this decision, considerations should be given to the feasibility of conducting each design type**.** For a variety of reasons, it may be impractical or impossible to seriously consider both experimental and quasi-experimental types. Section 4.4, included a list of questions and a basic decision-making tree (Exhibit 4-5) to help state agencies and their evaluators determine whether a QED is an option. If it is, an additional set of considerations should then be reviewed to further shape their decision. Exhibit 4-10 describes some of these considerations.

Exhibit 4-10. Important Considerations for Selecting an Impact Study Design

Questions to Consider	Experimental Studies	Quasi-Experimental Studies
How confidently can this design type detect the impacts of our RESEA intervention?	High level of confidence. Well-designed and well-executed experimental designs can demonstrate that the RESEA intervention caused the impacts on claimant outcomes. Experimental studies are eligible to receive CLEAR's High causal evidence rating.	Moderate level of confidence. Well-designed and well-executed QEDs can demonstrate that the RESEA intervention was associated with the claimant impacts, but cannot totally eliminate the possibility that outcomes changed because of some other factor. Most QEDs are only eligible to receive CLEAR's Moderate causal evidence rating (at most).
What sample size is needed to detect impact using this design type?	Sample required is smallest of the design types. Depending on the intervention tested and outcomes measured, your state will need to randomly assign several thousand to tens of thousands of claimants.	Sample required is larger, often several times larger, than for experimental designs.
What level of effort is required to implement this design type?	Level of effort is high. Implementation usually requires training staff on study procedures, monitoring random assignment and fidelity to the intervention, and other logistical and managerial tasks.	Level of effort is low to moderate. For retrospective QEDs, evaluators will need to collect available data. For prospective QEDs, evaluators may need to implement study procedures, such as setting and monitoring the selection cutoff for a regression discontinuity study or staggering implementation for an ITS study.
How difficult is the analysis associated with this design type?	Analysis requires staff with relevant statistical expertise, but is less complex than QED analysis.	Analysis is complex and iterative requiring skilled statistical staff.
Can results of this design type be easily understood?	Most experimental study results are easy for many different audiences to understand.	Some QEDs use complex analytical tools whose results can be challenging for the lay reader to understand. State agencies interested in using QEDs should discuss how to make QED findings as easy to understand as possible.
Are quick results needed?	Experimental studies can take 4-5 years to produce findings.	Retrospective QEDs, which use data you have already collected, can produce results in 2-3 years. Prospective QEDs produce results in 4-5 years.

If the state agency decided that a QED is appropriate for its evaluation after thinking through the above considerations, the state agency will also need to decide on which specific quasi-experimental design to use. Exhibit 4-11 compares the three QED types described in this chapter.

Exhibit 4-11. Considerations for Assessing Quasi-Experimental Designs

Questions to Consider	Matching	Regression Discontinuity	Interrupted Time Series
Can this design create a credible comparison group for RESEA?	Infrequently. Only in cases where the selection occurs such that there are claimants who did not receive the intervention who are very similar to those who did.	Yes, if selection was made using profiling scores.	Yes, but only if sequencing of intervention rollout is not related to things likely to influence outcomes, including staff capacity.
Can this design test the whole program and/or particular components?	Yes.	Retrospective design: No. Prospective design: Yes.	Retrospective design: Likely no. Prospective design: Yes.
What sample sizes are required for this design?	Larger than for experimental design.	Several times larger than for experimental design.	Much larger (usually several times larger) than for experimental design.

A final consideration that state agencies should remain aware of is for which CLEAR study quality rating(s) each impact study design is eligible. CLEAR has rigorous standards that it uses to assess the quality of a study's evidence. Exhibit 4-12 provides an overview of the highest possible CLEAR causal evidence rating of the quality of a study's evidence for which each impact study design is eligible.

Exhibit 4-12. Highest Possible CLEAR Causal Evidence Rating, by Study Design

Design Type	Highest Possible Rating
RCT	High
QED – Interrupted Time Series	High
QED – Matching	Moderate
QED – Regression Discontinuity	Not Yet Rated by CLEAR

Note: CLEAR does not have formal standards for regression discontinuity designs at this time, but recognizes that it is a strong design for measuring impact. CLEAR currently relies on the standards developed under the What Works Clearinghouse for regression discontinuity designs.

4.6 Combining Impact Studies with Other Study Types

State agencies may choose to combine several study design types in their evaluation to maximize what they might learn about the intervention they are testing. For example, an evaluation of a new intervention might benefit from an implementation study and an outcomes study plus an impact study (see Chapter 5). The implementation study will document program operations and may produce important insights which help evaluators and readers contextually interpret the corresponding impact study's results. Similarly, an outcomes study might collect and analyze program performance in the near term (e.g., a snapshot of the intervention's ability to meet targets at a point of time) or compare a new intervention's ability to meet outcomes to more established programs.

Conducting supplemental studies can more completely determine the potential of a particular program or intervention and provide credible information to improve or redesign it. In particular, different types of designs are more appropriate for collecting different types of data, and no one design type can provide a complete picture of an intervention. An outcomes or impact study is necessary for a quantitative accounting of program outputs, whereas an implementation study provides contextual data on experiences, structures, and decision-making.

The types of studies included in an evaluation are also determined by the capacity and experience of the program and the evaluator. State agencies with the capacity to do so could conduct small outcome studies routinely at relatively modest cost. In contrast, an intervention that has been previously studied or extensively piloted may be ready for impact study testing. Such an evaluation would include a study of the effect of the program on long- and short-term outcomes.

5. Selecting an Evaluation Design: Outcome and Implementation Studies

Chapter 3, discussed several types of research questions that may be of interest for RESEA evaluations. Then Chapter 4 described how impact evaluations can be used to answer research questions about the effectiveness of RESEA interventions. Now, Chapter 5 discusses designs for answering outcomes and implementation questions. Questions addressed by these types of studies include understanding the demographic or economic context in which the RESEA intervention (i.e., the whole program or components of the program) operates, the process by which claimants flow through the intervention, how those claimant flows differ across American Job Center (AJC) locations, how RESEA and Workforce Innovation and Opportunity Act (WIOA) activities are coordinated, or what happens to claimants after they exit the program. A state agency's general interest in how its RESEA intervention operates, why it operates as such, and how it performs motivates the types of studies discussed in this chapter.

Outcomes studies typically try to understand operational results or claimant outcomes following participation in the RESEA intervention.[27] In the RESEA context, key outcomes include employment and earnings. As such, an outcomes study of an RESEA intervention might analyze whether claimants participating in the RESEA intervention enters employment or attains certain levels of earnings. **Implementation or process studies** can assess a range of questions including the level of resources used to conduct the intervention, how the intervention operates, or how many claimants receive various kinds of services. As noted in Chapter 1, the toolkit uses the terms "implementation study" and "process study" broadly and interchangeably. These terms refer to studies about the implementation of a program and also aspects related to it, such as studies of participation patterns.

Often, a comprehensive evaluation will combine impact, outcome, and implementation analyses to answer questions about how an intervention works, what outcomes a claimant is experiencing, and the extent to which the intervention is causing those outcomes. By providing a detailed account of the intervention and its impacts, this more comprehensive type of evaluation can build evidence about the effectiveness of a particular intervention and assist in its replication. Though outcome and implementation studies can provide important insights, they *cannot* generate causal impact evidence—the kind needed to meet the evidence requirements called for in the Social Security Act (see Chapter 4 for more information on impact studies, which can meet these requirements).

After reading this chapter, state agencies will know more about:

- ✓ Outcomes studies and
- ✓ Implementation studies.

5.1 Outcomes Studies

An outcomes study measures intervention participants' outcomes and how those outcomes compare to expectations or established targets. Evaluators conducting outcomes studies can analyze the observed characteristics of claimants and assess those characteristics against

[27] For a lengthier discussion of approaches to implementation and outcome research, see: Werner, A. (2004). A guide to implementation research. The Urban Institute.

intervention goals, across intervention implementations or locations, or over time. For example, a common challenge for the RESEA program is that a significant percentage of claimants do not appear for their first scheduled meeting. Though many state agencies are aware that this is a problem, they may not have quantitative information to explain the magnitude of the problem. State agencies could conduct an outcomes study using retrospective data to identify all claimants who were scheduled for a first meeting but did not appear at the meeting, looking at variations across locations and time, and compare these to their target goals.

CLEAR's Descriptive Study Review Guidelines

CLEAR uses the descriptive study review guidelines to review quantiative studies that do not attemp measure the causal effects of a program or policy. Based on CLEAR's definition, descriptive studies include outcomes studies. While CLEAR's review of descriptive studies do not culminate in a rating, the review guidelines are still valuable for understanding the characteristics that contribute to a high quality descriptive study. More information about CLEAR's descriptive study review guidelines can be found at: https://clear.dol.gov/reference-documents/quantitative-descriptive-guidelines

State agencies may find outcomes studies valuable when trying to assess whether the intervention is making progress toward its intended outcomes and obtain feedback on their program's design or implementation. The kinds of outcomes typically of interest to an RESEA intervention are those labor market outcomes that the program intends to improve, namely employment and earnings. For example, an outcomes study focused on claimants' entry into employment could compare employment rates, time to employment, or earnings against performance benchmarks over time. The findings may prompt a state agency to study a particular intervention component or design feature further, perhaps through an implementation study (described below) to understand the inner workings of the intervention, or through an impact study to estimate the intervention's effects on the outcomes of interest. State agencies may be interested in other outcomes, as well.

Exhibit 5-1 (at the end of this chapter) discusses two broad types of outcomes studies. **Longitudinal studies** analyze outcomes over time. They can analyze the extent to which certain claimant-level characteristics change over time, such as following participation in the RESEA intervention. Such studies can also analyze the extent to which aggregate outcomes (e.g., average UI duration) for groups of claimants change over time. Specifically, pre-post analyses can examine how claimant outcomes at the start and end of the program differ. **Cross-sectional studies** analyze outcomes for a specific point in time. This type of study typically compares outcomes (e.g., employment and earnings) against pre-determined performance benchmarks or across geographic locations within a state.

To conduct a high-quality outcomes study, states need reliable and accessible data on outcomes. For a longitudinal study, data will preferably be measured at the individual claimant level at both baseline and at a pre-determined interval after enrollment in the RESEA intervention (e.g., 3 months or 6 months after enrollment into RESEA). A cross-sectional study may only require data from a specific point in time, which will vary depending on the questions of interest. To conduct a pre-post analysis, the intervention would have been implemented in a prior period, and the outcomes data should be available from both at the start of the program and after the program. Typically, the outcomes of interest for RESEA evaluations are available in state administrative data sources. Chapter 6 discusses data quality and accessibility in greater detail.

Though valuable for certain purposes such as analyzing performance measures, outcomes studies do have certain limitations. Outcomes studies are not designed to tell how *effective* an RESEA intervention is, that is, a state agency will not be able to make causal claims about their RESEA intervention's impact on claimant outcomes based on findings from an outcomes study. While an

outcome study provides information on claimants' outcomes, it cannot provide an estimate of how those outcomes differ from what they would have been were it not for the intervention—which is what is required to estimate *impact*. Unlike impact studies, outcomes studies do not have a comparison group that did not receive the intervention and to which the outcomes for the intervention group can be compared. Without an estimate of what outcomes would be in the absence of the intervention, outcomes studies cannot claim to estimate how the intervention changed outcomes. For example, a pre-post analysis may show increases in employment, but this does not indicate impact because there would have been some improvement in employment outcomes expected even without the RESEA intervention.

5.2 Implementation or Process Studies

Implementation studies analyze "what happened and why." Such studies examine the program's design, implementation context, administration, and operational processes. An implementation study typically answers questions about the extent to which a program operates as planned and whether the program reaches the intended target population(s) with intended services. In the RESEA context, an implementation study's findings are particularly useful for determining whether an intervention is being carried out in a manner consistent with its goals, design, or other planned aspects. Implementation studies can also provide valuable insight into the organizational, social, and economic contexts in which the intervention operates. Additionally, an implementation study can examine program implementation from the perspective of RESEA claimants or RESEA staff, including adherence to program rules and data entry practices. U.S. DOL has established guidelines for designing a high quality implementation study through its Clearinghouse for Labor Evaluation and Research (CLEAR).[28]

CLEAR Implementation Study Review Guidelines

CLEAR's review guidelines for implementation studies focus on determining the appropriateness of the study design, sampling strategy, data sources, data collection, and analysis for addressing the research questions. States may find CLEAR's implmentation study review guidelines helpful for understanding the elements that promote high-quality implmentation studies. More information about CLEAR's implementation study review guidelines can be found at: https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0

Implementation studies typically involve systematic collection of both qualitative and quantitative data. Possible data sources include interviews with stakeholders (e.g., intervention staff, claimants, and community partners), focus groups with stakeholders, intervention observations, and collection of programmatic documents and data. Interviews and focus groups can provide insights into staff's experiences operating the intervention and claimants' experiences within the intervention. For interviews and focus groups, state agencies and their evaluators will have to consider who will be included, why it is important to include them in the study, and how they will be recruited to participate. Implementation studies may also draw on intervention administrative data, which can provide information to answer questions about how the intervention functions in practice, including how claimants use (or do not use) services and length and duration of services.

Depending on the study design, evaluators may choose to use a statistical sampling method (such as random sampling) or purposive sampling (i.e., based on specific criteria) to select which sites (e.g.,

[28] Clearinghouse for Labor Evaluation and Research (July 2014). *Operational Guidelines for Reviewing Implementation Studies.* https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0

WDB or AJC offices) to include in the study. The evaluator and state agency will also have to decide which stakeholders (e.g., RESEA intervention staff, claimants, and state staff) will be included in the study. Using a statistical sampling strategy can help to increase the likelihood that the data collected from the study sample are representative of the larger group about which the study would like to draw conclusions. For example, findings from an implementation study about an RESEA intervention's design and operations which only collected data from a particular region might miss important variation in services that are occurring elsewhere. For this reason, findings from that study may not be representative of the state as a whole.

As noted earlier, implementation or process studies broadly refer to studies about the implementation of an intervention and aspects related to that implementation. Studies that focus on specific aspects of implementation include fidelity studies and participant flow studies. **Fidelity studies** examine whether intervention operations are consistent with legislative intent, regulations, agency objectives, and/or the intervention model. For example, a fidelity study of an RESEA intervention might examine how the state agency targets services to appropriate claimant populations and whether those efforts yield a pool of intervention claimants that aligns with the intervention's goals. **Participant flow studies** examine how participants progress through the various stages of a program. In the case of RESEA, a participant flow study would involve tracking claimants as they move through various stages of the RESEA intervention (e.g., scheduling their RESEA meetings, arriving at the AJC and completing a meeting). Exhibit 5-2 (at the end of this chapter) lists and describes these types of studies that examine aspects of implementation.

While fidelity and participant flow studies emphasize different aspects of program implementation, these studies are not necessarily mutually exclusive. State agencies may find it helpful to combine different aspects of the study types in their implementation study. For example, an implementation study of an RESEA intervention could focus on the operational processes and how claimants move through the RESEA intervention (a participant flow study*)* and how that is related to adherence to the intervention's logic model (a fidelity study). In general, implementation studies may be conducted alone or conducted to complement other parts of an evaluation. State agencies may find that implementation studies provide valuable contextual insight when interpreting results from outcomes or impact studies.

However, while implementation studies can provide valuable insights for understanding the RESEA intervention, they also have limitations. Implementation studies cannot answer questions about the impact of an intervention nor about the mechanism behind any observed outcomes. Whereas a well-executed impact analysis (as described in Chapter 4) can tell state agencies about the extent to which their interventions improved claimant outcomes, an accompanying implementation study cannot attribute those impacts to any particular element of the intervention. State agencies and their evaluators can only use implementation studies to provide some suggestive explanations for outcomes or impacts. Another limitation of implementation studies is their external validity or the degree to which the collected and analyzed data are representative of other contexts. Well-executed implementation studies can have internal validity but very limited external validity (i.e., information may not be extrapolated to other contexts).

Exhibit 5-1. Types of Program Evaluation – Outcomes Studies

Purpose: An outcomes study compares individual participant outcomes against goals, across programs or locations, or over time.

Data Requirements: All outcomes studies require individual-level outcome data on the population or sample of interest. Longitudinal outcomes studies may need baseline data on claimant outcomes and characteristics at program enrollment (baseline) and then follow-up data after program enrollment.

Type & Description	Purpose and Potential Value	Cost, Type of Evidence Generated, and Other Considerations
LONGITUDINAL STUDY Analyzes outcomes over time, in some cases for specific cohort(s) of individuals.	• **Trend analysis**—By tracking participant outcomes over an extended period of time, analyses can identify patterns of changes in program outcomes. For example, analyses can track improvement or worsening in employment outcomes at the state or sub-state level. • **Pre-post analysis**—Pre-post analyses require longitudinal data that cover the period before and after the event under analysis, be it participation or implementation of the RESEA intervention. At the individual level, pre-post analyses can examine how claimant outcomes change during participation in the intervention. These analyses are most appropriate when considering outcomes that can be measured during both periods and for which you can expect stability over time. For example, changes in claimant well-being, though not a primary outcome of interest, could be measured following participation in the RESEA intervention. Alternatively, states could consider pre-post analyses that examine changes in outcome levels for repeated cross-sections of claimants. For example, if a state expanded the RESEA intervention operations to a new geographic area, the state could analyze changes in UI duration among RESEA-eligible claimants in that geographic area before and after intervention expansion. In some instances, pre-post outcomes studies may be a first step in developing an evidence base for a new intervention, laying the groundwork for a later, more rigorous impact evaluation that can attribute changes in outcomes to the intervention.	• Can be more expensive than studies providing point-in-time outcome estimates because data needs to be collected at multiple time points. • Can be expensive to collect pre-intervention data, but intervention participation and short-term outcome data are generally collected as part of program performance tracking and would therefore pose fewer cost barriers. • Viewed (pre-post analysis) as a practical way to look at changes in outcomes before and after an intervention. • Can show whether the RESEA intervention is meeting its target goals. • Cannot provide estimates of program impacts.
CROSS-SECTIONAL STUDY 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.	• Cross-sectional analyses could include comparisons of outcomes across groups or locations. This may be appropriate to identify under- or over-performing implementations of the intervention. • Cross-sectional analyses could include performance assessments, in which outcomes for a particular AJC or state are assessed relative to performance objectives (e.g., employment retention). This may be appropriate to ensure that the RESEA intervention is meeting its goals at a particular point in time.	• Considered inexpensive to conduct, given that data are readily available through routinely collected program performance data. • Provide insights about RESEA intervention and how outcomes vary across locations or groups of claimants. • Do not provide estimates of intervention impacts.

Exhibit 5-2. Types of Program Evaluation – Implementation or Process Studies

Purpose: Studies of program implementation document how a program or intervention operates in relation to its theory of change, goals, across locations, or over time.

Data Requirements: A defined set of qualitative and quantitative program- and participant-level data related to the topics covered in the study. An understanding of potentially available interview subjects and documents and the types of information that can be gathered from them.

Type & Description	Purpose and Potential Value	Cost, Type of Evidence Generated, and Other Considerations
IMPLEMENTATION STUDY (GENERAL) Documents program operation or compares it against goals, across locations, or over time.	• 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 context, administration, and operational processes of programs. • Can be done across all sites implementing a program, within a purposive selection, or for a representative sample. • May focus on examining program operations and/or program outcomes at one location. Typically conducted when a program is unusual (innovative and high-performing, or troubled and poor-performing). • Performance analyses assesses program outputs against performance goals. Assessments could extend over time and may involve assessing program improvement. Generally used for new programs focusing on a single program in one place or a few locations. May be appropriate for state agencies with limited prior experience providing the reemployment services featured in RESEA interventions.	• Can incur fairly high labor costs for site-based information collection. Costs depend on methods used, number of sites or stakeholders included as part of study sample, availability of pertinent data, and whether the evaluator is local, and type of analysis conducted. • Cannot assess impact. • Can be used to inform future program development or replication by describing program features that did or did not work well. • Can provide a valuable first step before conducting a more rigorous evaluation (e.g., impact evaluation).
FIDELITY STUDY Assesses whether program operations are consistent with legislative intent, regulations, agency objectives, and/or program model.	• Determines the extent to which program operations adhere to the program model as it was originally designed. • May be appropriate for studying adherence to RESEA eligibility determination or faithful implementation of a previously evaluated reemployment service.	• May be less expensive to conduct, given that data are readily available through routinely collected program performance data, though cost may increase depending on number of locations included in the study and the volume of data to be analyzed. • Cannot assess impact. • Can explain how closely on the ground program operations and practices follow the program model and help RESEA program administrators address variation in how staff are implementing the intervention across sites.
PARTICIPANT FLOW STUDY Examines how participants progress through the different stages of a program.	• Addresses intervention participation by following a cohort(s) of claimants through intervention stages (application, enrollment, assessment and/or orientation, program activities). • Analyzes measures such as enrollment or participation rates, proportions of enrollees who participate in certain kinds of activities, duration, and completion rates. • May be appropriate for tracking UI claimants' progress through the various stages of RESEA intervention services.	• May be less expensive to conduct, given that data are readily available through routinely collected program performance data, though costs will vary based on design. • Cannot assess impact. • Can explain how claimants progress through the RESEA intervention and can help identify areas of improvement.

6. Data Needs for RESEA Evaluations

All evaluations use data to document intervention operations, claimant characteristics, intervention services, and claimant outcomes. For that reason, high-quality and complete RESEA-relevant data are a critical component of the evaluation.

Data needs are determined by the research questions and evaluation design type. The evaluator will be responsible for most data-related evaluation tasks. However, state agency staff overseeing the evaluation and other members of the state agency's evaluation team will play a proactive role in identifying required data, assessing its suitability, and ultimately obtaining it for the evaluation. The information discussed in this chapter will help state agencies better provide this support by:

- ✓ Describing data types, sources, and other data-related concepts.
- ✓ Determining how to use data for evaluation.
- ✓ Reviewing considerations for developing a data collection plan.
- ✓ Becoming familiar with best practices and common challenges.

6.1 Understanding Data Concepts

"Data," refers to any information collected for the purposes of conducting an analysis of an intervention. Evaluations of reemployment services interventions often rely on multiple types of data from multiple sources. This section defines the data concepts and discusses the types and sources of data the evaluator will likely need in order to conduct an evaluation of an RESEA intervention.

Types of Data

Evaluators commonly divide data into two broad types: quantitative and qualitative.

Quantitative data are those data that are numerical (e.g., the number of intervention claimants, the dollar value of post-intervention earnings). The approach to analyzing quantitative data often relies on mathematical or statistical techniques. Most administrative data—that is, data usually collected in the regular administration of the intervention—are quantitative in nature or are coded to be so.[29] Typical analytic strategies for using quantitative data involve using statistical analysis software (e.g., SAS, STATA, SPSS, Python) to review, clean, and analyze the data, according to the methods decided upon in an Evaluation Design Report or other analysis plan (described in Appendix I).

Qualitative data are not readily captured with numbers, predefined categories, or scales. As such, qualitative data can often be richer in detail, particularly when presented in text or narrative form (e.g., descriptions of program operations, stakeholder perceptions of the program). Analysis of qualitative data relies on techniques typically requiring some sort of structured interpretation by the evaluator. For example, a claimant's responses to an in-person interview can be qualitative in nature, and an evaluator could review multiple responses to identify key themes. Using certain kinds of analytic software (e.g., NVivo, Atlas.Ti, Dedoose), the evaluator can review responses to in-person interviews, identify common themes, and report on patterns or relationships between these

[29] For example, the name of the industry claimants most recently worked in (e.g., healthcare or manufacturing) is qualitative information, but most states have assigned each industry a numerical code. Meanings of each coded value may be contained in a data dictionary that is maintained by the state's IT department.

themes. Analysis of these themes can enhance understanding of the program under evaluation, which offers additional context when interpreting the results of quantitative analyses conducted, for example in an impact analysis.

Sources of Data

Data (quantitative or qualitative) can come from multiple sources including administrative records, interviews, focus groups, case notes, and surveys. For example, to answer research questions related to earnings, evaluators might access state employment and wage records. To answer questions about the usefulness of job search services, evaluators might contact a claimant directly to ask them about the types of job search services they accessed and their perceptions of the services.

The four data sources typically used for an evaluation of RESEA interventions are administrative; interviews, focus groups, and observations; document review; and surveys. Though evaluators may use data from many different sources, the richness of employment services and UI data suggests that administrative data are an excellent source of data for teams evaluating RESEA interventions. Exhibit 6-1 lists the pros and cons of relying on each source, which are further discussed in the sections that follow.

Exhibit 6-1. Pros and Cons of Data Used for Evaluation, by Source

Data Source	Pros	Cons
Administrative records	• Cost-effective • Can be specific to the intervention • Can be used to answer employment and earnings questions	• Require expertise from data manager • Only include data collected through intervention administration or maintained by other state/federal agencies (e.g., earnings data)
Interviews, focus groups, or observations	• Versatile • Captures nuance and detail • Can highlight important lessons about implementation and perception of the program	• Subjective • Require new data collection instruments • Can be time-consuming to collect and analyze
Document reviews	• Inexpensive	• May not reflect actual implementation
Surveys (of claimants, staff, etc.)	• Source for data not otherwise available	• Costly to administer • Require new data collection instruments

Administrative Records

Administrative records data refers to information routinely collected as a part of the regular administration of program or intervention activities. State agencies (e.g., employment services, UI, employment security agencies) often record information about the administration of their programs. The information collected by those agencies can be used to facilitate RESEA evaluations.

For evaluations of RESEA interventions, administrative data would likely be collected on:

- **UI claims and benefits.** State UI agencies maintain administrative data on each claim as well as data on subsequent payments to those claimants. State UI agencies collect information on claimant demographics and characteristics, which evaluators may use for a variety of purposes. For instance, they may use the data to conduct descriptive analyses of claimants who engage with RESEA services or to develop matched comparison groups for certain QEDs (as described in Chapter 4). The amount of benefits paid to UI claimants is a

key outcome variable for evaluations of RESEA, so evaluators may opt to collect administrative data on benefits paid in order to measure the extent to which participation in the RESEA intervention has an impact on UI benefit duration.

- **Participation in program services and activities.** To the extent that states maintain administrative records of employment services offered and/or delivered to claimants selected to participate in RESEA, evaluators may use those administrative data for several purposes. Data on services can inform analyses of intervention implementation to understand claimants' experiences in the intervention. For experimental impact evaluations, evaluators may also use claimant data from the RESEA management information system to monitor implementation of study procedures (e.g., random assignment process) and to compare the level of participation in services and activities by intervention and comparison group members.
- **Employment and earnings.** Impact evaluations of RESEA interventions will need to report on whether the intervention improves employment outcomes for claimants. State UI agencies typically offer one common source for data on these outcomes. To administer the UI program, state UI agencies collect data from employers on quarterly wages paid to employees. Evaluators may work with state UI agencies to collect those quarterly wage data in order to measure employment and earnings impacts from the RESEA intervention.[30] Another potential source of data on employment and wages is the State Wage Interchange System (SWIS) Data Sharing Agreement, which allows states signing the Agreement to share quarterly wage records with one another.[31]
- **Claimant characteristics.** As discussed above, state UI agencies typically collect information on claimant demographics and other characteristics. In addition, state agencies that administer RESEA may collect additional information about claimants that may be useful in an evaluation, such as claimants' profiling scores, level of education, employment history/industry, and scores on skills assessments. These data may be used to develop additional analyses, for example to detect whether there are differential impacts of an intervention according to particular claimant characteristics.

Interviews, Focus Groups, and Observations

Evaluators may engage with RESEA program administrators, staff, or claimants directly to collect data through in-person, telephone, or virtual interviews and observations. Evaluators often collect qualitative data through these sources using a technique called purposive sampling, in which evaluators use a specific criteria, such as claimant demographic characteristics or RESEA staff who perform particular functions, to select an appropriate sample to include in the data collection. For example, evaluators may select a sample of RESEA staff and ask them about their procedures and practices during a semi-structured conversation. If evaluators are permitted to observe RESEA intervention service delivery, they may document the observed services using a structured data collection or observation instrument. Typically, evaluators use these kinds of data to describe

[30] One limitation of these state UI wage records is that they often are not available for all claimants included in an evaluation. For example, claimants who obtain employment out of state are not included in a state's wage records; federal, military, and contractor employees and the self-employed also are not included.

[31] Experience suggests that it can be very challenging and time-consuming to obtain SWIS data sharing agreements. States that are interested in obtaining SWIS agreements should contact their Federal Project Officer to learn more about the process and limitations around it.

program implementation, which could be useful for understanding the results of separate impact analyses or for comparing RESEA intervention implementation across locations.

Document Review

To collect data on RESEA intervention design and client flow, evaluators may review program documentation to extract key information. For example, states' RESEA grant applications and plans or procedure manuals often describe several aspects of their intervention design. Through systematic review of these documents, evaluators can supplement analyses of intervention implementation.

Surveys

In some cases, an effective way to collect data for a program evaluation is to ask key informants to complete a survey. For example, to understand how an RESEA intervention is implemented, evaluators may create a list of implementation-related questions and ask RESEA staff to respond in an online survey. Similarly, evaluators could survey claimants who were assigned to RESEA services to learn more about their experience with the RESEA intervention. Surveys could even be used to measure outcomes in an impact study by collecting information about outcomes of interest, such as information about claimants' employment status and wages. Almost any kind of information could be collected using a survey, but there are several drawbacks to that approach:

- **Cost.** Survey data are often costly to collect. A survey requires development of a survey instrument, identifying a sample of respondents, and diligent administration of the survey and follow-up with respondents. Surveys of RESEA claimants may be particularly costly.
- **Data quality.** Surveys do not always generate high-quality data for certain types of information. For example, when respondents are asked to answer questions about events that occurred in the past, they may not remember those events accurately. In addition, respondents may not answer every question in the survey.
- **Response rate.** It can be challenging to get respondents to begin and complete surveys. Likely response rates vary with context, but they are often quite low. Simple requests to respond do not typically yield strong response rates, and persistent phone efforts or in-person follow-up might be required to raise response rates to a sufficient level. Many researchers dedicate significant resources to incentives that will increase response rates. If the evaluator plans to use survey data to measure program impacts, they will need to ensure that both the intervention and comparison groups respond at comparably high rates. It can be particularly challenging to bolster comparison group response rates, as they did not receive the intervention and may be less familiar with the study. Surveys with low response rates can suffer from non-response bias, in which responders to the survey differ in important ways from non-responders. This bias can be diminished with the use of particular weighting methods, but this will usually require the assistance of a statistical consultant. Even still, the degree of non-response will likely affect the validity of the survey results.

As previously noted, surveys may be an effective way to collect evaluation data. However, given the considerations above, strategies based wholly or largely on surveys are often unattractive to evaluators, particularly those conducting impact evaluations.

Other Data-Related Concepts

When working with data for an evaluation, there are several other key concepts the state agency will want to understand before collecting the data of interest for an evaluation:

- **Individual-level data versus aggregate-level data.** Data can be collected at different levels. When evaluators collect data on a particular topic (e.g., employment), that information can be reported either at the *individual level,* meaning separately for each claimant (e.g., individual employment status), or at the *aggregate level,* meaning combined for a group of claimants (e.g., average employment rate). State agencies interested in conducting impact evaluations will need to collect individual-level data in order to determine the intervention's impacts on claimant outcomes. Outcomes and process studies often rely on aggregate data to describe in a more general fashion the program's operations and its ability to meet targets.
- **Personally identifiable information (PII).** Both qualitative and quantitative data can contain PII, meaning information that could be used to identify an individual person, such as a claimant's name, address, or Social Security number. For example, if an evaluation collects an administrative data set that includes RESEA intervention claimants' earnings and SSNs, then the data are said to be "personally identifiable." Evaluators are obligated to protect the identities and privacy of the persons from whom data are being collected.
  - One way to do that is to "de-identify" the data, by assigning a unique arbitrary number ("identifier") to each person that replaces their PII. Continuing the example from above, with the data set that includes wages and SSNs, the evaluator could create a new data element that assigns such an identifier to each claimant and then delete the SSNs from the data file. The new data set is said to be "de-identified". The evaluator can use the new data element to differentiate each intervention claimant and his or her wages. Another way to protect claimants' identities and privacy is for the evaluator to maintain a single file that contains both the PII and unique identifiers on a secure server to which only a limited number of people on the evaluation team would have access. The evaluator still would be able to use this file to link the PII to the wage data, if needed.
  - Beyond direct identifiers, the study might consider other variables that could lead to identification, particularly in relatively smaller samples. In particular, the evaluator should consider whether it is necessary to give exact dates (e.g., date of birth). Often there is little loss from deleting these data elements or making them less specific (e.g., removing the specific day and/or month). In general, such decisions involve careful balancing of potentially weaker evaluation results and the risk of disclosure. Consulting with an expert in this area is often appropriate.
- **Baseline data**. The term "baseline" data refers to information collected at the point at which RESEA-eligible claimants are enrolled in the study but before they receive any of the services that are being evaluated. A claimant enters the study at the time that selection for the RESEA intervention occurs. Baseline data provides a snapshot of the claimant's characteristics and is often the data used to create comparison groups in QEDs. Workforce evaluations typically collect information on study participant demographics (e.g., age, race, and gender), employment history, education history, and prior public benefits receipt (including UI receipt, duration of benefits, and benefit amount). Evaluations of RESEA interventions can likely obtain much of this data from claimants' UI applications.
- **Follow-up data.** The term follow-up data refers to information collected about outcomes after the claimant enters the study. To make comparisons across claimants, evaluators must collect data uniformly for all study participants (i.e., claimants in the intervention group <u>and</u> the comparison group) and for a specified length of time after the claimant entered the study. The length of time over which outcomes are measured is typically referred to as the

"follow-up period." Measuring outcomes for the same length of time for all claimants allows evaluators to (1) aggregate claimant outcomes data to estimate an average effect of the program, (2) make meaningful comparisons between claimant outcomes based on their study group (intervention versus comparison), and (3) discuss study findings in a way that is easy for the audience to understand. If evaluators did not collect follow-up data for a uniform time period, differences observed in the outcomes of claimants could be caused by factors other than the intervention. For example, if an evaluator measures the earnings of Claimant A after six months and Claimant B after 12 months, Claimant B might have higher earnings because they had the opportunity to earn wages for a longer period of time or because they had more time to find work, not because the services received by Claimant B were more effective.

It is important that the evaluator define the follow-up period as the length of time measured from when the claimant is selected for the RESEA intervention rather than from when the claimant exits UI. This definition is crucial so that meaningful comparisons can be made between RESEA and non-RESEA claimants. For example, suppose that an RESEA claimant exits UI after one month and a comparison group (non-RESEA) claimant exits after five months from RESEA selection; both then remain employed for several years. In this example, the RESEA claimant's outcome is better, in having returned to work more quickly (UI duration=1 month vs. 5 months). But if the outcome is measured as employment since UI *exit* and not selection for RESEA, the two claimants would appear to have similar outcomes because both are employed at any given number of months after exit. Translating this example to a higher level, if an intervention is effective in returning RESEA claimants to work more quickly than non-RESEA claimants, measuring outcomes from the point of exit from UI, rather than selection for the RESEA intervention, will make the RESEA intervention look less effective than it actually is.

6.2 How Can Data Be Used for Evaluation?

Chapter 3 reviewed how to develop research questions for an evaluation. The research questions will inform the kinds of analyses that an evaluator will need to perform, and thus what type(s) of study the evaluator will conduct. This section begins by briefly describing what each type of analysis entails and common approaches to the corresponding data collection.

Implementation or Process Analysis. To understand how the RESEA intervention operates, an evaluator may want to analyze the intervention's design, characteristics of claimants, procedures used by state agency staff, interaction with other interventions, or the claimant's experience. For example, an implementation study may explore the services most commonly delivered to claimants or at which stage of the intervention claimants are likely to exit. From the kinds of data sources listed in the previous section, an evaluator may need to collect:

- Administrative data from state UI claims records and/or the RESEA program. Examples include claimant demographics from UI claim records, employment services offered and delivered, and attendance at RESEA meetings.
- Qualitative data from intervention observations or interviews with intervention staff and/or claimants.
- Data extracted from review of intervention design documents.
- Survey data collected from intervention staff and/or claimants.

Outcome Analysis. To assess an intervention's performance against outcome goals, such as labor market outcomes, the evaluator will need to collect data on individual-level intervention claimants' outcomes. The evaluator may want to analyze an interventions performance at a key point in time, changes over time, or comparisons across locations. From the kinds of data sources listed in the previous section, an evaluator may need to collect:

- Administrative data on claimant outcomes (e.g., state UI duration information, service engagement records).
- Survey data on claimant outcomes.

Impact Analysis. To understand how an RESEA intervention affects claimant outcomes, the evaluator will need to collect individual-level data for all claimants enrolled in the study. Study participants will include claimants who received the intervention (i.e., the intervention group) and claimants who did not receive the intervention but are similar enough to intervention claimants that they can serve as comparison group members. For the analysis, the evaluator will compare outcomes intervention group members to those for comparison group members. From the kinds of data sources listed in the previous section, an evaluator may need to collect:

- Administrative data from the RESEA intervention, such as employment service data to compare the services received by intervention group members scheduled for, and who attended, RESEA meetings versus the comparison group.
- Data on claimants' characteristics as of the claim (i.e., at baseline). This is particularly important for QED evaluations to ensure that intervention and comparison group members are similar.
- Administrative data on claimant outcomes, such as state UI wage records to compare employment outcomes among the intervention group members versus outcomes among the comparison group members.
- Survey data on claimant outcomes, such as data on earnings reported by claimants in both the intervention and comparison group.

Basic Data Requirements for RESEA Impact Evaluation

Section 306 of the SSA stipulates that an evidence-based RESEA intervention is one that has been shown to reduce claimants' UI duration as a result of improved employment outcomes.[32] This implies that evaluations will need data on shorter- rather than longer-term employment outcomes. That is, evaluations should focus on employment that occurs soon enough after the start of a claim that it could reduce UI duration. Evaluations will also need data on UI benefit receipt duration.

[32] Social Security Act, 42 U.S.C. § 306 (2018)

The list of data elements in Exhibit 6-2 below represents likely data elements state agencies may choose to incorporate into evaluations of their RESEA interventions. Collection of all these data elements would allow evaluators to perform several different analyses. However, at the very minimum, the following information should be collected:

- Claimant identifiers (i.e., SSN) and characteristics information (e.g., sex, age, gender, occupation, job tenure, base period earnings) from initial claim associated with the study.
- Information for the claim associated with the study, including claim application and approval dates.
- If applicable, claimant random assignment date and random assignment status.
- For the approximately 12-24 months prior to study enrollment, from state administrative data, each claimant's:
  - Prior UI claim information
  - Prior earnings information
- For the two to four quarters after the current claim's approval date, from state administrative data, each claimant's:
  - UI claim and duration information
  - Employment and earnings
  - Services provided
  - Nonmonetary issues raised and denials

Basic Data Elements Required for RESEA Experimental Evaluations

Evaluators will need the following data to conduct data analysis for experimental evaluations:

- Administrative data on UI claim and benefits paid.
- Administrative data on earnngs from state UI agencies.
- RESEA program data, including information on the intervention being tested.
- (For random assignment studies) The group that each claimant was assigned to.

This information would allow a state agency and its evaluator to generate an impact estimate for its RESEA intervention, though having data on services received would be helpful for analyzing the pattern of impacts.

State agencies and their evaluators should note that these data likely come from more than one source. To complete an impact analysis, evaluators would need to link individual-level data from different data sets to create a single analytic file. For this, they would use a unique claimant identifier (e.g., SSN), that is a data element common across the data sets. This merged file would then need to be checked for completeness—evaluators will not be able to use data fields for which large percentages of information is missing.

Finally, it is important to note that the data elements summarized above can support only a basic analysis of UI receipt and employment and earnings outcomes. State agencies, however, may have wider or more precise interests that require additional data. For example, the list of data elements above does not include information on claimants' hours worked or hourly wages. Nor does it include information on job quality, such as benefits, shifts, or guaranteed hours per week. Finally, the minimum requirement is that employment and earnings be recorded quarterly; state agencies may be interested in learning about impacts at the monthly or weekly level.

Identifying Data Elements for an Evaluation

Exhibit 6-2 lists common data elements needed for an evaluation of RESEA interventions. "Data elements" refers to the specific variables included in the analyses conducted for an evaluation. The table organizes these data elements into topic areas, such as claimant characteristics and claim information. For each topic area, the table also identifies the kinds of analyses from Section 6.2 for which these data elements may be applicable and the likely data source(s) from which the data could be collected. This table is a good starting point, but state agencies will need to update it to reflect the information needed to answer their specific research questions of interest.

Exhibit 6-2. Common Data Elements Needed for RESEA Evaluations

Topic Area	Data Element	Applicable Analysis	Likely Data Source
Claimant characteristics[a, b]	• Age • Gender • Race • Prior employment tenure • Prior occupation • Prior industry • Prior earnings • Educational attainment • Region	• Implementation/process analysis of claimants' service receipt • Outcome analysis • Impact analysis	• Program administrative data
Claim information	• Application date • Approval date • Maximum benefit amount • Prior claim dates • Nonmonetary issues • Denial • Denial date	• Outcome analysis • Impact analysis	• Program administrative data
Program services	• Service name • Service referral • Service attendance • Eligibility review results	• Implementation/process analysis • Impact analysis	• Program administrative data • Interview or observation
UI duration	• UI dates • Claim associated with UI dates • UI benefit	• Outcome analysis • Impact analysis	• Program administrative data • Administrative data on outcomes
Employment outcomes	• Employed at follow-up period	• Outcome analysis • Impact analysis	• Program administrative data • Administrative data on outcomes
Earnings outcomes	• Earnings at follow-up period	• Outcome analysis • Impact analysis	• Program administrative data • Administrative data on outcomes
Perceptions of program	• Claimant self-reported perceptions • Staff self-reported perceptions	• Implementation/process analysis	• Surveys • Interviews

[a] Information on claimant characteristics should be taken from the first UI application associated with the study. For example, if a random assignment study is being conducted, the state agency and its evaluator would want to use the demographic information from the UI application closest to the random assignment date so that the evaluator can create a snapshot of the study group's characteristics prior to their receiving services.

[b] Information on prior employment, earnings, and UI claims will help ensure that an equivalent comparison group has been created. Typically, evaluators collect 12-24 months of prior information, though this can vary based on the research questions of interest and the intervention being tested.

6.3 Considerations for Developing a Data Collection Plan

The sections above reviewed the kinds of data sources and data elements typically used in an evaluation of RESEA interventions. In tandem with the development of research questions and an evaluation design, evaluators will identify the data required and develop and execute a data collection plan. This section discusses several topics and decision points that evaluators will need to consider when developing that plan.

Data Availability

A first step for evaluators is to identify the sources from which they will collect each data element. If some required data elements are not available in existing administrative sources, in some cases it may be possible to add new data fields to the state agency's administrative systems to capture that information. For example, a new data field may be added to the state agency's UI claim data system for indicating each claimant's random assignment status. For those data not readily available from existing sources, evaluators may elect to conduct a survey or other primary data collection.

For each data source, it will be important for the state agency and its evaluator to consider the extent to which the collected data fully cover the population to be included in the analysis. In other words, having identified a source for a given data element, are data available from that source for each individual intended to be included in the study? This is not always the case. For example, state UI wage records include earnings data, but these state-level records do not include earnings information for federal or military workers or out-of-state workers. Evaluators may be able to seek out additional data sources to fill these gaps or account for these kinds of limitations in the analysis.

Data Access Requirements

Having identified a viable source for a given data element, the evaluator should consider any challenges in accessing the data. In general, these considerations apply primarily to administrative data sources, which are often subject to restrictive access requirements. In these cases, it is important to identify exactly who will be able to acquire the data needed for evaluation. If a given administrative data source is maintained by a third party or another state agency, the evaluator might require the state agency or RESEA program administrator's support to coordinate data collection with those other entities.

The state agency or evaluator may need to enter into a **data sharing agreement** with whichever entity maintains the data. Data sharing agreements can be negotiated between the agencies that govern the data of interest and external parties that want to use the data. Common elements of these agreements include:

- Specific data to be shared.
- Rules associated with storing, sharing, and using those data.
- Name and information for a data custodian at both parties to the agreement.
- Time period for the agreement.
- Any requirements for destruction of shared data upon completion of the evaluation.

Timing of Data Collection

Once the state agency and its evaluator have identified an available source of data and have acquired access, it is important to ensure that the evaluator can collect desired data elements at the appropriate times. Typically the evaluator will need to collect evaluation data for two time periods: baseline and follow-up. Section 6.1 defined the concepts of baseline and follow-up data. As a

reminder, **baseline data** is information collected at the point when claimants file their UI claim and/or are enrolled in the study, and before they receive any intervention services.

Follow-up data is information collected at a specified time point after the claimant enters the study. As Section 6.1 described, the evaluator will need to ensure that the follow-up data are collected consistently (i.e., for the same length and using the same methods) for all study participants. As discussed earlier, this consistency is critical for evaluators to be able to make meaningful comparisons between claimants.

Evaluators might collect follow-up data at a single or multiple points in time (e.g., at intervention completion, two quarters after enrollment), depending on the research questions and outcomes being examined. For evaluations of RESEA, evaluators will generally need to collect follow-up data on employment outcomes no later than the first two quarters (approximately 26 weeks) after a claimant is selected for RESEA. As specified in Unemployment Insurance Program Letter (UIPL) No. 1-20, RESEA interventions are required to reduce UI duration through increased employment. For employment to shorten UI duration, it must occur before the claim would exhaust.[33]

Data Quality

Not all data sources or data elements will contain data of the quality needed for the evaluation, where "quality" captures completeness, accuracy, consistency, etc. There are two essential factors to consider that erode data quality: missing data and inconsistent data values.

Missing Data

Missingness most often occurs when surveys are used to collect data, but respondents refuse to answer particular survey items or complete the survey at all. Fortunately, RESEA impact and outcome evaluations are likely to rely wholly on administrative data, for which missing values are less common. However, some values may be missing in administrative data, as well. State agencies' UI claim forms may collect applicant characteristics in fields that are optional, which the respondent may elect not to fill in. Counselors may also occasionally fail to enter data about service receipt. If data are missing for a large percentage of claimants for a particular data element, the evaluator may not be able to use that element in the analysis or may have to flag some findings with a caveat reflecting the limitations in the analysis due to "missingness."

For example, suppose the state agency wants to conduct a QED study for which administrative data is needed to create a matched comparison group (an approach discussed in Chapter 4, Section 4.4). Certain baseline data elements will be needed in order to generate that matched comparison group, such as gender, age, and educational attainment. For example, educational attainment is an important predictor of employment outcomes. If data on educational attainment are missing for a large proportion of claimants, it will be analytically challenging to demonstrate that the evaluator has created a reliable matched comparison group. Furthermore, it will likely not be possible to conduct a sub-group analysis of impacts by educational attainment.

Inconsistent Data Values

"Inconsistent" data elements are those that are not defined or reported in the same way for every individual in the data set. Inconsistent data may also include outliers— that is, data points that differ significantly from other values and may consequently interfere with analyses.

[33] https://wdr.doleta.gov/directives/attach/UIPL/UIPL_1-20_acc.pdf

For administrative data sources, two potential scenarios might commonly result in inconsistent data. First, if local workforce staff are required to enter data into an administrative database but they are not properly trained, the resulting data may be inconsistent across staff, making it difficult to use in an evaluation. Second, if a data element is recorded in an open text box, different staff might enter the information into the text box in very different ways (e.g., for educational history: "some college," "15 college credits," "about 1 year of college"). To analyze such data, evaluators would have to read each data entry and then code the entry as a defined value (e.g., "some college but no degree"), requiring expensive processing time. It is unlikely that data entered in open text boxes would be included in an evaluation.

Cost of Data Collection and Analysis

Data collection options vary considerably in cost. The state agency and its evaluator should consider the following when thinking about data-related costs:

- **Are the data already being collected, or are new collection efforts required?** In general, using data that already exist will be less costly than collecting new data. For example, accessing existing administrative data costs less than collecting new data through a survey, which requires time and effort to develop and field a survey instrument.
- **Are internal staff available to extract data that can be shared with the evaluator?** The evaluator will likely need to work with IT, performance management, or other agency staff in charge of managing RESEA-related data. Internal staff might need an hour or two to discuss data needs with the evaluator, one or two days to write programs to extract the needed data, several hours to extract and check each requested data extract, and two or three hours over the course of the evaluation to answer evaluator questions about the data. Data will need to be shared securely, likely over a secure file transfer portal (FTP). Staff may need an hour or two to give the evaluator access to the agency's secure FTP and troubleshoot any questions.
- **How long will it take to analyze the data collected?** The amount of time the evaluator needs in order to prepare the data for analysis and then conduct the analysis will contribute to the cost. For example, if the evaluator will collect qualitative responses to interview or survey questions (e.g., open-ended responses), consider what strategies will be used to analyze the data. Depending on the research question, qualitative data may require substantial amounts of time for interpretation and analysis.

Data Security

All evaluations will need to establish data security protocols in line with evaluation industry standards and that meet human subjects protection requirements established by the Federal Policy for the Protection of Human Subjects ("Common Rule") and discussed in Chapter 8, Section 8.4. The state agency and the RESEA program administrator should consider whether an evaluator is using encryption software to restrict data access to authorized users only. Also consider the steps an evaluator is taking to securely store data. If any data have been collected in physical copies, such as paper surveys or flash drives, the evaluator should keep those securely stored, perhaps in a locked drawer or filing cabinet.

To the extent the state agency and its evaluator need to transfer data, consider options to securely make those transfers. In general, sending data, including PII, over email is not secure. Use encrypted file transfer protocols. Speak with the evaluator early in evaluation planning to ensure that the evaluator can securely store, transfer, and use claimant data.

6.4 Best Practices and Common Challenges

Data collection can seem complicated. There are many data sources to choose from and numerous factors to consider when deciding on a data collection plan. Adhering to some good practices and being aware of common challenges can simplify the process. This section discusses some of those best practices and challenges.

Best Practices

Chapter 2, Section 2.2 covered approaches a state agency can take to build capacity related to understanding and using data within the RESEA program team. This section revisits some of those approaches and discusses other best practices when using data for evaluation.

Managing a State Data System

Much of the data required for an evaluation will come from the state agency's own data information systems. For example, RESEA services are often recorded in state workforce data systems by RESEA program staff as services are offered or delivered. To produce top-quality data for evaluation, the state agency responsible for the evaluation should follow these best practices:

- **Revise or develop data entry protocols for staff to follow when entering data.** These protocols should specify data entry rules with sufficient detail to support consistent data entry practices across staff and outline any consequences for improper data entry. Protocols should be distributed to all staff.
- **Create a data dictionary.** This resource could be created as part of a data entry protocol and would include a definition for every data element in the state agency's data systems. This includes any variables that have coded values based on qualitative information. State agency staff tasked with managing the data systems may already have a data dictionary that can be shared with the RESEA program staff.
- **Train staff on data entry protocols.** Training should demonstrate standard data entry practices using the protocols and answer any staff questions about the developed protocols.
- **Routinely monitor data quality.** After training has been provided, state agencies should routinely monitor data entry, checking for errors, inconsistency, missing data, and other issues. State agencies can do this using site-level data quality reports. When it becomes clear that some sites or staff are not entering data in accordance with data entry protocols, agency staff should intervene to provide additional training to mitigate protocol non-compliance and reinforce expectations and standards.

Managing the Evaluator

The evaluator will manage much of the data collection work, but the state agency can take some simple steps to facilitate that process.

- **List required data elements.** Encourage the evaluator to provide a list of required data elements early.
- **Connect the evaluator with IT and data staff.** Put the evaluator directly in touch with IT or data management staff. This will allow state agency staff who manage information data systems to begin developing a plan for providing those data to the evaluator. Over the course of several conversations, the evaluator can directly communicate any data needs, and data managers can communicate any data limitations or set expectations on their capacity to share data.

- **Develop routines for extracting and transferring data.** Standardize a process to produce the data extract files the evaluator will need routinely over the lifespan of the evaluation.
- **Test the process of creating data extracts.** Extract test data files from data systems. State agencies and their evaluators may need to exchange several test files before settling on a final data extraction and transfer process. Completion of these early data extraction steps is necessary to begin an evaluation.

Challenges

Every evaluation is different, and state agencies may run into different challenges with the implementation of their RESEA evaluation. A few common data-related challenges reported from previous evaluations are listed below:

- **Linking individual records across data sets.** As discussed above, evaluation data will likely need to be collected at the individual level. That is, data files need to contain a separate record for each claimant included in the evaluation. At the same time, data for each claimant will likely need to be collected from multiple data sources, such as separate administrative data systems. Evaluators need to be able to match an individual claimant's data record from one data file to that same claimant's data record from a different file (e.g., linking RESEA intervention activity with UI claim record). To do this, data systems need a common unique identifier for each claimant across data files.

  Some state agencies, however, will find that a single identifier does not exist across all of the data systems needed for their evaluation. In that case:

  - The most reliable individual-level identifier is the SSN because it is unique to each person and does not change over time.
  - If the SSN is not collected by one or more systems, a customer identification number may be acceptable if (i) it is unique to each claimant; (ii) it does not change over time (e.g., a claimant who filed a claim in October 2016 would have the same identification number if the claimant also filed claim filed in October 2017); and (iii) it is used across all data systems of interest or can be connected to an identifier used in other of data systems.
  - If no unique numerical identifier is available, the state agency and its evaluator can use a combination of claimant name and date of birth. This alternative is less than ideal and should only be used when no other alternatives present themselves. Using this combination as an identifier introduces a small risk information on the wrong person is included in the study, given that neither name nor date of birth is unique (e.g., it is not hard to imagine two John Smiths born on the same day).
- **Processing UI data.** UI records are often recorded at the event level (e.g., at the weekly UI benefit payment level) and may be updated from week to week as state UI agency staff rectify claimants' under- or overpayments. Choosing an evaluator with experience processing this data is important for correctly summing UI benefit and duration information, as well as for correctly interpreting this information.
- **Aligning claim, service, and nonmonetary decision dates.** It is important to align claim, service, and nonmonetary decision dates in order to accurately reflect how long claimants receive UI, what services they engaged in, and for how long. Because different state agency staff members enter each type of data (UI staff, employment services staff, and automatic entry from computer systems), it can be challenging to align these dates in a way that

reflects program operations, especially if data entry protocols are not standardized. For example, suppose employment services staff do not enter data on the day the service is delivered, but they enter services information after a nonmonetary decision has been made. Those entries may make it appear as though the claimant received services after receipt of UI benefits ended. Such entries would misrepresent what actually happened in the program, raise questions from the evaluator, and take significant resources to investigate and fix. Promoting strong data entry protocols at the data entry stage limits the number of costly data questions that will occur at the analysis stage.

- **Handling unanticipated levels of missing data.** Many data fields will have some percentage of missing data. When a particular data field is missing a large percentage of data, it may not be able to be used in analysis. For example, if the state agency is interested in learning about how intervention impacts vary by age but values for the age variable are missing for 50 percent of the sample, the analysis will be constrained to only those cases where data are available. If data are consistently missing for too many cases, the analysis may not be generalizable to the entire claimant population.

  It is also important to emphasize that missing data problems must be addressed "upstream," at the point when claimants apply for UI and/or enroll in the RESEA intervention. Robust data entry training and quality control procedures at the outset help to ensure that any missingness reflects cases when there were no applicable data to record (e.g., UI earnings data are missing because no earnings were received) and not because of a data collection error.

- **Addressing unanticipated use of open text fields.** In the interest of time, staff may enter a wide range of information into open text fields rather than entering data in the specific fields the study means to capture. For example, staff might type "middle-aged white man" instead of choosing the options "male" in the "Gender" field, "age 50-64" in the "Age" field, and "White" and "Non-Hispanic" in the "Race" and "Ethnicity" fields, respectively. Open text fields allow for wide variety in responses. Standard statistical analysis software (e.g., SAS, STATA, and SPSS) is not designed to process such qualitative data efficiently and cannot easily generate frequencies using them. In contrast, data fields coded as radio buttons or drop-down menus are much more efficiently processed by these software. Processing qualitative data for numerous study participants is very expensive and time-consuming. If a large percentage of state agency staff continue to enter important data in open text fields, the evaluator may find it cost-prohibitive to use that data in the analysis.

- **Conveying information about data to the evaluator.** Accurately translating program-related activity into administrative data is complicated, and a full discussion of that process is beyond the scope of this chapter. When developing an evaluation plan, it is important to talk to experts in the state agency to understand exactly how data are recorded. Some key questions to ask include how data records are created and whether or not data are ever overwritten. The evaluator should be provided with as much written documentation as possible. For example, the state agency may wish to provide a data dictionary or record layout to its evaluator to the extent that those materials exist or can be easily created.

- **Defining an appropriate follow-up period.** As discussed earlier, defining the follow-up period is an essential part of ensuring that outcomes are being measured as needed to answer the evaluation's research questions. The evaluator will help determine an appropriate follow-up period for the study. There are two things to keep in mind when defining the follow-up period:

- The period begins for each individual claimant at the point of the claimant's entry into the program. Thus, Claimant A, who enters the RESEA program on January 1, 2017, has a follow-up period that begins on January 2; and Claimant B, who enters the RESEA program on May 15, 2017, has a follow-up period that begins on May 16.
- The timeframe for which follow-up data are collected for each individual claimant should be the same specified length of time, measured from the claimant's point of selection for RESEA. Thus, continuing with the example above, a state agency that elects to collect follow-up data at 10 months would collect data from Claimant A on November 1, 2017 and from Claimant B on March 15, 2018.

7. Selecting an Evaluator

Selecting a qualified, independent evaluator is one of the most important decisions a state agency will make when planning its RESEA evaluation. A qualified evaluator must be able to demonstrate experience in labor evaluations, expertise in conducting the type of evaluation the state agency is pursuing, and independence. The evaluator is separated from the RESEA program intervention in such a way as to be able to objectively and transparently report evaluation findings regardless of whether those findings are favorable, unfavorable, or null. (See section 7.1 for more details.) This chapter provides insight on how to select a qualified, independent evaluator, regardless of whether a state agency plans to select an evaluator from internal state staff or an external organization such as a partner university or a research firm. A highly qualified, independent evaluator can help state agencies produce a strong evaluation that generates meaningful results, meets CLEAR standards, and minimizes risks to the RESEA program and the claimants it serves.

As discussed in this chapter, when engaging in the process of selecting an evaluator, state agencies should aim to:

- ✓ Define the qualities, skills, and competencies needed in a potential evaluator.
- ✓ Determine the role state agency staff will play in an evaluation.
- ✓ Understand key steps and considerations for soliciting, funding, selecting, contracting, and working with an external evaluator.

7.1 Defining Evaluator Qualifications and Skills

Whether a state agency is considering an internal or external evaluator, there are key qualities, skills, and competencies that any evaluator will need in order to successfully execute a rigorous, high-quality evaluation.

Evaluator Independence

Evaluator independence is a core principle in the research field. Both the Social Security Act and DOL's Evaluation Policy require that the evaluator be independent and specifically that "evaluation functions [be insulated] from undue influence."[34] In practice, independence means that evaluators will lead the study and objectively and transparently report on the evaluation and its findings. The evaluator should not be subjected to any inappropriate pressure (perceived or real) from the state agency or other RESEA stakeholders to alter reporting on evaluation findings. For instance, state agencies cannot ask the evaluator to make edits to a report in order to make evaluation results seem more positive. Independence is most readily achieved from outside the state agency, although it may be feasible internally. If a state agency selects qualified state agency staff to serve as its evaluator, the state agency will want to firmly establish that the in-house evaluator is separate from the RESEA program and not subject to undue influence that can bias the way that the evaluator conducts and reports the evaluation.

Evaluator independence does not mean that state agency staff are not involved in the evaluation. State agency staff will want to be in regular communication with their evaluator to help make sure the evaluation stays on track, to facilitate access to administrative data, to verify proper protection of study participants, and to check that reports and other written deliverables meet professional

[34] United States Department of Labor. (November 2013). *U.S. Department of Labor Evaluation Policy.* https://www.dol.gov/asp/evaluation/EvaluationPolicy.htm

and agency standards (addressed in Chapter 8). Again, though state agency staff can and should provide constructive comments and feedback on reports, the evaluator is the author of the evaluation's reports and has final say on findings and language. Meaning, among other things, that state agency staff cannot seek changes that would skew the reporting (e.g., in order to make the results seem more positive).

What Skills Are Needed to Conduct an Evaluation?

The skills an evaluator will likely need will vary depending on the evaluation design type the state agency wants to use. For example, an impact study is typically technically complex and involves collecting quantitative data, thus requires an evaluator with skills in statistical analysis. In contrast, an implementation study may not involve any statistical analysis and relies more on qualitative data (e.g., interviews and observations). An implementation study, thus, requires that an evaluator has strong communication and interpersonal skills, qualitative analysis expertise as well as research methods. Also, an implementation study will need an evaluator with experience collecting and analyzing qualitative data. Exhibit 7-1 details the skills needed to conduct each of the evaluation design types.

Exhibit 7-1. Evaluator Skills Needed, by Evaluation Design Type

Implementation Study	Outcomes Study	Impact Study
• Designing qualitative research projects. • Developing interview protocols and observation guides. • Understanding of relevant external evaluations, research methods, and state programs • Conducting interviews with a broad array of stakeholders, including senior-level state agency staff. • Analyzing qualitative data using a structured coding guide or qualitative data analysis software (e.g., NVivo). • Synthesizing information collected to identify best practices, opportunities for program redesign and process improvement, relevant policy recommendations, and implications for future research. • Summarizing research findings in a formal report.	• Designing quantitative research projects. • Understanding of relevant external evaluations, research methods, state (and relevant external) data systems, and performance benchmarks. • Developing data collection protocols, e.g., surveys. • Conducting statistical analysis of large data sets using quantitative data analysis software (e.g., STATA, SAS, and SPSS). • Synthesizing information collected to identify best practices, opportunities for program redesign and process improvement, relevant policy recommendations, and implications for future research • Summarizing research findings in a formal report.	• Designing and leading a labor-focused impact study of the proposed design type (e.g., random assignment, matching, and regression discontinuity). • Training in statistical analysis methods relevant for the study. • Creating a statistically equivalent comparison or control group. • Implementing random assignment procedures (for RCT designs), including: – Building, testing, and incorporating a random assignment algorithm into existing state data systems. – Monitoring random assignment to ensure that assignment is truly random across a number of factors and that claimants are receiving services in line with their random assignment status. – Training frontline staff in evaluation procedures and protocols. • Creating complex analytic data sets using information from multiple data sources. • Synthesizing information collected to identify best practices, opportunities for program redesign and process improvement, relevant policy recommendations, and implications for future research • Summarizing impact findings for multiple audiences (technical and non-technical). • Summarizing research findings in a formal report.

7.2 Establishing the Role of State Workforce Agency Staff in Evaluation

The state agency itself may have the capability in-house to conduct the planned evaluation. It is possible that the expertise needed to conduct the evaluation may reside within the state's Labor Market Information unit, elsewhere in a research division of state agency, or within another designated state agency. Remember, though, that the evaluator must be independent of the RESEA program.

When considering whether or not to rely wholly on in-house staff, the following should be considered:

- Are there state agency staff (insulated from the RESEA program) with the requisite skills and experience to lead the evaluation and function as the Project Director and/or Principal Investigator? Does that staff person have time and availability to oversee the evaluation, given other duties?
- Can the state agency field an evaluation team that is cohesive and can be committed to working on the study?
- Would it likely be more or less expensive to conduct the study in-house?

If in-house staff members have the required independence, expertise, and experience, the state agency can move forward with an internal evaluator. If in-house staff members do not have the right mix of qualifications, or if their independence cannot be assured, it may make sense to obtain external evaluation services. This is particularly true if a state agency is planning to conduct an evaluation with higher technical requirements, such as an impact evaluation.

Regardless of whether a state agency uses an in-house or external evaluator, state agency staff will play an important role in the evaluation. State agency staff members are a critical source of information about the RESEA intervention during many stages of the evaluation. Exhibit 7-2 provides the delineation of responsibilities between the state agency and its evaluator for each potential evaluation activity. It is important to note that the nature of the evaluation activity may vary depending on the evaluation design type.

Exhibit 7-2. Evaluation Activities and Division of Responsibility

Activity	State Agency	Evaluator
Determining what to test, the basic research questions, main outcomes of interest	✓	
Identifying evaluation design type	✓	
Developing a timeline that considers the study follow-up period	✓	
Selecting an evaluator	✓	
Completing an evaluability assessment	✓	✓
Refining research questions, outcomes measures, data sources, required sample sizes	✓	✓
Determining evaluation procedures		✓
Establishing the evaluation timeline	✓	✓
Producing an Evaluation Design Report		✓
Aligning data forms across partners (if needed)	✓	✓
Formalizing data use agreement, data transfer process	✓	✓
Modifying existing data systems (if needed)	✓	
Establishing random assignment process (if needed)		✓
Making changes to program operations (if needed)	✓	✓
Training staff		✓
Refining research questions of interest	✓	✓
Developing data collection protocols		✓
Enrolling claimants into the study	✓	✓
Collecting evaluation data (ongoing)		✓

Activity	State Agency	Evaluator
Monitoring evaluation activities	✓	✓
Preparing data for analysis, including data cleaning, robustness checks, initial descriptive analysis		✓
Conducting analyses according to plan		✓
Drafting and revising reports		✓
Reviewing drafts of reports	✓	

7.3 Selecting an External Evaluator

After having considered the expertise, independence, and hours of effort that are required to carry out the desired evaluation—and compared those requirements againast the resources available within the agency—a state agency may find that it is preferable to contract with an external evaluator to conduct the study. Some state agencies have limited research departments (e.g., low levels of funding) and do not the appropriate expertise to conduct an evaluation in-house.[35] Given these internal research resource constraints, state agencies may find it necessary to obtain an external evaluator.

Third-Party Evaluator Types and Sources

There are three general categories of third-party evaluators:

- Research/evaluation firms;
- Independent research consultants; and
- University centers, faculty, or affiliated academics.

State agencies can find listings for evaluation firms and consultants on the websites of professional associations, such as the American Evaluation Association (eval.org) and Association for Public Policy Analysis and Management (appam.org).

Information about university-affiliated research centers, faculty, or academics can be found on university websites.

Also, state agencies can use CLEAR to identify potentially qualified evaluators by searching for names of authors or institutions that conducted studies of similar interventions.

Once engaged, the evaluator will manage the evaluation to its completion. As such, the evaluator must be able to overcome expected and unexpected challenges, maintain timelines that are sometimes aggressive, conduct complex data analyses, and develop high-quality evaluation reports. In addition, for state agencies interested in conducting an impact study, it will be critical to consider whether a potential evaluator has the qualifications, experience, and dedicated time to manage the complexities of conducting impact studies (RCTs, QEDs). Impact studies generally require a significant degree of technical expertise as well as independence to ensure that the evaluation is objective and the results can be trusted as impartial.

This section discusses considerations for finding and selecting a third-party evaluator, including:

- Creating a statement of work or performance work statement.
- Funding the evaluation.
- Creating an evaluation timeline.
- Reviewing evaluator qualifications and experience.

[35] Chocolaad, Y. & Wandner, S. (2017). *Evidence-building capacity in state workforce agencies: Insights from a national scan and two state site visits.* National Association of State Workforce Agencies (NASWA). https://wdr.doleta.gov/research/FullText_Documents/ETAOP-2017-13_Evidence_Building_Capacity_in_State_Workforce_Agencies_Report.pdf

Additionally, this section discusses two approaches to obtaining a third-party evaluator:

- Soliciting bids for an evaluator.
- Partnering with universities.

Creating a Statement of Work (SOW) or Request for Proposal (RFP)

State agencies interested in obtaining external evaluation help will need to create a "statement of work" document, also referred to as the performance work statement or scope of work, which is typically included in an RFP if that is required for procuring an external evaluator. While different terms may be used, the goal is the same: to draft clear, written guidelines for the evaluation tasks that the evaluator will perform. This section of the toolkit uses the term statement of work.

The SOW provides a high-level description of the state agency's program evaluation requirements, including research questions, evaluation tasks, and any reporting the state agency would like the evaluator to carry out. Building on the state agency's preliminary design parameters (e.g., possible interventions, research questions, and evaluation designs), the SOW will describe key aspects of the state agency's proposed evaluation, including:

- **Intervention** being tested, including relevant policy background and a logic model (if available).
- Evaluation's **purpose, parameters** (e.g., timeline, whole program versus component, geographic locations)**, and key research questions** to meet the state agency's evaluation need.
- Preferred **evaluation type**(s) (such as impact study, outcomes study, or implementation study).
- High-level **requirements related to the research design** (such as anticipated data sources, minimum data collection requirements, and minimum reporting requirements) recognizing that potential evaluators will use their expertise and creativity to develop a technical proposal with specific methodological details to meet the state agency's evaluation requirements.
- Requirements for **protecting participants' rights** (such as Institutional Review Board review; see Chapter 8, Section 8.4 for more information).

The SOW should also address the following important practical considerations:

- **Evaluation budget.** While evaluators will develop a detailed budget, it is important for the state agency to develop a realistic cost estimate that will help it to understand how to budget for the evaluation year-over-year until the completion of the evaluation. Having a cost estimate will also help make sure that the state agency provides the evaluator with information about the level of resources available for the evaluation. If a state agency is funding the evaluation incrementally, such as through options on a base contract or through successive contracts, understanding the total anticipated budget to complete the evaluation will help the state agency develop a procurement strategy that can meet those needs. (See the section on Funding the Evaluation below for more details.) Various methods exist to developing cost estimates. State agencies may wish to examine historical rates and information from previous or similar projects, use an established calculator for similar services, or consider other methods in consultation with their state procurement office. Exhibit 7-3 describes items that should be considered in developing a budget.

Exhibit 7-3. Factors Affecting an Evaluation Budget

Factor	Considerations
Evaluator qualifications	More experienced evaluators are more likely to have the knowledge and expertise required to conduct a high-quality evaluation and as a result can be more expensive than less experienced evaluators. Because of the trade-off between cost and experience, many evaluators will propose an evaluation team that uses a mix of senior, mid-level, and junior staff to meet quality needs while maintaining cost-competitiveness.
Sample size	Sample size refers to the number of participants included in the evaluation. A larger sample size can yield more precise findings. However, as the sample size increases, some of the evaluation costs may also increase, such as costs related to participant recruitment, surveying, data collection and analysis, and staff time needed for evaluation procedures. While the evaluator will help assess the study's specific sample size needs, having a general idea of whether the sample size will need to be large or small can help state agencies assess the evaluation budget.
RESEA program data systems	All evaluations require some kind of data collection to answer research questions. The RESEA program needs to have systems in place to record program data. The evaluator needs to have systems in place to collect data from the program and other sources. In some cases, existing data systems can be used for the evaluation. In other cases, data systems will need to be modified to accommodate the evaluation (e.g., additional variables need to be captured in the management information system; the system needs to be modified to conduct random assignment). It is likely that the evaluator may need to work across several data systems to capture the information needed. If anticipated modifications to existing data systems are needed, state agencies should budget funds to help them do so.
Data collection: Administrative data	Administrative data from sources beyond the program are often critical for the evaluation analysis (e.g., federal, state, and local workforce agencies, including UI). Collecting data from these sources may have an associated cost that may include data transfer and storage, time spent orienting evaluators to data fields, data cleaning, addressing data issues, analysis, and building study-specific reports.
Data collection: Surveys	If the research design involves conducting surveys (e.g., surveys of staff or claimants) to gather information that cannot be found in administrative data, this work will need to be considered and planned for in the budget. Costs for conducting surveys and follow-up surveys include staff time to develop and revise the survey questions, time to track down potential respondents, potentially financial incentives for survey completion (if used), postage (if the postal service will be used to deliver the survey), technology to deliver and support the instrument, and staff time for synthesis of collected information. Survey costs will increase with sample size and should be carefully considered, as they tend to make up a significant portion of an evaluation budget.
Travel	Evaluation teams typically visit program sites (one to several times) to better understand how programs work on the ground and to monitor evaluation procedures. If the evaluator is not local, the evaluation budget should account for at least some transportation, lodging, meals, and incidentals for one to two evaluators over the lifecycle of the evaluation.
Project Management	The evaluator will need to stay in regular contact with the state agency staff who are overseeing the evaluation throughout the period of performance of the study. Planning for regular check-in meetings to discuss how the RESEA intervention's implementation is unfolding and how critical evaluation activities are progressing is important to ensuring a successful project experience. Because much of program development and evaluation design is iterative, the evaluation budget will need to account for routine check-in meetings, progress reports, and several rounds of revisions to major deliverables.

- **Timelines.** Setting an appropriate timeline will be critical to ensuring the feasibility of the planned evaluation project. (See the Creating an Evaluation Timeline subsection later in this chapter for more information about estimating a timeline). In general, state agencies should anticipate that the evaluation, particularly one using an impact study design, will require multiple years. State agencies may find it helpful to review sample timelines from prior research and evaluation projects to help them map a realistic tentative schedule. When reviewing the evaluation timeline from a potential evaluator, it is important for state agencies to consider whether the evaluator's proposed timeline is reasonable and feasible for the evaluation design. Researchers with little or no experience often underestimate the amount of time needed for the evaluation's various phases.

- **Protection of participant rights.** The primary purpose of an Institutional Review Board (IRB) is to protect the welfare of human subjects used in research. Remember to factor appropriate time into the schedule for review, approval, and continuing review, to ensure sufficient protections are in place. It is possible that the IRB will determine that the evaluation does not pose any risk to human subjects and is exempt from a full review. However, the exemption determination must be made by the IRB, not by the state agency, the evaluator, or DOL. IRB reviews can take considerable time and may have implications to both the evaluation schedule and budget. Chapter 8, Section 8.4 provides information on the IRB process that can help state agencies plan adjustments accordingly.

SOWs should reflect the type of evaluation to be conducted. Exhibit 7-4 outlines the different types of evaluations (discussed in greater depth in Chapter 4 and Chapter 5) and the different design elements to consider and include in an SOW.

Exhibit 7-4. Key Design Elements for Evaluations, by Type

Key Elements	Impact Study	Outcomes Study	Implementation Study
General purpose	• Estimates the difference in individual outcomes attributable to a specific intervention.	• Compares individual outcomes against goals, across programs or locations, or over time.	• Documents program operation or compares it against goals, across locations, or over time.
Study purpose	• Specific purpose statement(s) aligned with overall purpose of the evaluation design type.		
Study parameters	• Sites or areas. • Target population. • Outcomes addressed. • Observation period.	• Sites or areas. • Target population. • Outcomes addressed. • Observation period.	• Sites or areas. • Elements addressed. • Timing and length of study.
Types of key research questions	• Questions to assess the extent to which the program effects participants' outcomes.	• Questions to assess whether the program is meeting its objectives.	• Questions to assess program delivery process such as: – How is it operating? – Is it operating as planned? – Level of participation?
Analysis approach	• Research design: – Experimental (RCT). – Quasi-experimental design (QED). – Rapid cycle. • Theory-based. • Sampling methods and sampling frame. • Statistical methods (tests by which impacts are determined to be statistically significant).	• Research design: – Longitudinal. – Cross-sectional. • Sampling methods and sampling frame. • Statistical methods.	• Research design: – Case study. – Process study: Fidelity analysis, Participant flow analysis. – Performance study. – Site comparison study. • Qualitative research methods such as interviews, focus groups, document review, and observations.
Data collection/ Data sources	• Surveys. • Administrative records. • Other data collection.	• Surveys. • Administrative records. • Other data collection.	• Field data collection such as interviews, focus groups, and observations. • Document review such as program records and internal reports on participation and completion. • Other data collection.

Funding the Evaluation

The SSA allows state agencies to use up to 10 percent of their RESEA funds on evaluation activities.[36] State agencies may wish to supplement RESEA evaluation funds with state resources or other sources of evaluation funds, provided that doing so does not present a conflict of interest. Because RESEA is funded annually and most evaluations will take multiple years to implement and complete, state agencies may wish to work with their procurement office to carefully plan their approach to funding the evaluation over multiple years. Regardless of who conducts the evaluation, state agencies should ensure that any contracts for evaluation are awarded in accordance with applicable procurement rules.

This section discusses two ways RESEA evaluations conducted by an external evaluator might be funded. The options available to state agencies will depend on the procurement rules to which they are subject.

Fund the evaluation in parts as new RESEA funds become available each year. Having decided on a set of research questions and preferred evaluation design, state agencies can choose to divide the evaluation into its many tasks and fund each one as resources become available. For example, state agencies using this incremental approach could:

- Use Year 1 funds to support evaluator selection, creation of the Evaluation Design Report and evaluation procedures, and staff training in evaluation procedures and protocols.
- Use Year 2 funds to start the study (i.e., begin enrolling claimants into the study), including random assignment (if applicable), data collection, and study monitoring.
- Use Year 3 funds to support ongoing evaluation activities, which may include continuing random assignment, data collection, and study monitoring.
- Use Year 4 funds to support data analysis and reporting.

A benefit of this option is that state agencies will be able to ensure that they have adequate funds to support the tasks that are contracted for each year. It also provides more flexibility to adjust plans over time. Depending on the available funds for the year, state agencies may be able to reduce or increase the number of evaluation related tasks that are contracted. State agencies will be able to determine annually how to best maximize their evaluation funds.

A potential challenge of this approach is that, depending on local or state procurement rules, state agencies may be required to re-compete their RESEA evaluation contract each year. State agencies that are allowed to have follow-on or sole source types of contracts (discussed below) will not need to re-compete their RESEA evaluation contract. If state agencies do need to re-compete their RESEA evaluation contract, the time and cost associated with procurement may be significant. There is also a distinct possibility that the original evaluator will not win or rebid on subsequent contracts. State agencies can specify in their evaluation contract that the state agency owns all of the information and products produced by the evaluation, but there are significant drawbacks associated with changing your evaluator. Institutional knowledge about the evaluation and the intervention being tested will likely be lost, even if each evaluator diligently documents its decisions and work activities. There will be a burden on state agency staff overseeing the evaluation to ensure all materials are securely transferred to the new evaluator. Finally, a new evaluator will need time to become knowledgeable about an evaluation that is already in progress and may wish to change aspects of the evaluation that cannot be changed. For example, the original evaluator will make many important decisions related to baseline data collection. Once study enrollment begins,

[36] Social Security Act, 42 U.S.C. § 306 (2018)

baseline data collection forms cannot be changed without excluding the initial sample from the analysis.

Exhibit 7-5. Pros and Cons of Funding the Evaluation Incrementally

Pros	Cons
• Funds evaluation as financial support becomes available.	• May result in multiple evaluators, if a different evaluator is selected for each stage. • Creates additional administrative work with each procurement cycle. • May delay evaluation if procurement process has delays.

Fund initial evaluation activities, but in the evaluation contract include "optional" evaluation tasks that can be exercised at later dates if additional funds become available. Under this scenario, state agencies would fully fund the evaluation activities that can be supported by the first year of funding (for many small and medium state agencies, this would mean funding the Evaluation Design Report and creation of evaluation procedures). However, state agencies would have bidders also provide details and draft budgets for the remaining scope of work, including evaluation launch, data collection, data analysis, reporting, and dissemination. These tasks would be included in the contract as options that the state agency could choose to fund when future funds become available. Including optional tasks allows state agencies to select an evaluator based on the evaluator's approach to the entire evaluation and to maintain evaluator consistency without committing resources to which the state agency does not yet have access.

A state agency looking to fund its evaluation using the optional task approach will need to determine how much of the evaluation can be accomplished with the evaluation funds it is using from its RESEA grant in the first year. The tasks that can be completed within the first year will depend in part on which evaluation type(s) the state agency plans to pursue and the number of claimants the state agency anticipates serving. State agencies with limited funding may want to include the initial evaluation design, procedures, assessment of data quality, and evaluation instruments in their fully-funded SOW and leave evaluation implementation, data collection and analysis, and reporting as optional tasks. State agencies will want to make sure that their SOW specifies which activities are covered under the fully-funded contract and which activities are optional tasks. The RFP, partnering agreement, or SOW should ask potential evaluators to create a full technical proposal and budget for the fully-funded contract and a short narrative and projected budget for each optional task. Structuring the contract this way will help both the state agency and the potential evaluator consider the evaluation budget and timeline and determine what can and cannot be achieved.

Exhibit 7-6. Pros and Cons of Funding the Evaluation with Optional Tasks

Pros	Cons
• Funds core evaluation in one procurement cycle. • Allows additional tasks to be added by exercising options if financial support exists, with somewhat less burdensome procurement and administrative actions required. • Funds one evaluator for the entire evaluation.	• May make potential evaluators reluctant to submit a bid. • Difficult for evaluators to schedule and accurately plan out their work, which sometimes raises total cost.

Regardless of which option is chosen, state agencies will want to consult with their local procurement team, and their DOL Federal Project Officer, if appropriate, to answer specific questions about procurement in their state and using RESEA funds. State agencies should always follow state and federal procurement rules if obtaining third-party evaluation services.

Creating an Evaluation Timeline

Creating a draft timeline for critical evaluation activities, such as implementation of the basic study design, data collection procedures, and steps for report development and dissemination, can help ensure those evaluation activities are aligned with intervention activities. It can also ensure that the state agency, its evaluator, and other stakeholders are on the same page about the timing of the evaluation. State agencies will want to set a time frame during which they give the selected evaluator opportunities to draft and refine the evaluation design; develop research protocols and data analysis plans; and identify additional evaluation requirements, tasks, and deliverables provided in a detailed project timeline.

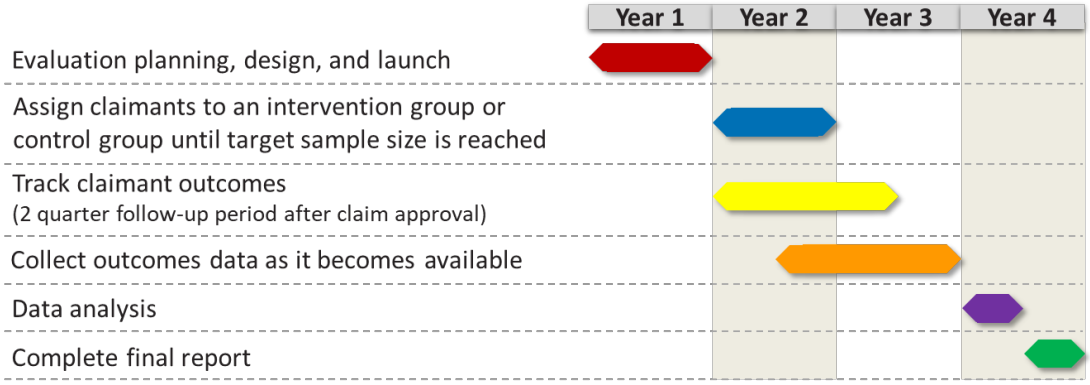
Every evaluation is unique, but Exhibit 7-7 provides examples of common tasks associated with each stage of the evaluation, the entity responsible for the task, and an estimated time frame in months. Tasks in the table appear in the order they would typically occur in evaluation implementation, although many tasks may be performed simultaneously. Some of the tasks listed may or may not be appropriate for all evaluations. Additionally, state agencies or their selected evaluator may identify other evaluation tasks.

Exhibit 7-7. Sample Evaluation Activities and Timing

Task	Who Is Responsible	Time It Takes
Planning the Evaluation		
Define the evaluation	State RESEA agency staff	<1–6 months
Select an evaluator	State RESEA agency staff	1–12 months
Conduct an evaluability assessment	Evaluator	2–3 months
Designing Your Evaluation		
Refine evaluation plans with the assistance of the evaluator	Evaluator, state RESEA agency staff, and state data staff	2–6 months
Submit application to IRB	Evaluator, in consultation with state RESEA agency staff and DOL	1–2 months
Write an Evaluation Design Report	Evaluator and state RESEA agency staff	2–4 months
Implementing the Evaluation		
Determine data requirements and modify systems (as needed)	Evaluator, state RESEA agency staff, along with state data and IT staff	2–4 months
Train agency staff	Evaluator	2–4 weeks
Build sufficient sample	Evaluator and state RESEA agency staff	Varies depending on evaluation design and sample size needed
Collect data	Evaluator and state RESEA agency staff	Varies depending on evaluation design and sample size needed
Monitor data and evaluation procedures (*prospective* studies only)	Evaluator and state RESEA agency staff	Ongoing, throughout entire data collection phase
Analyzing the Data		
Prepare data and conduct data analysis	Evaluator	2–3 months
Communicating Findings		
Write and publish interim report(s)	Evaluator	2–4 months
Write and publish Final Report	Evaluator	2–4 months
Prepare public use data set (if possible)	Evaluator	2–6 weeks

The time frames provided in Exhibit 7-7 are broad estimates for common evaluation tasks that state agencies should consider in building their high-level evaluation plan timeline. The actual time frame needed to complete the task will depend on factors such as the scale of the evaluation, the type of data to be collected, and whether or not IRB approval is needed. When developing the estimated timeline, it is important to keep in mind that evaluations are multi-year efforts. Additionally, evaluation tasks may happen concurrently. Exhibit 7-8 provides an illustrative example of a multi-year impact evaluation.

Exhibit 7-8. Illustrative Example of Impact Study Timeline



Finally, it is possible that the evaluator will update the state agency's preliminary timeline based on the evaluator's expertise and experience implementing similar evaluations. However, providing the evaluator with a sample timeline helps the evaluator to better understand how to best plan out the evaluation. After procurement, the state agency should work with its selected evaluator to finalize the evaluation's timeline.

Reviewing Evaluator Experience and Qualifications

When selecting an external evaluator, state agencies should consider the levels of and types of experience and qualifications of the candidates. In order to produce a high-quality evaluation, the evaluator will need to have experience conducting a workforce evaluation using the evaluation design type(s) that are of interest to the state agency, plus specialized education and training in evaluation. For example, when considering among evaluators who have submitted proposals for conducting an RCT, the evaluator the state agency chooses should have demonstrated expertise and experience conducting an RCT study, preferably within the labor force field and perhaps even within the UI field.

Again, it is also important that the evaluator is independent from the RESEA intervention. Section 7.1 discusses evaluator independence in detail.

In general, it is reasonable to expect a qualified evaluator to have:

- Experience implementing an evaluation of the proposed evaluation design type, preferably in the proposed content area (labor evaluations, and perhaps specifically program evaluations of reemployment services or UI programs).
- Senior staff with the appropriate technical evaluation and methods expertise, and a postgraduate education or an adequate amount of experience (for example, at least five years of experience) to demonstrate the technical skills necessary to implement the study.
- Capacity and resources to facilitate meetings, collect and analyze data, and write reports.
- Other specialized knowledge important to meet the state agency's requirements.

Overall, evaluator firm and staff qualifications are one of the most important factors in ensuring a high-quality evaluation. After deciding on the evaluator qualifications needed for their evaluation, state agencies should describe those qualifications in the RFP to ensure that only qualified applicants submit bids.

Exhibit 7-9 outlines some questions to ask about a potential evaluator's qualifications and experience and provides examples of evidence to examine.

Exhibit 7-9. Evaluator Selection Questions and Evidence of Experience

Topic	Questions	Examples of Evidence
Evaluator competencies and experience	• Does the potential evaluator demonstrate: – Competency through experience in successfully conducting the type of evaluation you want conducted? If so, how much experience? – Knowledge of and experience with the subject matter of your evaluation or a closely related topic (e.g., transitional job programs, employment services, other job training programs)? – Ability to recognize and overcome evaluation challenges?	• Summary of previous studies using your proposed evaluation design or in your content field. • Previous publications or other publicly available deliverables from these studies. • Summary of the potential evaluator's mission, history, and experience. • Record of previous clients with similar needs.
Staff qualifications	• Does the potential evaluator have: – A Project Director and a Principal Investigator (or similar lead roles) with appropriate education and experience? – A proposed evaluation team with a mix of seniority levels? (A mix of staff at senior, mid-, and junior-levels exhibits the evaluator's attempt to meet the staffing requirements for the evaluation cost-effectively.) – Other staff qualifications/skills needed to conduct the evaluation as you envision it (e.g., proficiency in data collection and analysis methods)?	• Resumes or CVs. • Publications written by senior staff demonstrating familiarity or skills in a particular evaluation type or context. • Staff members with postgraduate degrees in related subject areas. • Identification of staff by tasks described in the evaluation requirements.
Capacity and resources	• Does the potential evaluator have: – Sufficient capacity to carry out the tasks associated with the type of evaluation? For example, an evaluator conducting a multi-site RCT study would likely require more staff than an evaluator conducting an outcomes study at a single site. – Demonstrated capacity and resources to (1) collect data, (2) conduct quality checks of the data; (3) analyze data (e.g., statistical analysis tools), (4) interpret the results, (5) produce graphs and tables, (6) write reports, and (7) provide technical assistance on evaluation-related issues? – Administrative, IT, and publication support needed for the evaluation and subsequent deliverables?	• Details and examples of evaluations with similar tasks conducted by proposed evaluation team member(s). (This information might appear in staff resumes or a short biography included with the evaluator's technical proposal.) • Administrative and IT resources to meet the evaluation requirements (e.g., management and review processes, interview and site visit protocols, secure data collection and transfer sites). • Data collection software and/or analysis tools created and/or used, and examples from previous studies. • Sample products produced for previous clients.
Specialized knowledge	• Does the potential evaluator have specialized knowledge to meet the evaluation requirements (e.g., familiarity with relevant geographic, cultural, or other contextual elements)?	• Staff with credentials demonstrating appropriate skills, cultural competence, knowledge, and professional training to conduct the study according to the standards and principles of the evaluation profession.

Soliciting Bids for an Evaluator

Many state agencies decide to conduct their evaluations using the services of a third-party evaluator. The term third-party evaluators refers to evaluators that are external to the state and consist of private research firms or organizations, including university research centers. Third-party evaluators typically specialize in conducting evaluations and have a deep knowledge of all aspects of the evaluation process. State agencies use an RFP bidding process to advertise evaluation opportunities, receive and assess bids, and select awardees. In general, state agencies will need to work with their state procurement office to prepare the SOW (as discussed earlier in this chapter), articulate their required qualifications (corporate and staff), and provide a general budget estimate as the basis for the development of the actual RFP. Then, the state agency's two primary tasks are to publicize the RFP and assess the proposals.

Publicizing the Request for Proposals (RFP)

Once approved by the state, the RFP is usually posted in the public domain. Typically states have a process for formally announcing RFPs and requesting responses. However, to ensure the RFP reaches a wide range of potential evaluators, additional outreach may be useful to increase the number of bidders. All additional outreach should comply with state procurement process.

The following avenues may be appropriate to consider. Always be sure to include information on where to find the official state announcement and 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 state agency, including but not limited to LinkedIn, Twitter, and Facebook.
- Announce the RFP in local, state, and national evaluation newsletters or publications.

Assessing Bidder Proposals

Ideally, state agencies will receive several proposals in response to their RFP. The state procurement office likely has a process in place to review and assess proposals to select the winner. Most often, such a process uses weighted selection or evaluation criteria to identify the strongest technical proposal with the best staff qualifications. These rubrics allow state agencies to allocate more points to the qualifications or characteristics they value most, which will allow them to identify the candidates that likely will be a good fit for their needs. For example, an evaluator's institutional and staff experience are typical assessment factors, but the state agency may also value content knowledge or experience in their geographic area. The weight assigned to each factor should reflect the state agency's priorities.

When considering proposals, keep in mind the importance of making appropriate tradeoffs between quality and cost competitiveness. Bids may vary some in price, but all bids should be responsive to the key needs outlined in the RFP, and budgets should reflect the associated level of effort. The lowest-priced bid might not be offering the best technical methods for your evaluation. Remember, that if a potential bid seems too good to be true, it probably is. Do not hesitate to ask the evaluator more questions about the services it plans to offer, the professionals who will be providing these services, and their experience providing these services in the past.

Partnering with Universities

As noted in the 2017 NASWA report on state capacity building,[37] many state agencies 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 their ability to:

- Supplement and augment a state agency's research and evaluation staff with highly qualified evaluators, especially regarding QED and RCT evaluation approaches.
- Host and otherwise support the use of large data sets by providing secure computing facilities (social science research universities/centers).
- Obtain IRB approval for human subject's research, as many major universities host established IRBs registered with the U.S. Department of Health and Human Services.
- Bring other university, foundation, and research organization partners to the table. Major research universities often have existing social science networks and partnerships of their own that state agencies may be able to exploit.
- Benefit from local and ongoing sociodemographic and socioeconomic research that the university may conduct.

Such partnerships are options for conducting a specific, one-time study or for establishing 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 than contracting for a single evaluation, and such relationships afford many more options for external evaluations, as documented in the NASWA report. The key to forming partnerships is a mutually beneficial agreement or memorandum of understanding that documents each partner's role and responsibilities.

[37] Chocolaad, Y. & Wandner, S. (2017). *Evidence-building capacity in state workforce agencies: Insights from a national scan and two state site visits.* National Association of State Workforce Agencies (NASWA). https://wdr.doleta.gov/research/FullText_Documents/ETAOP-2017-13_Evidence_Building_Capacity_in_State_Workforce_Agencies_Report.pdf

8. Implementing the Evaluation

Whether the state agency selects an evaluator from within or outside the agency, the RESEA intervention staff should work with the evaluator to carry out the evaluation. Though the evaluator will be the primary party responsible for implementing the evaluation, the evaluator will need the state agency's help to implement numerous core evaluation activities. Having an understanding of these activities will allow the state agency to monitor the evaluator's work and assess whether the final product will meet the state agency's needs. This chapter focuses on these activities and discusses how the state agency is likely to work with the evaluator on the implementation of the evaluation.

The implementation of the evaluation will involve:

- ✓ Fostering collaborative relationships between state agency staff and the evaluator.
- ✓ Coordinating key evaluation implementation activities that require input from state agency staff.
- ✓ Creating a strong Evaluation Design Report.
- ✓ Protecting the rights of human subjects of research.
- ✓ Communicating the results of the evaluation, including ensuring evaluator independence and evaluation transparency.

8.1 Fostering Collaborative Relationships with an Evaluator

The evaluator will do much of the work to implement the evaluation, but the state agency's feedback, direction, and assistance will be needed to implement many parts of the evaluation plan. Viewing the relationship with the evaluator as an ongoing one that will be cultivated over time is vital to the success of the evaluation. The evaluator is responsible for responding to the state agency's needs and requests, and at the same time, the state agency should obtain and listen to the evaluator's sound recommendations regarding evaluation methodologies, data collection, and analysis, and maintain the key principles of transparency and independence of the evaluation.[38]

To strengthen the relationship between the state agency and the evaluator, the following steps can be useful:

- **Designate an RESEA state agency staff member to work closely with the evaluator.** It will be important to designate at least one RESEA staff member who can help facilitate conversations between the RESEA state agency and the evaluator. This staff member should serve as the primary point of contact for the evaluator. The staff member would be primarily responsible for staying up to date on evaluation tasks, monitoring evaluation activities, and reporting back to the RESEA program administrator. Ideally, this staff member would have some knowledge of evaluation concepts and also be able to coordinate with senior-level state agency staff about important evaluation decisions. As a liaison between the state agency and the evaluator, this staff member will be responsible for

[38] Evaluator independence—a critical principle in research—requires that evaluators lead the study and generate evaluation findings on their own, without undue influence. See Chapter 7 for more details.

helping the evaluator to understand and access RESEA program data, arrange access to state and local staff as needed, and answer questions about formal policy and its evolution.

- **Communicate regularly with the evaluator.** Conducting an evaluation requires careful coordination between the state agency and the evaluator. The state agency will need to ensure that the evaluator has access to RESEA program resources and staff sufficient to obtain information to answer research questions and conduct evaluation activities while still maintaining evaluator independence and objectivity (discussed in Chapter 7). It is important for the state agency to provide timely and detailed feedback to the evaluator. Regular communication allows the state agency and the evaluator to address any issues that arise and ensure that the evaluation is woven into and aligned with existing operations.
- **Clearly designate tasks.** There will be some evaluation-related tasks that are the responsibility of both the evaluator *and* the state agency. Other tasks will be the sole responsibility of the evaluator *or* the state agency. Chapter 7, Section 7.2 provided details about activities state agency staff will be heavily involved in, particularly accessing data.

Practical Steps to Maintain Communication with Evaluators

To maintain open lines of communication with evaluators, state agencies can take the following practical steps:

- **Establish regular conference calls or meetings** where the evaluator can update state agency staff on evaluation progress and the state agency staff can provide any updates on the study's implementation. State agency staff can choose to meet with their evaluator weekly, bi-weekly, monthly, or quarterly. Meetings with the evaluator will likely be longer and more frequent at the start of the evaluation and decrease in frequency and duration as the evaluation proceeds.
- **Request monthly written progress reports** wherein the evaluator describes key activities, status of existing activities, plans for upcoming activities, and any challenges encountered during the month.

8.2 Integrating the Evaluation into Existing Program Operations

The evaluator will likely design procedures for implementation of the evaluation that clearly articulate state agency staff members' roles and responsibilities. Evaluators will incorporate these procedures into existing program operations so that the evaluation can run as smoothly as possible. To do this, state agency staff need to inform evaluators about the RESEA intervention and answer questions about the intervention design, staff roles, and other items of interest.

Integrating evaluation activities into RESEA intervention activities has many benefits. It can reduce the evaluation's burden on state agency staff and adverse effects on intervention activities. It can limit claimants' awareness of evaluation activities while ensuring they are fully informed and able to consent to participate in the study. Integration also promotes staff buy-in to the evaluation.

This section covers several key points at which state agencies and their evaluators will need to work together to integrate evaluation activities into program operations:

- ✓ Designing a random assignment process (if conducting an experimental evaluation).
- ✓ Training state agency staff in evaluation procedures, including the random assignment process (if conducting an experimental evaluation).
- ✓ Developing a data extract.

Designing the Random Assignment Process

State agencies that are estimating the impact of their RESEA intervention using an experimental impact design (with random assignment of study participants) will need to work with the evaluator to determine the "point of random assignment," meaning when in the intake process claimants will be assigned to the intervention group or the comparison group.

Random assignment of claimants will likely occur once their UI claim has been approved, they have received their first UI benefit check, and they have been identified as eligible for RESEA. For state agencies interested in testing the effectiveness of offering different sets of services, random assignment will most likely occur before claimants are scheduled for their first RESEA meeting. For example, if the state agency plans to test the value of participating in the RESEA program compared to not participating (i.e., an evaluation of the whole program), claimants in the comparison group will not be scheduled for any first meeting. To ensure the comparison group does not receive the intervention (i.e., to prevent "crossover"), random assignment will need to take place before scheduling occurs. (See Appendix G for a random assignment checklist and Appendix H for an example random assignment flow chart.)

In addition to setting the point of random assignment, the evaluator will also need to establish a mechanism for assigning participants to the intervention group or comparison group. For instance, in a study of RESEA's predecessor, the Reemployment and Eligibility Assessment program (REA), state IT staff worked with the state's evaluator to insert random assignment into its standard REA scheduling algorithm.[39] Conducting random assignment manually or by clusters (i.e., assigning particular UI offices to administer one set of services versus another) is also possible, but requires that state agencies and evaluators monitor random assignment more carefully and think through additional considerations, such as potential assignment errors and selection bias. For more information on the random assignment process, please refer to Chapter 4.

Training Staff on Evaluation Procedures

Evaluators should have a plan for training state agency and local supervisors and staff, as well as training their own evaluation team, on specific study procedures. It is important that all staff be trained in study procedures in order to ensure the study is implemented consistently and as planned. Evaluators should train personnel who regularly communicate with RESEA claimants, including frontline staff at AJCs and their supervisors, state agency staff responsible for overseeing RESEA, and staff who handle data on evaluation procedures. State agencies should make sure to require a training plan in their formal statement of work for evaluators.[40]

Typically, evaluation procedures training will include:

- Information about the origins and importance of the evaluation.
- Instructions for implementing the evaluation in accordance with its design, including information about any changes in program implementation that are occurring during the study. For example, if a state agency is evaluating a specific intervention within the RESEA

[39] Klerman, J. A., Saunders, C., Dastrup, E., Epstein, Z., Walton, D., and Adam, T., with Barnow, B. S. (2019). *Evaluation of impacts of the Reemployment and Eligibility Assessment (REA) Program: Final report*. Abt Associates. Retrieved from: https://www.dol.gov/sites/dolgov/files/OASP/evaluation/pdf/REA%20Impact%20Study%20-%20Final%20Report.pdf

[40] See also the RESEA evaluation TA webinar, "Writing a Statement of Work (SOW) for an Evaluation" for further details on what should be included in a statement of work.

program using a random assignment design, evaluation training will explain how to deliver services to claimants in accordance with their random assignment status.

- Information on the logistics of the random assignment procedure, if the evaluation uses a random assignment design. The evaluator will explain how to randomly assign study participants, when to conduct random assignment, and how to communicate the results to study participants.
- Directions on human subjects protection and data security protocols for staff responsible for handling, storing, and transferring paper or electronic records, as protecting claimants' privacy is a priority of any DOL-sponsored study.
- Information that needs to be conveyed to claimants about the study. Ideally, the evaluator will give staff scripts to follow to convey that information clearly and consistently.

To supplement the staff training, state agencies should consider asking evaluators to produce written manuals that staff can refer to throughout the study's duration and be available to answer staff questions as necessary.

Developing Data Extracts from State Administrative Data

As discussed in Chapter 6, the evaluator may need access to data extracts from the state's RESEA program or UI data systems. To ensure the collected data meet the evaluation's needs, the state agency should communicate with the evaluator about those data needs, what the state's data systems can offer, and how the state agency can extract and securely transfer the data to the evaluator for analysis. Developing data extracts will likely involve several conversations, and the state agency's data management staff will produce test files containing data that the evaluator can use to run trial analyses. The state agency and the evaluator will ultimately settle on a set of data elements to include in future data files transferred for analysis.

8.3 Creating an Evaluation Design Report

The first task of the evaluator will be to develop a written plan that describes the technical approach to be taken to implement the evaluation. Typically called an evaluation design report (EDR), it describes the planned methods before analysis begins. To increase transparency, best practices involve providing the final EDR to U.S. DOL as part of the report on RESEA activities and publicly releasing it (e.g., on the state agency's or evaluator's web site). Changes thereafter are not uncommon, but they are explicitly documented and justified—sometimes in a revised EDR, in an addendum to the EDR, or in the evaluation's reports. The EDR should build on the statement of work developed for the study (which is usually is included in the Request for Proposals) as well as on any preliminary plans the state agency created in planning for the evaluation.

EDRs should include information on the following:

- The study purpose and parameters.
- The intervention being tested, including an intervention-specific logic model.
- Specific research questions that are measurable, relevant to the intervention. (Chapter 3 has information on how to develop research questions.)
- Appropriate and reliable outcomes that can be measured through available resources in a time frame that is reasonable given the intervention.

- Evaluation method(s), including the evaluation design type, methods for creating an intervention and comparison group, and data collection processes and sources. (Chapter 4 and Chapter 5 have more information on evaluation design types.)
- Data elements (i.e., inputs or activities expected to produce the primary outcomes) and proposed outcome measures that are valid, appropriate, and reliable.
- Data analysis plan and methods, including suitable strategies for mitigating any threats or risks to interpretation of findings (e.g., inconsistent or missing data) and overcoming any limitations to the extent possible.
- Timeline and milestones for intervention and evaluation activities. The EDR should note how evaluation activities and timeline dovetail with intervention activities and timeline.
- Reporting details to convey evaluation progress, results, and findings.

The EDR provides details on all of the technical aspects of the evaluation, but it also serves as a guide to program staff on how various evaluation activities link to or coordinate with program operations.

Exhibit 8-1 describes the common elements typically included in an EDR. An independent evaluator is typically responsible for writing the EDR. However, it will be beneficial for state agency staff to understand the type of content that evaluators include in the EDR in order to anticipate questions for which evaluators may require the input of state agency staff. Additionally, Appendix I provides more details on the format of an EDR as well as an EDR template.

Exhibit 8-1. Elements Included in the Evaluation Design Report

Element	Summary Description
Program purpose and logic model	Describes the purpose of the RESEA intervention that the evaluator will test (this can be an entire RESEA program or particular aspects of it). The Evaluation Design Report should provide details on each of the program's organizational and service delivery components (if more than one) and explain the theory of change. Presents a logic model that reflects the inputs, activities, outputs, outcomes, and context for the specific intervention being tested. Specifies the expected outcomes or impacts from the intervention and possible effects of other activities on the variables of interest beyond those in the evaluation.
Evidence base (literature review)	Reviews, synthesizes, and summarizes the existing evidence to understand gaps in RESEA-relevant research or replicate a comparable study. The evaluator should include research on relevant past interventions and evaluations, including findings and a discussion of research designs used, if applicable. Describes how the evaluation will enhance the state's RESEA program and/or contribute to the workforce evaluation literature.
Study purpose and what will be tested	Articulates the specific purpose of the study and describes what aspects of the program the evaluator will test. Notes what program sites or geographic areas will be included in the study as well as what population will be studied.
Key and detailed research questions	Identifies the research questions. Aligns research questions with the intervention's logic model. Includes descriptive/process questions, general outcome questions, and/or impact questions (i.e., hypotheses about expected outcome changes due to the intervention), depending on the type of evaluation being conducted.
Evaluation type	Describes the evaluation design type(s) selected as well as the methods the evaluator will use. Provides the justification for the approach selected. Describes the overarching evaluation goals or objectives and explains their appropriateness to the evaluation design and the intervention.
Participants, samples, and units of analysis	Describes the unit(s) of analysis (which will most likely be RESEA-eligible UI claimants); the eligibility or exclusion criteria for RESEA; the overall population targeted or from which generalizations will be made; whether the evaluation will be conducted on the entire population or a sample; and, if a sample, whether it is representative.

Element	Summary Description
Data collection plan	Describes how the evaluator will collect and use data to measure outcomes of interest. Identifies which data elements will support evaluation activities, identifies the source for each element, and describes efforts to collect and securely store data while protecting claimants' PII. Evaluators and RESEA staff will need to pay careful attention to protecting PII, and plans should specify when and how data will be destroyed after the completion of evaluation activities. Describes whether the evaluation will engage in new data collection efforts (e.g., staff or participant survey), use existing administrative sources (secondary data collection; e.g., from a management information system), or some combination thereof. Includes a detailed timeline for data collection activities, including timelines related to creating data sharing agreements, obtaining test data extracts, and conducting analysis. If new data collection is proposed, the plan includes drafts of data collection protocols/instruments that will address the research questions.
Analysis plan	Describes an analysis plan based on the research questions and type of evaluation selected. For example, impact studies may include sampling plans that describe the purpose, method of sampling, and anticipated sample sizes. Random assignment impact studies should include the power calculations to use; describe the process for random assignment of participants to intervention and comparison groups; and demonstrate any impact formulas and other analytical assumptions. Describes the analytical software or other tools appropriate to the evaluation design. Discusses validity/threats and mitigation strategies—whether issues of internal and external validity exist, threats to validity and their implications, and strategies to mitigate selection bias, if needed.
Reporting	Describes how the evaluator plans to share evaluation results. The standard final deliverable for rigorous evaluation efforts is a comprehensive final report that describes all aspects of the evaluation, including the intervention tested, the population and outcomes of interest, and the methods used to measure outcomes. A final report tends to be a long and dense, relatively technical document. If a state agency is interested in shorter reporting deliverables written for a wider or lay audience, they can ask the evaluator to write briefs, presentations, or other products that will communicate results. The EDR should identify all written deliverables, any specific provisions associated with each, due dates for deliverables, ongoing progress, and handling data sets and a de-identified public use data set(s) at the conclusion of the study (as appropriate).

8.4 Protecting the Rights of Human Research Subjects

A key evaluation project activity is protecting the rights of human research subjects; that is, the claimants included in the study. To measure the effectiveness of the RESEA intervention, the evaluation requires collecting and storing detailed information from and about those claimants. Treating this information responsibly and protecting the privacy of these data and human subjects' other rights will be one of the state agency's most fundamental responsibilities. It is also a requirement for all federally- funded research.

The following subsections provide an overview of the federal requirements and guidelines for protecting human research subjects—including the role of an Institutional Review Board (IRB), the IRB

Key Terms

- **Human subject.** A living individual about whom an investigator (whether professional or student) is conducting research: (1) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (2) Obtains, uses, studies, or analyzes, or generates identifiable private information or identifiable biospecimens. (Source: Code of Federal Regulations 29CFR 21.102(e)(1)).
- **Personally identifiable information.** Information 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. PII includes names, Social Security numbers, birthdates, addresses, and other related contact information.
- **Informed consent.** Process of providing an individual and their legally authorized representative information (as required by federal, state and local laws) for making a voluntary and informed decision whether or not to participate in a study.

process, and key other protection processes for which the state agency and the evaluator are responsible.

Human Subjects Protections

Federally-funded research and evaluation involving human subjects must comply with federal and state laws and regulations governing the ethical treatment and the rights of research participants. Most of these regulations are based on the 1979 **Belmont Report**,[41] a summary of ethical principles and guidelines for protecting research participants. The Belmont Report acknowledges both the social benefit that research has produced and the ethical issues it has sometimes posed. The report states that research studies must be guided by three ethical principles: **respect for persons**, **beneficence**, and **justice**. Evaluators must carry these principles into their research protocols and activities.

Ethical Principles

- **Respect for persons.** Consent to be in a study should be informed and voluntary.
- **Beneficence.** Do no harm; maximize benefits of research and minimize risks to participant.
- **Justice.** Fair and equitable distribution of the benefits and burdens of research.

The regulation most commonly applicable in evaluation is the Federal Policy for the Protection of Human Subjects, also known as "**the Common Rule.**"[42] Federal and state privacy laws also usually govern federally funded research and evaluation.

In order to comply with all principles and applicable laws, each organization that conducts research must have or hire an IRB to ensure the protection of human subjects. IRBs serve as an independent and objective ethics committee responsible for overseeing research that involves human subjects. The IRB reviews and monitors steps researchers must take to protect human subjects.

Ensuring human research subjects are protected in accordance with federal, state, and local laws has very practical implications for the evaluation timeline and budget. It thus affects other major features of a study. This means that conducting research is not as easy as simply gathering and analyzing data; it requires a plan for study recruitment, consent, data collection, data management, analysis, and reporting in order to ensure that its human research subjects are protected. Even studies that use administrative records and never survey or interview participants directly must comply with human subjects considerations and applicable privacy laws. For example, human subjects considerations should factor into a state agency's decisions on:

- Selecting an evaluator with sufficient experience and training[43] to protect human research subjects and manage the IRB review process.
- Structuring data collection efforts such that data are transferred, stored, and analyzed securely and risks to human subjects are minimized.

---

[41] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research.*http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html

[42] Federal Policy for the Protection of Human Subjects, 45 C.F.R. § 46 2018.https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML

[43] Anyone conducting human subjects research (even research exempt from human subjects protection requirements) needs to periodically complete human subjects training. The type and frequency of training required will depend on institutional policy and the IRB. One common source of such training is the CITI Program: https://about.citiprogram.org/en/homepage/.

- Allowing sufficient time for IRB review in the evaluation and intervention timelines.
- Having sufficient resources to carry out the evaluation, including training the study team and monitoring evaluation implementation.
- Implementing technology solutions (e.g., encryption software, secure file transfer protocol [FTP]) to securely transfer and store participant information to comply with applicable federal and state laws governing such data.

IRB Process Overview

To ensure appropriate protections are in place and maintained, IRBs provide an independent and objective review of human subjects research. Prior to giving its approval, an IRB may stipulate modifications to the study's design, methods, or procedures to better protect the human subjects. IRB review must occur prospectively—that is, evaluators are required to obtain IRB approval before the study begins and evaluation procedures are implemented. Prospective IRB approval is also required before the state agency or the evaluator may make any modifications to IRB-approved research protocol.

IRBs are also responsible for monitoring active research projects by periodically 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 applicable rules to protect research participants from harm. Each IRB has its own policies and requirements, but Exhibit 8-2 summarizes the main roles and responsibilities of IRBs.

Exhibit 8-2. Typical IRB Responsibilities and Study Criteria for Obtaining IRB Approval

Responsibility	Criterion
The IRB must: • Review and approve prospective 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.	The study must: • Minimize risks to participants. • Show benefits to society and/or participants outweigh risks. • Select participants equitably to distribute burden. • Obtain and document informed consent process (unless eligible for waiver). • Monitor welfare of participants for safety and complaints. • Minimize risks to privacy and confidentiality. • Implement additional safeguards to protect rights and welfare of those likely to be vulnerable to coercion or undue influence (e.g., children, individuals with impaired decision-making capacity, "persons who are economically or educationally disadvantaged").

IRBs are registered and regulated by the U.S. Department of Health and Human Services. In general, most institutions that regularly conduct research (e.g., medical centers, universities, research organizations, and state agencies) have their own IRBs. Commercial IRBs are also available for a fee.

All institutions, and some government agencies, engaged in federally-funded human subjects research must have a Federal Wide Assurance for the Protection of Human Subjects number to show that they have committed to compliance with the federal regulations. If an evaluator or prospective evaluator is not sure whether the organization has a Federal Wide Assurance number

that can be checked at the on-line database of the Department's Office for Human Research Protections.[44]

Typical IRB Review Levels

- **Exempt** may have different meanings depending on the IRB, but generally it means the study is exempt from the federal regulatory requirements.There are many complex categories of exemption that only a trained IRB professional should make. "Exempt" does not necessarily mean no review needed, because the study still has to comply with ethical standards as well as applicable laws such as privacy laws. Exempt-level studies present minimal risk to its subjects and might involve collection of anonymous data, non-sensitive topics, educational tests, or observations in public settings.
- **Exempt with Limited IRB Review** is a level of review that allows exemption of surveys or interviews involving sensitive data as long as the IRB reviews the data security procedures to protect the data.
- **Expedited.** Similar to Exempt status, Expedited Review does not require review by the full IRB committee. Studies at this level pose minimal risks to subjects and might collect data in a manner that is not anonymous.
- **Full Board.** Studies requiring review by the full IRB committee pose more than minimal risk to subjects and/or may involve vulnerable populations. Full Board reviews require more time than do Exempt or Expedited reviews.

An evaluator should possess demonstrated experience with the overall process of receiving approval for a study before engaging any human subjects. The IRB will let the evaluator know whether the study is "Exempt" from review (see the box above for details) or is not exempt and needs to undergo review by the IRB. Note that even Exempt studies need to comply with local policies, ethical standards, and privacy laws governing the data, which may dictate the consent process is required. The IRB should have experience reviewing social science studies and, ideally, have experience reviewing labor or workforce studies. Some IRBs may focus more on reviewing biomedical studies and be a less than ideal fit to review studies related to labor and workforce.

Federal and state regulations and specific IRB's policies define the criteria for exempting a study from IRB review or assigning the level of IRB review required. These criteria are based on considerations including but not limited to:

- **Level of risk to subjects of the study.** Risk is considered both overall and relative to the potential social benefit of the research. The biggest risk to claimants in evaluations of RESEA interventions will likely be the risk of a data breach.
- **Vulnerability of the population in the study.** IRBs may require additional protections for studies that include vulnerable subjects in their sample. For example, children and prisoners are considered especially vulnerable groups.

Generally, all evaluations that use federal funds need to engage with an IRB for an initial review of the study, its design, and its materials to determine whether the evaluation is Exempt or requires IRB review, and if so, at what level. In general, the IRB, not the evaluator, determines whether the proposed study is Exempt (but evaluators and funding agencies should check the existing policy on who can make IRB exemptions and determinations).

The IRB review process could take from one to several months (see Chapter 7 for more information on timeline). The time it takes to prepare an application for IRB review, have it reviewed, respond

[44] https://ohrp.cit.nih.gov/search/fwasearch.aspx?styp=bsc

to comments or requests for revisions, and receive approval depends on the IRB's process and on the complexity of the evaluation and the level of risk it poses to participants.

IRB reviews generally consider four major areas:

- The study's procedures for informed consent, a process by which potential human subjects become aware of the risks and benefits associated with participating in research in order to make a voluntary and informed decision.
- How the evaluator will protect subjects' privacy and confidentiality.
- The plan for data security.
- How the evaluator will handle adverse events and unanticipated problems.

These factors have very real, on-the-ground implications for how the evaluation will unfold. As the sponsor for the evaluation, state agencies will want to be aware of and may be involved in executing these aspects of the study protocol.

Informed Consent

To collect information from claimants participating in the study (e.g., through a survey, interviews, focus groups), the evaluator needs to obtain their legally effective written informed consent (or consent of a legally authorized representative such as parent or guardian if the participant is a minor). Obtaining informed consent includes giving prospective study participants sufficient time and opportunity to consider participation and minimizing "undue influence," such as offering an excessive reward for participating in the study, and "coercion," such as an implicit or explicit threat that a participant can lose access to a service if they do not agree to participate.

Informed consent is a *process*, not just a form to be signed, that begins with explaining the study in language appropriate to the participants (an eighth-grade reading level is usually the standard, but it is important to check with the IRB). If participants are not proficient in English, the consent form and data collection instruments need to be translated by a certified translator. The information provided to prospective participants during the consent process must help them understand the implications and risks of participation. It is essential to disclose all relevant information honestly and to give each potential participant the opportunity to ask questions and receive answers to their questions.

Informed consent processes, including scripts for any verbal consent procedures, should be fully described in IRB submission. Waivers of written consent and waivers of consent are possible but need to be granted by the IRB.

Understand the Required Elements of Consent

Unless the IRB has approved a waiver of consent, evaluators must obtain informed consent from study participants before collecting information from or about them. State RESEA intervention administrators don't need to be involved in the informed consent process. But they should make sure that their evaluator has steps in place to obtain informed consent. State RESEA intervention administrators should also become familiar with the required elements of consent:

- Statement explaining research purpose, duration of participation, procedures to be followed, and voluntary nature of participation.
- Description of risks and benefits related to participation.
- Statement describing extent to which confidentiality will be maintained.
- Information about whom to contact for questions about the research and research subjects' rights.

Privacy and Confidentiality

Once claimants consent to the study, the evaluator is required to protect their privacy and the confidentiality of their information by preventing its accidental disclosure or loss. Protecting privacy means collecting data in such a way that claimants' information or PII is not seen or overheard by others. For example, claimants might take surveys in a private space where others cannot view their answers.

Protecting confidentiality of participant information means:

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

As a reminder, RESEA data is confidential UC data, making it subject to federal regulations that require data users to uphold the confidentiality of "any UC information which reveals the name or any identifying particular about any individual or any past or present employer or employing unit, or which could foreseeably be combined with other publicly available information to reveal any such particulars."[45] Disclosure of UC data, including wage records and UC claims information, must meet the appropriate privacy and confidentiality protections outlined in 20 CFR Part 603.

Disclosure of confidential UC information to a third party for purposes of conducting an RESEA evaluation is allowable provided the appropriate agreements are in place between the public official responsible for the state's involvement in the evaluation and the evaluator. If the public official responsible for the state's involvement in RESEA evaluation is not a party to the data sharing agreements, then disclosure to the evaluator must align with the informed consent parameters found in 20 CFR Part 603.

Many state workforce agencies and RESEA programs already have procedures in place to protect claimants' privacy and confidentiality of their information. State RESEA programs should work closely with their evaluator to ensure claimants' privacy and confidentiality are protected during the evaluation, particularly when new data collection instruments are deployed. An experienced evaluator can be a valuable resource on how to best do this.

Data Security

State RESEA agencies and their evaluators have a collective responsibility to protect participant information by ensuring that data are **collected, transferred, stored, and analyzed securely**. All parties will need to work together to create a climate of accountability and responsibility for the data collected, shared, and analyzed. See Chapter 6 for more information on data security.

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 concerning the protection of study participants' rights and welfare include the following:

- A participant is visibly upset by the questions during a survey.
- The evaluator or local staff learn that a participant is at risk of harm.

---

[45] Federal-State Unemployment Compensation (UC) Program: Confidentiality and Disclosure of State UC Information, 20 C.F.R. §§ 603.3-603.12 2006. https://www.govinfo.gov/content/pkg/CFR-2012-title20-vol3/pdf/CFR-2012-title20-vol3-part603.pdf

- A participant or his/her parent/guardian has serious concerns about the study. [46]
- Study procedures were not followed (e.g., claimant data was inappropriately shared, stored, or analyzed).
- Study data are lost (e.g., consent forms, paper-and-pencil surveys).

The selected evaluator should be prepared for such potential problems. State agency staff also may want to have specific procedures in place for the state agency staff, the evaluator, and other key stakeholders to work together to resolve such issues should they arise.

8.5 Communicating Evaluation Results

Though the final results of the evaluation will likely not be available for several years, the state agency should give careful consideration to how to disseminate findings during the evaluation and after it is completed. Reports are a common dissemination product that summarize the various aspects of the intervention tested, the methods for testing the intervention, and the evaluation results. Many state agencies ask evaluators to produce **interim reports** (e.g., on implementation of program procedures), the **final report** that discusses all evaluation findings, and **other dissemination products** (e.g., briefs, articles, presentations, social media). Reports that come out on a periodic basis, as parts of the study are complete or when findings are ready, can increase the value and relevance of the study by making timely findings available to decision-makers. Periodic reporting can also help the evaluation project continue to maintain engagement with stakeholders throughout the period of performance. Finally, depending on the type of study conducted, the evaluator also may prepare and submit a **public use data set** as part of final reporting deliverables.

Reports convey the evaluation's findings to relevant stakeholders, including the state agency, other state stakeholders, U.S. DOL, the larger workforce community, and potential funders. U.S. DOL's CLEAR routinely reviews labor evaluations, using the details in the final publicly available report. The state agency could make reports publicly available on its website and disseminate them through other channels. Making the report publically available will allow other states and interested stakeholders to learn from the evaluation and support continuous learning in the RESEA community. Ultimately, the state agency and the RESEA program administrator will need to determine what types of reporting activities they want the evaluator to complete and how the evaluation's results will be disseminated.

Communicating the State Agency's Requirements for Reporting Study Findings

State agency staff play an important role in selecting an evaluator to conduct the study—whether a third-party evaluator via an RFP, a university or other organization partner to conduct the study jointly, or an in-house entity. In all cases, the state agency will need to clearly communicate expectations for how the evaluator will report the evaluation findings. These can include full written reports, short briefs, podcasts, and presentations. Note that additional reporting requirements increase evaluation costs and stretch timelines. It will be important to balance the value of the information desired with cost considerations.

Key things to communicate in an RFP or other written instructions to the evaluator include:

- How many reporting products and what types, what formats, and their approximate due dates/time frames.

[46] Please note that minors in involved in research studies would need additional protections.

- For progress reports, how often they are written (e.g., monthly, quarterly) and what aspects of evaluation activity do they need to cover.
- For interim reports and the final report, how many interim reports and what they should cover, when drafts must be delivered for review, and when final versions incorporating comments and revisions are due.

State agency staff will also review evaluator reports and other products. As discussed in Chapter 7, given that the evaluation is to be "independent," such review should aim to be constructive. In their review, state agencies can provide feedback on whether the evaluator has delivered the products specified in their statement of work and that the delivered products follow appropriate and agreed upon methods, are clearly written, and accurately describe the RESEA intervention. Beyond that, independent evaluations necessitate that evaluators provide their interpretation of the findings of the evaluation. Sometimes those findings are uniformly and strongly positive; that is, the intervention as it is currently operated works well and strongly improves outcomes. Other times, the evaluation results may be mixed or even negative (e.g., the intervention has no impact on claimants' outcomes). State agency staff can and should comment on how those findings are framed and suggest alternative interpretations. The evaluation is a product of the evaluator, not of the state agency. The evaluator is the author of the evaluation's reports and has final say on the findings and language. State agency staff cannot seek changes that would skew the reporting (e.g., in order to make the results seem more positive).

Types of Dissemination Activities

This section describes various types of reporting that state agencies can request from evaluators, including their typical content and place in an evaluation's timeline.

Interim Evaluation Reports

These reports convey findings from the program evaluation prior to completion of the final analyses. Interim reports may include short- and medium-term claimant outcomes or analyses of program implementation. For multi-year evaluations, interim reports are often submitted at the evaluation's mid-point, but they can also be submitted annually. An interim report may also require multiple rounds of revisions to ensure that it meets quality standards for public release.

Some interim reports may discuss data collected during implementation of the evaluation, such as information from site visits, interviews, and focus groups. Other interim reports could analyze outcomes for early cohorts of study participants, such as the first group of training completers in an evaluation of a training intervention. An interim report might discuss preliminary outcomes to the extent that these are available, but it does not serve as the final assessment of the intervention's impact on claimant outcomes, which is the role of the final report.

Final Report

This report serves as the capstone to the evaluation. If the evaluation includes multiple study types (e.g., an implementation study and an impact study), an evaluation may have multiple different reports (one for each study) that may be finalized at different points in the timeline. Final reports are generally prepared for public release. For an experimental impact evaluation, the final report presents program impacts on key outcomes. Such a report is completed only after allowing for time to collect follow-up data on participant outcomes. For example, if a state agency wants its evaluator to track participant outcomes 12 months after program enrollment, the deadline for the final report would need to fall after the 12-month follow-up period is over for the last claimant enrolled, plus time for analysis and writing.

Key topics typically covered by the final report include the following:

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

The final report should contain sufficient details that another evaluator could reproduce the evaluation. Given the amount of information that needs to be covered, the final report will be longer than interim reports. A final report should also be written in such a way that non-technical audiences will be able to understand and consume the information.

Compared with other report types, a final report takes time for analysis, writing, and revisions. The amount of time needed will depend on the sample size, amount of data, complexity of analyses, and the desired level of detail. Once the final report is drafted, it should be reviewed for correctness and quality.[47]

Briefs

Briefs are shorter documents designed for a broad and public audience. These dissemination products tend to focus on a single or a couple of findings and the insights they yielded and highlight less the methods and technical aspects of the study. Briefs (and presentations, discussed next) can be especially useful in communicating results to stakeholders who are less familiar with the intervention and/or evaluations in general, including policymakers, outside funders, and the general public. Given their target audience, it is important these products are short, visually interesting, and written in a style that is not overly technical or dense.

Presentations

Like briefs, presentations are shorter than reports, using a combination of bulleted text and visuals to succinctly convey high-level information about the intervention, evaluation design, and key findings. Presentations are helpful for sharing information both internally and publicly. Presentations might be a good approach for communicating findings from the evaluation to a lay audience. The level of detail and technical information contained in the presentation will vary depending on the audience.

[47] OUI's guidance has suggested states contact CLEAR when reports are publically released so that CLEAR can include them in evidence reviews.

Appendices

A. **SSA Authorizing Legislation**

B. **Unemployment Insurance Program Letter No. 1-20**

C. **Reemployment Services and Eligibility Assessments (RESEA) Evaluation Resource List**

D. **Logic Model Template**

E. **Data Availability Questionnaire**

F. **Evaluation Timeline Overview and Worksheet**

G. **Random Assignment Checklist**

H. **Sample Random Assignment Flowchart**

I. **Evaluation Design Report Template**

J. **Glossary of Terms**

Appendix A. SSA-2018 Authorizing Legislation

SEC. 306. [42 U.S.C. 506] GRANTS TO STATES FOR REEMPLOYMENT SERVICES AND ELIGIBILITY ASSESSMENTS.

(a) IN GENERAL.—The Secretary of Labor (in this section referred to as the "Secretary") shall award grants under this section for a fiscal year to eligible States to conduct a program of reemployment services and eligibility assessments for individuals referred to reemployment services as described in section 303(j) for weeks in such fiscal year for which such individuals receive unemployment compensation.

(b) PURPOSES.—The purposes of this section are to accomplish the following:

(1) To improve employment outcomes of individuals that receive unemployment compensation and to reduce the average duration of receipt of such compensation through employment.

(2) To strengthen program integrity and reduce improper payments of unemployment compensation by States through the detection and prevention of such payments to individuals who are not eligible for such compensation.

(3) To promote alignment with the broader vision of the Workforce Innovation and Opportunity Act (29 U.S.C. 3101 et seq.) of increased program integration and service delivery for job seekers, including claimants for unemployment compensation.

(4) To establish reemployment services and eligibility assessments as an entry point for individuals receiving unemployment compensation into other workforce system partner programs.

(c) EVIDENCE-BASED STANDARDS.—

(1) IN GENERAL.—In carrying out a State program of reemployment services and eligibility assessments using grant funds awarded to the State under this section, a State shall use such funds only for interventions demonstrated to reduce the number of weeks for which program participants receive unemployment compensation by improving employment outcomes for program participants.

(2) EXPANDING EVIDENCE-BASED INTERVENTIONS.—In addition to the requirement imposed by paragraph (1), a State shall—

(A) for fiscal years 2023 and 2024, use no less than 25 percent of the grant funds awarded to the State under this section for interventions with a high or moderate causal evidence rating that show a demonstrated capacity to improve employment and earnings outcomes for program participants;

(B) for fiscal years 2025 and 2026, use no less than 40 percent of such grant funds for interventions described in subparagraph (A); and

(C) for fiscal years beginning after fiscal year 2026, use no less than 50 percent of such grant funds for interventions described in subparagraph (A).

(d) EVALUATIONS.—

(1) REQUIRED EVALUATIONS.—Any intervention without a high or moderate causal evidence rating used by a State in carrying out a State program of reemployment services and eligibility assessments under this section shall be under evaluation at the time of use.

(2) FUNDING LIMITATION.—A State shall use not more than 10 percent of grant funds awarded to the State under this section to conduct or cause to be conducted evaluations of interventions used in carrying out a program under this section (including evaluations conducted pursuant to paragraph (1)).

(e) STATE PLAN.—

(1) IN GENERAL.—As a condition of eligibility to receive a grant under this section for a fiscal year, a State shall submit to the Secretary, at such time and in such manner as the Secretary may require, a State plan that outlines how the State intends to conduct a program of reemployment services and eligibility assessments under this section, including—

(A) assurances that, and a description of how, the program will provide—

(i) proper notification to participating individuals of the program's eligibility conditions, requirements, and benefits, including the issuance of warnings and simple, clear notifications to ensure that participating individuals are fully aware of the consequences of failing to adhere to such requirements, including policies related to non-attendance or non-fulfillment of work search requirements; and

(ii) reasonable scheduling accommodations to maximize participation for eligible individuals;

(B) assurances that, and a description of how, the program will conform with the purposes outlined in subsection (b) and satisfy the requirement to use evidence based standards under subsection (c), including—

(i) a description of the evidence-based interventions the State plans to use to speed reemployment;

(ii) an explanation of how such interventions are appropriate to the population served; and

(iii) if applicable, a description of the evaluation structure the State plans to use for interventions without at least a moderate or high causal evidence rating, which may include national evaluations conducted by the Department of Labor or by other entities; and

(C) a description of any reemployment activities and evaluations conducted in the prior fiscal year, and any data collected on—

(i) characteristics of program participants;

(ii) the number of weeks for which program participants receive unemployment compensation; and

(iii) employment and other outcomes for program participants consistent with State performance accountability measures provided by the State unemployment compensation program and in section 116(b) of the Workforce Innovation and Opportunity Act (29 U.S.C. 3141(b)).

(2) APPROVAL.—The Secretary shall approve any State plan, that is timely submitted to the Secretary, in such manner as the Secretary may require, that satisfies the conditions described in paragraph (1).

(3) DISAPPROVAL AND REVISION.—If the Secretary determines that a State plan submitted pursuant to this subsection fails to satisfy the conditions described in paragraph (1), the Secretary shall—

(A) disapprove such plan;

(B) provide to the State, not later than 30 days after the date of receipt of the State plan, a written notice of such disapproval that includes a description of any portion of the plan that was not approved and the reason for the disapproval of each such portion; and

(C) provide the State with an opportunity to correct any such failure and submit a revised State plan.

(f) ALLOCATION OF FUNDS.—

(1) BASE FUNDING.—

(A) IN GENERAL.—For each fiscal year after fiscal year 2020, the Secretary shall allocate a percentage equal to the base funding percentage for such fiscal year of the funds made available for grants under this section among the States awarded such a grant for such fiscal year using a formula prescribed by the Secretary based on the rate of insured unemployment (as defined in section 203(e)(1) of the Federal-State Extended Unemployment Compensation Act of 1970 (26 U.S.C. 3304 note)) in the State for a period to be determined by the Secretary. In developing such formula with respect to a State, the Secretary shall consider the importance of avoiding sharp reductions in grant funding to a State over time.

(B) BASE FUNDING PERCENTAGE.—For purposes of subparagraph (A), the term ‘‘base funding percentage’’ means—

(i) for fiscal years 2021 through 2026, 89 percent; and

(ii) for fiscal years after 2026, 84 percent.

(2) RESERVATION FOR OUTCOME PAYMENTS.—

(A) IN GENERAL.—Of the amounts made available for grants under this section for each fiscal year after 2020, the Secretary shall reserve a percentage equal to the outcome reservation percentage for such fiscal year for outcome payments to increase the amount otherwise awarded to a State under paragraph (1). Such outcome payments shall be paid to States conducting reemployment services and eligibility assessments under this section that, during the previous fiscal year, met or exceeded the outcome goals provided in subsection (b)(1) related to reducing the average duration of receipt of unemployment compensation by improving employment outcomes.

(B) OUTCOME RESERVATION PERCENTAGE.—For purposes of subparagraph (A), the term ‘‘outcome reservation percentage’’ means—

(i) for fiscal years 2021 through 2026, 10 percent; and

(ii) for fiscal years after 2026, 15 percent.

(3) RESERVATION FOR RESEARCH AND TECHNICAL ASSISTANCE.—Of the amounts made available for grants under this section for each fiscal year after 2020, the Secretary may reserve not more than 1 percent to conduct research and provide technical assistance to States.

(4) CONSULTATION AND PUBLIC COMMENT.—Not later than September 30, 2019, the Secretary shall—

(A) consult with the States and seek public comment in developing the allocation formula under paragraph (1) and the criteria for carrying out the reservations under paragraph (2); and

(B) make publicly available the allocation formula and criteria developed pursuant to subclause (A).

(g) NOTIFICATION TO CONGRESS.—Not later than 90 days prior to making any changes to the allocation formula or the criteria developed pursuant to subsection (f)(5)(A), the Secretary shall submit to Congress, including to the Committee on Ways and Means and the Committee on Appropriations of the House of Representatives and the Committee on Finance and the Committee on Appropriations of the Senate, a notification of any such change.

(h) SUPPLEMENT NOT SUPPLANT.—Funds made available to carry out this section shall be used to supplement the level of Federal, State, and local public funds that, in the absence of such availability, would be expended to provide reemployment services and eligibility assessments to individuals receiving unemployment compensation, and in no case to supplant such Federal, State, or local public funds.

(i) DEFINITIONS.—In this section:

(1) CAUSAL EVIDENCE RATING.—The terms ''high causal evidence rating'' and ''moderate causal evidence rating'' shall have the meaning given such terms by the Secretary of Labor.

(2) ELIGIBLE STATE.—The term ''eligible State'' means a State that has in effect a State plan approved by the Secretary in accordance with subsection (e).

(3) INTERVENTION.—The term ''intervention'' means a service delivery strategy for the provision of State reemployment services and eligibility assessment activities under this section.

(4) STATE.—The term ''State'' has the meaning given the term in section 205 of the Federal-State Extended Unemployment Compensation Act of 1970 (26 U.S.C. 3304 note).

(5) UNEMPLOYMENT COMPENSATION.—The term unemployment compensation means ''regular compensation'', ''extended compensation'', and ''additional compensation'' (as such terms are defined by section 205 of the Federal-State Extended Unemployment Compensation Act of 1970 (26 U.S.C. 3304 note)).

(b) Report.--Not later than 3 years after the date of enactment of this Act, the Secretary of Labor shall submit to Congress a report to describe promising interventions used by States to provide reemployment assistance.

(c) Adjustment to Discretionary Spending Limits.--Section 251(b)(2) of the Balanced Budget and Emergency Deficit Control Act of 1985 (2 U.S.C. 901(b)(2)) is amended by adding at the end the following: (E) Reemployment services and eligibility assessments.—

(i) In general.--If a bill or joint resolution making appropriations for a fiscal year is enacted that specifies an amount for grants to States under section 306 of the Social Security Act, then the adjustment for that fiscal year shall be the additional new budget authority provided in that Act for such grants for that fiscal year, but shall not exceed--

(I) for fiscal year 2018, $0;

(II for fiscal year 2019, $33,000,000;

(III) for fiscal year 2020, $58,000,000; and

(IV) for fiscal year 2021, $83,000,000.

(ii) Definition.--As used in this subparagraph, the term `additional new budget authority' means the amount provided for a fiscal year, in excess of $117,000,000, in an appropriation Act and specified to pay for grants to States under section 306 of the Social Security Act.''.

(d) Other Budgetary Adjustments.--Section 314 of the Congressional Budget Act of 1974 (2 U.S.C. 645) is amended by adding at the end the following: (g) Adjustment for Reemployment Services and Eligibility Assessments.--

(1) In general.—

(A) Adjustments.--If the Committee on Appropriations of either House reports an appropriation measure for any of fiscal years 2022 through 2027 that provides budget authority for grants under section 306 of the Social [[Page 132 STAT. 132]] Security Act, or if a conference committee submits a conference report thereon, the chairman of the Committee on the Budget of the House of Representatives or the Senate shall make the adjustments referred to in subparagraph(B) to reflect the additional new budget authority provided for such grants in that measure or conference report and the outlays resulting therefrom, consistent with subparagraph (D).

(B) Types of adjustments.--The adjustments referred to in this subparagraph consist of adjustments to--

(i) the discretionary spending limits for that fiscal year as set forth in the most recently adopted concurrent resolution on the budget;

(ii) the allocations to the Committees on Appropriations of the Senate and the House of Representatives for that fiscal year under section 302(a); and

(iii) the appropriate budget aggregates for that fiscal year in the most recently adopted concurrent resolution on the budget.

(C) Enforcement.--The adjusted discretionary spending limits, allocations, and aggregates under this paragraph shall be considered the appropriate limits, allocations, and aggregates for purposes of congressional enforcement of this Act and concurrent budget resolutions under this Act.

(D) Limitation.--No adjustment may be made under this subsection in excess of--

(i) for fiscal year 2022, $133,000,000;

(ii) for fiscal year 2023, $258,000,000;

(iii) for fiscal year 2024, $433,000,000;

(iv) for fiscal year 2025, $533,000,000;

(v) for fiscal year 2026, $608,000,000; and

(vi) for fiscal year 2027, $633,000,000.

(E) Definition.--As used in this subsection, the term `additional new budget authority' means the amount provided for a fiscal year, in excess of $117,000,000, in an appropriation measure or conference report (as the case may be) and specified to pay for grants to States under section 306 of the Social Security Act.

(2) Report on 302(b) level.--Following any adjustment made under paragraph (1), the Committees on Appropriations of the Senate and the House of Representatives may report appropriately revised suballocations pursuant to section 302(b) to carry out this subsection.''

Appendix B. Unemployment Insurance Program Letter No. 1-20

ADVISORY: UNEMPLOYMENT INSURANCE PROGRAM LETTER NO. 1-20

TO: STATE WORKFORCE AGENCIES
FROM: JOHN PALLASCH /s/ Assistant Secretary

SUBJECT: Expectations for States Implementing the Reemployment Service and Eligibility Assessment (RESEA) Program Requirements for Conducting Evaluations and Building Program Evidence

1. **Purpose.** To provide States with guidance and expectations regarding the implementation of the RESEA evaluation and evidence requirements.

2. **Action Requested.** The Department of Labor's (DOL's) Employment and Training Administration (ETA) requests State Workforce Administrators to provide information contained in this Unemployment Insurance Program Letter (UIPL) to appropriate program and other staff in the state's workforce system. This information should be shared with staff in, but not limited to, the Unemployment Insurance (UI) program, workforce programs administered under the Workforce Innovation and Opportunity Act (WIOA), including the Wagner-Peyser Employment Service, and workforce information/labor market information programs..

3. **Summary and Background.**

- Summary – In accordance with the statutory provisions for RESEA contained in the Social Security Act (SSA), states are expected to begin conducting evaluations of RESEA interventions and service delivery strategies no later than fiscal year (FY) 2020 to support building new evidence on effective RESEA interventions that all states can rely on in designing and delivering the RESEA program.

- Background – On February 9, 2018, the President signed the Bipartisan Budget Act of 2018, Public Law No. 115-123 (BBA), which amended the SSA and created a permanent authorization for the RESEA program. The RESEA provisions are contained in Section 30206 of the BBA, enacting new Section 306 of the SSA.

  Section 306 of the SSA includes a tiered evidence approach for the RESEA program to encourage states to use evidence-based strategies, where they exist, and to conduct evaluations and build evidence for other interventions and service delivery strategies. The goal is to ensure that each state employs RESEA interventions and service delivery 2 strategies that are based on rigorous causal evidence from evaluations rated as "high or moderate causal" and are shown to reduce benefit duration as a result of improved employment outcomes. In addition, states using interventions or service delivery strategies without such evidence must be under evaluation at the time of use to determine their effectiveness in achieving this goal. Over time, as the RESEA program uses this tiered evidence approach, [48] states will add to the evidence base; grow the workforce system's understanding of what interventions work well, for whom, and in what contexts; and

[48] "Tiered evidence" refers to a policy tool that allows federal agencies to tie federal funding to strategies with evidence, to encourage the use of interventions that have strong evidence of success and test promising new ideas. With the RESEA program, the legislation ties certain levels of future funding to interventions with moderate or high causal evidence ratings, to encourage the use of those interventions that have stronger evidence that they "work," and requires interventions without those ratings to be to under evaluation at the time of use.

expand the use of interventions with strong evidence of success. Similar tiered evidence models are used across federal government programs such as the Department of Health and Human Services' Home Visiting Program.

The statute, in section 306(c), SSA, requires states to use RESEA grant funds for evidence-based interventions or service delivery strategies that reduce the average number of weeks participants receive benefits by improving employment outcomes, including earnings. Specifically, it requires the following with regard to evidence building and evaluations:

"(c) EVIDENCE-BASED STANDARDS.—

"(1) IN GENERAL.--In carrying out a State program of reemployment services and eligibility assessments using grant funds awarded to the State under this section, a State shall use such funds only for interventions demonstrated to reduce the number of weeks for which program participants receive unemployment compensation by improving employment outcomes for program participants.

"(2) EXPANDING EVIDENCE-BASED INTERVENTIONS.--In addition to the requirement imposed by paragraph (1), a State shall –

(A) for fiscal years 2023 and 2024, use no less than 25 percent of the grant funds awarded to the State under this section for interventions with a high or moderate causal evidence rating that show a demonstrated capacity to improve employment and earnings outcomes for program participants.

(B) for fiscal years 2025 and 2026, use no less than 40 percent of such grant funds for interventions described in sub-paragraph (A); and

(C) for fiscal years beginning after fiscal year 2026, use no less than 50 percent of such grant funds for interventions described in subparagraph (A).

"(d) EVALUATIONS.—

"(1) REQUIRED EVALUATIONS.--Any intervention without a high or moderate causal evidence rating used by a State in carrying out a State program or reemployment services and eligibility assessments under this section shall be under evaluation at the time of use.

"(2) FUNDING LIMITATION.--A State shall use not more than 10 percent of grant funds awarded to the State under this section to conduct or cause to be conducted evaluations of interventions used in carrying out a program under this section (including evaluations conducted pursuant to paragraph (1)."

ETA provided preliminary guidance with regard to these provisions for FY 2019 in UIPL No. 07-19 (https://wdr.doleta.gov/directives/corr_doc.cfm?DOCN=8397).

This guidance provides information on the new RESEA evidence-based requirements and provides definitions of high and moderate causal evidence. High or moderate causal intervention ratings are based on how many good quality studies show positive impacts of that intervention. To provide states a solid foundation on the meaning of good quality studies, this guidance presents a description of how DOL rates studies' quality of evidence through its Clearinghouse for Labor Evaluation and Research (CLEAR) (see Section 6,

below). The guidance also discusses the standards for rating intervention effectiveness and identifies relevant interventions that currently meet those standards.

In addition to setting standards and intervention ratings, this guidance also suggests RESEA components that are in need of expanded evidence and includes a discussion of evaluation approaches and strategies for carrying out evaluations. Finally, the guidance points to resources that are available to states to better understand and use existing evidence and to help states initiate rigorous high-quality evaluations to build evidence on the effectiveness of interventions in their RESEA programs.

While the intent is that states will implement interventions and service delivery strategies supported by rigorous evidence, there is not yet a large body of such evidence related to the new parameters for the permanent RESEA program. States must begin conducting rigorous studies to produce new evidence that helps determine the success of the interventions and service delivery strategies that meet the goals of the RESEA program.

4. Expectation that States Begin RESEA Evaluations No Later than FY 2020.

Reemployment evaluations, to date, have focused mainly on broad categories of services or services at a program level. These evaluated programs have similarities to RESEA, but also many differences. A primary goal of the RESEA legal requirement for evaluations is to expand the evidence base by conducting new high-quality evaluations of states' RESEA programs, particularly to build evidence about specific program components or activities.

Congress, as reflected in the provisions of section 306, SSA, intended the evidence base for RESEA to expand and to improve the program through state use of evidence-based interventions with high or moderate causal ratings. While there is a modest and growing 4 evidence base from which to synthesize and draw conclusions about RESEA interventions' effectiveness, there is an immediate need to grow and expand it to address new RESEA program components. Previous evaluations of the Reemployment and Eligibility Assessment (REA) program, the predecessor to the RESEA program, were based on the whole program and the need now is to develop and expand evidence on more well-defined activities, program components, and service delivery approaches that states use in operating the RESEA program. Development of a culture of continuous improvement and evidence building around the RESEA program will strengthen it over time and improve reemployment outcomes for unemployment compensation (UC) claimants.

To meet Congressional intent with regard to causal evidence ratings in the tiered evidence approach and to ensure states' ability to comply with the evidence and evaluation provisions in the statute, states are expected to begin evaluating RESEA interventions and service delivery strategies as soon as feasible and no later than the end of FY 2020 for the following reasons:

- The requirement that states use only interventions with high or moderate causal evidence ratings or have them under evaluation is in effect in FY 2020;
- RESEA, while modeled in part after the former REA, is a different program and includes the actual delivery of reemployment services in addition to the foundational elements of the REA program, so evidence beyond evaluations of the REA program is needed;
- Expanded evidence is needed to ensure that states have sufficient evidence to support program delivery when the minimum percentage requirements for use of interventions with high or moderate causal ratings begin in FY 2023;

- In the new RESEA state plan, required in FY 2020, states must articulate a description of their evaluation structure for RESEA interventions without a high or moderate causal evidence rating; and
- Rigorous impact evaluations sufficient to achieve a high or moderate causal rating are most often multi-year in length, and states need to begin conducting evaluations now to obtain sufficient evidence to support delivery strategies and interventions of their RESEA programs in FY 2023 and beyond.

5. Evaluation Parameters.

Given that a key goal of the RESEA program is reduced average duration of UC benefit receipt as a result of improved employment outcomes, states' RESEA impact evaluations must include duration of UC and employment as primary outcomes:

- Unemployment Compensation Duration: This outcome is measured as the number of weeks RESEA participants receive UC; and
- Employment: For RESEA participants, employment and earnings outcomes can be measured in the second full calendar quarter following the start of a participant's UC claim similar to the WIOA measures, or sooner in the claims cycle to the extent that data is available.

States are also encouraged to propose additional outcomes that could provide early indications that the RESEA program is working as intended. Examples of outcomes that states might consider include increased participation in or completion of the RESEA program activities, or time to reemployment following the start of RESEA interventions.

States should consider, when feasible, coordinating their RESEA evaluations with their WIOA-mandated evaluation projects which can create economies of scale and generate synergies across programs. States' new evaluations must meet evidence standards for study quality and find favorable impacts with at least a reasonable degree of statistical confidence to allow the intervention under examination to potentially qualify for a high or moderate rating, as defined in Section 8, below. The goal of this evidence-generating approach is to provide states operating RESEA programs with a sufficient number of new studies that meet these standards, which can support, along with current evidence, the statutory requirement for states to use interventions demonstrated to be effective.

DOL recognizes that all findings, whether positive, negative, or null, are important contributions to the evidence base, and DOL is committed to learning from and using evaluations and data to inform program improvements. As such, it is both critical and expected that all evaluations conducted of RESEA interventions be publicly available, regardless of the outcomes. States are also encouraged to share links to their publicly posted, completed evaluations with CLEAR to ensure their inclusion in future evidence reviews.

RESEA evaluations will play an important role in building the reemployment evidence base, and in helping states and other program decision-makers make more informed choices about how to bundle RESEA program components and strategies to best meet the needs of the people being served by them.

6. Clearinghouse for Labor and Evaluation Research (CLEAR).

A first step in identifying interventions with high and moderate causal ratings is determining which existing studies provide evidence about them that is relevant and credible. DOL will

leverage CLEAR to identify evaluations in the evidence base that are relevant to the RESEA program and determine which impact studies have high, moderate, or low causal evidence ratings. DOL established CLEAR to make research on labor topics more accessible to practitioners, policymakers, researchers, and the public so that evidence can inform policy and program decisions. To achieve this goal, CLEAR conducts systematic evidence reviews of research and evaluation reports on labor topics, and then reviews and summarizes those studies. CLEAR also rates studies that estimate causal impact.

CLEAR currently has over 600 studies summarized across 18 labor-related topic areas, including "Reemployment," and is continually growing. The "Reemployment" evidence review identifies, summarizes, and determines the quality of existing causal evidence on reemployment service delivery strategies intended to promote reemployment of UC claimants while also reducing UC receipt duration. Under the "Reemployment" topic area, CLEAR has reviewed 45 publications published between 1978 and 2018, and has developed one-page summary profiles of and ratings for each of these studies. The reviewed studies use "causal" designs, otherwise known as "impact studies," and assess the effectiveness or impact of an intervention. These studies identify how a particular intervention changes 6 claimants' outcomes relative to a comparison group, such as those that receive a different intervention or those that did not receive the intervention.

Many impact studies use random assignment designs. Such designs randomly (i.e., through the functional equivalent of a coin toss) assign some eligible individuals to a "treatment group" or groups that may participate in the intervention and others to a "control group" that do not participate in the intervention. These designs use random assignment to prevent systematic pre-existing differences between the two groups from creating bias in an evaluation. Thus, systematic differences in outcomes between the two groups can reasonably be attributed to the intervention. Other causal impact studies may use "quasi-experimental" designs that estimate impact, but do not use random assignment. Instead, quasi-experimental designs use administrative data and statistical techniques to identify a comparison group that is similar to the treatment group to act as a control group.

The credibility of the evidence from an impact study depends on how it is designed and carried out. Currently, CLEAR has established causal evidence guidelines, [49] which identify the criteria CLEAR uses to assess the strength of a study's causal evidence. CLEAR's causal evidence ratings are an indicator of the quality of the study and the level of confidence you can have that the study's findings truly reflect the causal impact of the intervention studied and not some other factor. CLEAR also has guidelines for high-quality, quantitative, descriptive, and implementation studies, but does not currently assign evidence ratings to those types of studies. [50]

CLEAR currently assesses its causal evidence ratings based on the rigor of the study as follows.

- Studies receive a "high" rating for causal evidence if there is confidence that the study's estimated effects are "solely attributable to the intervention being examined."
- Studies receive a "moderate" rating for study quality if there is "some confidence that the estimated effects are attributable to the intervention studied, but there might be other contributing factors that were not included in the analysis."

---

[49] Find CLEAR's causal evidence guidelines here: https://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_V2.1.pdf

[50] Find CLEAR's quantitative descriptive guidelines and guidelines for reviewing implementation studies here, under "Reference Documents": https://clear.dol.gov/about

Studies that do not meet the criteria for a high or moderate rating receive a "low" rating, which indicates that it is "not possible to be confident" that the estimated effects are attributable to the intervention studied. In these instances, other factors likely contributed to the estimated effects. Moving forward and as described below, CLEAR will assign causal evidence ratings to new RESEA studies based on both study quality and effectiveness of the intervention examined in a study, as appropriate. DOL will publicly and transparently post information about this process on the CLEAR website when future evidence reviews begin.

7. <u>The Need for Expanded Evaluations of Interventions.</u>

The statute, in section 306(i)(3), SSA, defines an intervention as "a service delivery strategy for the provision of State reemployment services and eligibility assessment activities." In operating RESEA programs, states bundle or mix multiple components and activities together in ways that may vary in their details or emphasis. They may also vary in the strategies or approaches for carrying them out. For instance, all RESEA programs include a claimant selection component, but states may vary in how they select claimants for participation. For evidence rating purposes, an evaluation intervention may be a whole program or any component of it.

RESEA is different from any of the interventions for which evidence currently exists. It shares some elements with earlier programs, particularly with REA, but it also has new elements. For example, it places a greater emphasis on more intensive reemployment services for claimants and states now have greater freedom in deciding how to select claimants. States need to develop a substantial body of high-quality evidence about the effectiveness of RESEA strategies and components. Exhibit 1 lists components for which evidence needs to be built, in order to meet the basic requirement of demonstrating effectiveness and to provide meaningful findings to help states design and implement their RESEA programs. Other gaps in the evidence base are expected to emerge as more is learned about states' current RESEA programs. As new evidence is produced, the list of interventions that have been demonstrated to be effective will be refined.

8. <u>Criteria for Causal Evidence Ratings.</u>

Section 306(c)(2), SSA, conditions funding for RESEA programs on states using interventions either demonstrated as effective with a "high or moderate causal evidence rating" or being under evaluation. Beginning in FY 2020, the definitions established below will be in effect and explain how an intervention can qualify for a high rating or a moderate rating.[51] These ratings examine available evidence and determine whether the interventions have favorable impacts on both employment and benefit duration outcomes. [52] The high and moderate causal evidence standards described below rely on evidence of impact exclusively from studies that received a high or moderate rating for study quality in CLEAR. These studies are identified in the definitions as credible studies.

- ***High:*** For an intervention to qualify for a high causal evidence rating, there must be at least two credible impact studies of the intervention (as reviewed by CLEAR) that have each

[51] As the evidence base grows, more information will be available to help distinguish which approaches have the strongest evidence of effectiveness. At that time, the standards for evidence of effectiveness may evolve as well, in order to help better support those distinctions.

[52] Specifically, the ratings criteria are based on interventions' estimated impacts on (1) reduced UC duration and (2) increases either employment rates or earnings, as measured in the second full calendar quarter after the claim began.

found favorable impacts on employment and UC duration, with a strong degree of statistical confidence. [53]

- ***Moderate:*** For an intervention to qualify for a moderate causal evidence rating, there must be at least one credible impact study of the intervention that found a favorable impact on employment and one credible impact study of the intervention that found a favorable impact on UC duration. Again, these ratings of the study or studies are as reviewed by CLEAR. Each study must have at least a modest degree of statistical confidence.[54] The findings on employment and benefit duration may both come from the same study or from different studies.

DOL also defines two additional categories: "potentially promising" and "no rating."

- ***Potentially Promising:*** A potentially promising rating indicates that there is some suggestive evidence that an intervention may be effective. Such interventions are candidates for further evaluation that possibly would allow the intervention to qualify for a higher rating. For an intervention to qualify for a potentially promising causal evidence rating, there must be one impact study reviewed by CLEAR (irrespective of the causal evidence rating it received)[55] that has found significant favorable impacts on either employment or UC duration with at least a moderate degree of statistical confidence.[56]
- ***No Rating:*** All interventions that do not qualify for any of the three ratings above will receive no rating, regardless of the rating given by CLEAR for the quality of studies of that intervention. These may be interventions for which no impact studies have been conducted, interventions with an impact study that have not been reviewed by CLEAR yet, or interventions whose studies have been reviewed by CLEAR but have not shown any favorable impacts.

---

[53] A "strong statistical confidence" is defined as an estimated impact that is "statistically significant" (different from zero) at the 5% level. (p

[54] A "modest degree of statistical confidence" is defined as an estimated impact that is "statistically significant" (different from zero) at the 10% level. (p<.10). This means that there is less than a 10% chance that the study's results are due to chance and not actually the intervention. Impact estimates must meet that threshold for both outcomes—UC duration and employment.

[55] CLEAR also rates some studies as low. These studies are not used when considering whether an intervention is eligible for a high or moderate effectiveness rating. However, studies rated as low can contribute to a potentially promising rating. The potentially promising rating indicates that some suggestive evidence exists that an intervention might be effective. While evidence from a low rated study is not a strong basis for concluding that an intervention is effective, it can suggest that the intervention may be worth considering for more rigorous testing. Some studies that are rated as low may still be considered promising and thus, a candidate for further evaluation.

[56] As noted earlier, for moderate effectiveness ratings, a "modest degree of statistical confidence" is defined as an estimated impact that is "statistically significant" (different from zero) at the 10% level.

9. High and Moderate Causal Ratings for Existing and Future Interventions Beginning in FY 2020.

Existing impact studies of approaches to speed the reemployment of UC claimants typically focus on broadly defined sets of services and activities. CLEAR's 2018 ***Reemployment Research Synthesis*** on reemployment interventions[57] identified the following broad intervention categories from the existing evidence base that are relevant to RESEA: [58]

- Reemployment and Eligibility Assistance (REA): The REA program, the predecessor to RESEA, provided claimants up to three mandatory in-person sessions in which workforce staff assessed their continued eligibility for UC, provided them with labor market information, and supported their development of a reemployment plan. In some cases, they also provided referrals to reemployment services. Failure to attend REA sessions without good cause affects continued receipt of UC.
- Job Search Assistance (JSA): JSA interventions provide claimants assistance and training in job search techniques, including job search workshops, preparing a resume, and interview training. The JSA interventions that included strong linkages between UI and workforce partners and required claimants at risk for benefit exhaustion to report for job search assistance demonstrated positive impacts.
- Profiling: Profiling interventions identify claimants at higher risk of exhausting UC and offer or require enhanced employment services. These services may include an orientation, providing labor market information, and referrals to job search training or resume training workshops. Claimants that fail to participate in required services without good cause lose UC.
- More Stringent Employer Contact Requirements: This type of intervention increases the amount of work search effort required of claimants to continue receiving UC, strengthen verification of work search efforts, or both.
- Less Stringent Employer Contact Requirements: This type of intervention reduces the amount of work search effort required of claimants to continue receiving UC, loosen verification of work search efforts, or both.

These broadly defined interventions often involve partially overlapping services and activities. Beginning in FY 2020, these interventions will receive effectiveness ratings using the definitions above. To the extent that the states' programs use interventions that have not received high or moderate evidence ratings, those states must be conducting high-quality impact evaluations using the CLEAR guidelines for study quality.

Interventions Receiving a High Rating

Of the interventions considered, only REA currently receives a high causal evidence rating. If a state's RESEA program has components that are sufficiently similar to the evaluated REA program components, a state can demonstrate that those pieces or components of its RESEA program are evidence-based, by referring to this intervention and its rating. While no evaluation

---

[57] Find CLEAR's Reemployment Research Synthesis here: https://clear.dol.gov/sites/default/files/CLEAR%20Reemployment%20Synthesis%20November%202018.pdf

[58] All descriptions are adapted from CLEAR's 2018 research synthesis, What do we know about the effect of reemployment initiatives?, which can be found at: https://clear.dol.gov/sites/default/files/ResearchSynthesis_Reemployment.pdf

of REA components that are part of a state's RESEA program is required, states are encouraged to continue to evaluate these interventions in order to build rigorous evidence in the context of the new RESEA program. As noted above, RESEA is not identical to the REA program and has different components, so REA interventions alone will not be sufficient to meet RESEA requirements, and states should continue to consider implementing new interventions. If a state's RESEA program includes other components that are not evidence-based, those components must be under evaluation at the time of use.

Interventions Receiving a Moderate

Applying the criteria above, JSA and profiling interventions receive a moderate causal evidence rating. If a state's RESEA program includes components like the JSA and profiling strategies described above in this guidance, a state can demonstrate the corresponding components of its RESEA program are evidence-based by referring to these components and interventions and their ratings. Again, while no evaluation of JSA components of a state's RESEA program are required, states are encouraged to continue to evaluate their interventions in order to continue to grow the base of evidence regarding their use in the RESEA program. Additional evidence on these interventions is still valuable and could result in raising the causal evidence rating for the interventions to the high category. If a state's RESEA program includes other components that are not evidence-based, those components must be under evaluation at the time of use.

Interventions Receiving a Potentially Promising Rating

Applying the criteria above, the component of requiring **more stringent employer contacts** receives a potentially promising causal evidence rating. This rating indicates that the component or intervention may be of interest to consider adopting or testing, as it might be effective. Additional evidence on these interventions might also support a change of causal evidence rating for the intervention. States implementing interventions with only a potentially promising rating must be evaluated at the time of use.

Interventions Receiving No Rating

Applying the criteria above, the component of **less stringent employer contacts** receives no rating. Additionally, the more detailed components included in Exhibit 1, as well as any additional RESEA interventions or program components not identified here as being demonstrated effective by current evidence, also currently receive no rating. Such interventions must be evaluated if states choose to implement them.

10. Evaluation Approaches.

Section 306(c), SSA, gives states time to evaluate RESEA interventions before the percentage requirements for use of interventions with high or moderate causal ratings begin in FY 2023. States' evaluations will need to meet the causal evidence standards described in Section 8 of this guidance. To help states' impact evaluations have the best chance of meeting CLEAR's standards, states are strongly encouraged to: (1) choose an experienced evaluator; (2) choose a simple impact study design (the simplest being random assignment with administrative data follow-up); and (3) take advantage of the evaluation technical assistance (EvalTA) guidance being provided (described in more detail in Section 13, below).

Importantly, there are multiple types of evaluations and evaluation-related activities that ultimately support a strong impact evaluation that produces high or moderate causal evidence. For example, it may be appropriate to conduct an evaluability assessment (discussed in more

detail below) or feasibility study before embarking on an impact evaluation to identify any challenges or barriers, such as data availability or limited sample size, to conducting an evaluation of a specific intervention or service delivery design. Alternatively, it may be desirable to pair both an implementation evaluation and an impact evaluation. DOL considers activities leading up to an impact evaluation that has the capability of producing a high or moderate causal rating to be interventions designated as "under evaluation." Examples of these activities include an evaluability assessment as described below, and an implementation study that helps refine the specific intervention and research questions to be addressed in the impact evaluation.

Evaluation Design

DOL encourages states' evaluation designs to specify use of, or be building evidence to move toward the use of, approaches capable of earning high or moderate quality ratings, with 13 the goal of producing both the strongest possible evidence and the highest possible causal evidence rating for the intervention being studied. As noted previously, impact evaluations are necessary to achieve these ratings; however, there may be other types of evaluations or pre-evaluation activities that should be conducted first or along with an impact study to maximize learning about the intervention.

The most common types of evaluation designs are:

- Impact Evaluation: This type of evaluation assesses the impact of a program or component of a program on outcomes, typically relative to a counterfactual situation. This evaluation provides some estimate of what would have happened in absence of the program or component of the program. Impact evaluation includes both experimental (i.e., randomized controlled trials) and quasi-experimental designs. These types of evaluations speak to the "does it work?" question.
- Outcome Evaluation: This type of evaluation measures the extent to which a program has achieved its intended outcome(s), and focuses on outputs and outcomes to assess effectiveness. Unlike an impact evaluation, an outcome evaluation cannot show causal impacts. An outcome evaluation can help answer questions like, "Did the program, policy, or organization do what it intended to do?"
- Process or Implementation Evaluation: These types of evaluations assess how the program or service is delivered relative to its intended theory of change, and often include information on content, quantity, quality, and structure of services provided. Process or implementation evaluations can be conducted on their own, but are often conducted along with impact and/or outcomes evaluations. Process or implementation evaluations can help answer questions like, "Was the program or policy implemented as intended?" or "How is the program, policy, or organization operating in practice?"
- Formative Evaluation: This type of evaluation, typically done before full implementation of a program, assesses whether a program or component of a program is feasible, appropriate, and acceptable before it is fully implemented. It may include some of the activities described above, such as process evaluation or outcome evaluation. However, unlike summative evaluation designs like impact and outcome evaluations, which seek to answer whether or not the program met its intended goal(s) or had the intended impact(s), a formative evaluation focuses solely on learning and improvement and does not answer questions of overall effectiveness.

Selecting an Evaluator

While DOL recognizes there is value in all types of evaluations, the RESEA evidence-generating approach specifically requires impact evaluations of interventions to help determine causal evidence ratings for those interventions. While states may have evaluation capacity within the agency operating the RESEA program, DOL recommends that states use an experienced and independent evaluator that can identify the most appropriate and rigorous 14 design to answer research questions and learn about the RESEA program and program components and interventions.

Deciding What to Evaluate

DOL recognizes that each state's RESEA program is a uniquely implemented bundle of different interventions and service delivery strategies, or components. However, this can make it difficult to know which program components, and how these components, are generating the observed outcomes of the intervention. Therefore, states are strongly encouraged to work with their independent evaluators to develop evaluations that seek to estimate the impact of individual RESEA program components and interventions, or to develop evaluations of whole programs where the components are well defined and the effectiveness of which could be evaluated at a later time through meta-analyses. Building this type of evidence will further states' understanding of the effectiveness of components that could be mixed and matched to develop a program that meets the needs of a specific state. However, DOL recognizes that evaluating only a component of the program implies the need for the evaluation to detect smaller impacts, which requires much larger samples. Again, states are encouraged to work with experienced evaluators and explore partnerships with other states to develop the most rigorous and appropriate designs to determine the effectiveness of program components. See Section 11, below, for more discussion on evaluation partnerships across states.

Pre-evaluation activities states can begin doing now to support getting to a firm evaluation plan include the following activities:

- States are encouraged to conduct **evaluability assessments** of their RESEA programs. Evaluability assessments will help states: define specific interventions that are evaluation-ready to test in a feasible, measurable way; ensure that the intervention and the component(s) to test are well-understood by all stakeholders; confirm availability of data and other operational resources needed to conduct an evaluation; and build consensus on evaluation goals to ensure results are relevant to stakeholders. Evaluability assessments also are useful for identifying a program's potential strengths and challenges with planning and executing an evaluation. For example, they may assess whether adequate access to information technology (IT) and data resources exist and are available to support the evaluation or if program staff has sufficient evaluation expertise. The results of an evaluability assessment refine a state's broad learning goals with more narrowly-focused research questions that explore the RESEA program's influence on a particular population's outcomes of interest, and identify the type of evaluation that can best answer these questions. Evaluability assessments highlight evaluation feasibility issues, such as operational gaps that must be addressed to successfully execute the evaluation, such as availability of IT resources and data availability, staffing to increase evaluation capacity, developing partnerships with organizations that have appropriate evaluation expertise, and creating evaluation procedures and training staff. Addressing these operational gaps identified through an evaluability assessment strengthens a state's ability to produce a high-quality evaluation that meets CLEAR standards. States can 15 find an Evaluation

Design Assessment Tool developed by IMPAQ International to support WIOA evaluations here: https://evalhub.workforcegps.org/resources/2018/09/07/19/53/Evaluation-DesignAssessment-Tool

- States are also encouraged to develop **logic models** when formulating evaluation plans. Logic models are graphical representations of interventions and how they operate. They are designed to show the following regarding a RESEA intervention:
  - Inputs – such as staff time, RESEA funds, and other resources used to deliver the program;
  - Activities – such as meeting with American Job Center (AJC) staff to create an individual reemployment plan, provision of reemployment services, conducting the eligibility assessment, and other activities the program regularly operates;
  - Outputs (the immediate results of the program) – such as improved job readiness skills or enhancing labor market knowledge; and
  - Outcomes (the expected short-term and long-term goals of the program) – such as reduced UC duration, faster return to employment, and improved earnings.

  Logic models define the inputs, activities, or other tangible activities that lead to outputs and outcomes for the RESEA program. Logic models and other similar program mapping activities demonstrate how RESEA interventions drive the change in outcomes for claimants. States can find more information about developing logic models for labor programs in a webinar titled *Fully Articulating Your Vision: Using Logic Models to Support Innovation:* https://evalhub.workforcegps.org/sitecore/content/global/resources/2015/05/07/11/07/ Fully_Articulating_Your_Vision_Using_Logic_Models_to_Support_Innovation

- States' evaluability assessments and logic models ultimately help states identify specific research questions that may be added to a **multi-year learning strategy or agenda**. Organizing learning priorities is an approach that is gaining traction across the Federal government, most recently supported in the Foundations for Evidence-Based Policymaking Act of 2018 (Public Law No. 115-435),[59] which requires Federal government agencies to produce evidence-building plans. Learning agendas can also serve as roadmaps to help states plan for immediate and future evaluations, by clarifying learning goals, research questions, the types of evaluations that would answer those questions, and the states' priorities in building evidence

11. <u>Strategies to Meet RESEA Evaluation Requirements.</u>

Some states may be interested in conducting their own individual impact evaluations. As indicated previously in UIPL No. 7-19, DOL encourages states to consider evaluation partnerships with other states, so that states may consider conducting pooled evaluations of similar RESEA interventions. This approach has the benefit of potentially yielding sample sizes large enough to demonstrate effectiveness. Smaller states, in particular, might benefit from this strategy. It also has the benefit of allowing states to pool their limited evaluation funding.

From previous research, we know that sample sizes required to detect impacts on labor market outcomes of the kind required by section 306, SSA, are large. DOL recognizes that many states do not have a sufficiently large number of RESEA-eligible claimants in a single year, and some states

[59] https://www.congress.gov/bill/115th-congress/house-bill/4174

do not have that many RESEA-eligible claimants in several years. A pooling strategy, including a well-defined intervention aligned across states, can help overcome this challenge. States can pool their data and yield samples large enough to detect effects and potentially demonstrate effectiveness on a more reasonable timeframe. As mentioned previously, states are encouraged to work with experienced evaluators and explore partnerships with other states to develop the most rigorous and appropriate study designs to evaluate program components and interventions.

12. Evaluation Resources.

In addition to the tools linked to above, DOL's CLEAR also has several additional tools to help states better understand the evidence in the Reemployment topic area.

- CLEAR's Reemployment Synthesis is a short, high-level, plain-language report that summarizes studies of interventions that are relevant to RESEA. It describes key takeaways from the reemployment evidence base, provides an overview of the interventions studied, and identifies gaps in the research. This report may be useful as RESEA program managers begin to focus on conducting evaluability assessments and efforts to build the evidence base.
- As a companion to the Synthesis, CLEAR's Reemployment Synthesis Supplement gives states a more detailed look at the information provided in the synthesis. It provides brief descriptions of the findings for all the reports reviewed in the Reemployment topic area. This supplemental tool also includes links that lead directly to the study profiles in CLEAR where more information about the specific studies and interventions is available. It also is organized by sections that correspond to the intervention categories identified in the Reemployment Synthesis that will receive ratings beginning in FY 2020, as described above.

CLEAR and its resources are at the links below:

- CLEAR: https://clear.dol.gov/
- Reemployment topic area: https://clear.dol.gov/topic-area/reemployment
- Reemployment Synthesis landing page: https://clear.dol.gov/synthesisreport/reemployment-synthesis, and a downloadable brief report: https://clear.dol.gov/sites/default/files/ResearchSynthesis_Reemployment.pdf.
- Reemployment Synthesis Supplement: https://clear.dol.gov/sites/default/files/ResearchSynthesis_Reemploy_Sup.pdf. 17

Another important resource for states will be DOL's ***RESEA Evidence Building and Implementation Study.*** In September 2018, DOL's Chief Evaluation Office awarded a three-year contract to Abt Associates, the Urban Institute, Capitol Research Corporation, and National Association of State Workforce Agencies ("the RESEA study team") to provide support on implementing the evaluation requirements in section 306, SSA. Among the tasks planned for the RESEA study team is an implementation study of states' RESEA programs. DOL is conducting this implementation study to examine how RESEA programs and strategies are operated, understand how states are bundling various services to improve outcomes for RESEA participants, and identify new, innovative, and potentially promising strategies being implemented. Findings from this implementation evaluation will inform an evaluation report that will further describe research and evaluation options for DOL and states to consider and will contribute to the RESEA evidence base. Reports from the study will be publicly available when completed. A brief description of DOL's RESEA study on the Chief Evaluation Office's

website is available here: https://www.dol.gov/asp/evaluation/currentstudies/Reemployment-Services-and-EligibilityAssessments-Research.htm.

Finally, it is DOL's intent that states also leverage other available evaluation capacity-building resources. These include, but are not limited to:

- Evaluation and Research Hub: A new community of practice, created with input from state and local workforce agency representatives across the country. While it is available to address the evaluation requirements of the WIOA, the resources included on the Hub can inform or support the evaluation needs of all ETA-funded programs. You can find it here: https://evalhub.workforcegps.org/about
- WIOA Evaluation Technical Assistance Tools: State and local workforce agencies participated in ETA's peer learning effort to share and disseminate evaluation resources as well as address questions such as, "Where and how do we start?" Key tools are included here:
  - Evaluation Readiness Assessment Tool: https://evalhub.workforcegps.org/resources/2018/09/07/19/45/EvaluationReadiness-Assessment-Tool
  - Evaluation Design Assessment Tool: https://evalhub.workforcegps.org/resources/2018/09/07/19/53/EvaluationDesign-Assessment-Tool
  - Evidence Says: Work-based Learning: https://evalhub.workforcegps.org/resources/2018/09/07/16/10/The-EvidenceSays-Work-Based-Learning
- Workforce System Strategies (WSS): A research clearinghouse that profiles evidence-based and emerging practices in workforce development to help the field make informed decisions about improving outcomes for job seekers and employers. Its resource library contains more than 1,200 profiles of evaluation reports, policy and practice briefs, and how-to guides. It is available here: https://strategies.workforcegps.org/announcements/2018/05/04/20/17/Connect-YourPeers-to-Workforce-System-Strategies.

13. Evaluation Technical Assistance (EvalTA)

As described above, a critical piece of the DOL's RESEA project is to provide EvalTA to states. The EvalTA will include a suite of tools and resources to help states meet evaluation and evidence-building needs for their RESEA programs. Experienced staff from the RESEA study team will develop and deliver EvalTA.

Beginning in summer 2019, the RESEA study team is offering generalized EvalTA, which has been informed by state feedback from webinars, clarifying calls, and a review of available documents on states' FY 2019 RESEA programs. The goals of this generalized EvalTA is to help states with the following: 1) gradually and continually increase their evaluation capacity so states are prepared to begin evaluation-related work by FY 2020; 2) to describe evaluation activities in their FY 2020 RESEA state plans; and 3) to meet evidence-related statutory requirements, both now and moving forward. The EvalTA will be provided through resources DOL will make broadly available, such as webinars, toolkits, briefs, templates, and videos, to explain key topics to improve states' understanding of basic evaluation concepts and begin to

plan and carry out evaluations. These resources will build on existing DOL evaluation technical assistance resources (e.g., for WIOA, as described above) and focus particularly on knowledge required for evaluations of RESEA program components and interventions that can meet evidence standards. When they are finalized, a schedule of EvalTA activities, as well as all resources developed through the EvalTA, will be available on the Reemployment Connections community of practice on WorkforceGPS (www.workforcegps.org).

In addition to this generalized EvalTA, the RESEA team will also offer customized EvalTA to individual states or small groups of states, as appropriate, that are planning or carrying out evaluations. Customized EvalTA is anticipated to begin in fall 2019 and is likely to include detailed verbal and written technical assistance to states at key points during individual evaluations. The RESEA study team may provide customized EvalTA in areas such as procurement and selection of an independent evaluator; selection of methods and development of evaluation design plans; monitoring random assignment and dealing with unanticipated issues; methods of analysis and reviews of analysis plans; reporting and dissemination; and other issues as appropriate and needed. Plans for customized EvalTA will be updated by DOL's RESEA project yearly, as states' needs are better understood and as new RESEA interventions and evaluations are planned.

While not every state is expected to need or participate in customized EvalTA, all states are encouraged to take advantage of the generalized EvalTA being offered. Previous experience from other tiered evidence initiatives across the government suggest that an adequate planning period combined with a comprehensive EvalTA strategy can improve evaluation quality. Together, these efforts can help meet one of the primary goals of this early phase of RESEA implementation, to expand the evidence base by supporting states in conducting new high-quality evaluations of RESEA program components and interventions.

14. Inquiries. For further information, please direct inquiries to the appropriate ETA Regional Office.

15. References.

- Section 306, Social Security Act, 42 U.S.C. § 506;
- The Bipartisan Budget Act of 2018, Public Law No. 115-123;
- The Foundations for Evidence-Based Policymaking Act of 2018, Public Law No. 115-435; and
- Unemployment Insurance Program Letter No. 07-19, *Fiscal Year (FY) 2019 Funding Allotments and Operating Guidance for Unemployment Insurance (UI) Reemployment Services and Eligibility Assessment (RESEA) Grants*, issued January 11, 2019.

Appendix C. Reemployment Services and Eligibility Assessments (RESEA) Evaluation Resource List

As a part of its evaluation technical assistance (EvalTA) efforts, the RESEA study team has compiled a list of key websites, studies, and tools that will help RESEA program administrators and their staff understand, use, and produce evidence about their RESEA interventions. This list will be updated as new resources are available.

The Clearinghouse for Labor Evaluation and Research (CLEAR)

CLEAR reviews studies in a variety of labor-related topic areas and can be found at https://clear.dol.gov. CLEAR categorizes studies into topic areas. Below are particularly relevant resources:

- **CLEAR Causal Evidence Guidelines***:* This document describes CLEAR's guidelines for rating the strength of causal evidence presented in causal studies. https://clear.dol.gov/sites/default/files/CLEAR_EvidenceGuidelines_V2.1.pdf
- **CLEAR Guidelines for Reviewing Implementation Studies**: This document describes the Clearinghouse for Labor Evaluation and Research guidelines for reviewing the technical qualities of and findings from implementation studies. https://clear.dol.gov/reference-documents/clearinghouse-labor-evaluation-and-research-clear-guidelines-reviewing-0
- **RESEA Topic Area Tab:** This page lists RESEA-related evidence resources, including RESEA intervention ratings. https://clear.dol.gov/reemployment-services-and-eligibility-assessments-resea
- **Reemployment Topic Area**: This topic area focuses on studies of interventions designed to promote faster reemployment of unemployment insurance (UI) claimants. Of particular note are REA impact evaluations. https://clear.dol.gov/topic-area/reemployment
- **Job Search Assistance Topic Area:** This topic area includes reviews of studies examining the effectiveness of job search assistance interventions which aim to improve participants' employment and earnings outcomes. https://clear.dol.gov/taxonomy/term/522

WorkforceGPS – Evaluation and Research Hub

WorkforceGPS is an online repository of technical assistance resources for the public workforce system. The Evaluation and Research Hub (https://evalhub.workforcegps.org) provides workforce staff with evaluation resources and peer learning activities to help them integrate evaluation into program operations. The following resources can be found on the Evaluation and Research Hub:

- **Workforce Innovation and Opportunity Act (WIOA) Evaluation Toolkit**: The WIOA Evaluation Toolkit is an extensive general guide on how to conduct workforce evaluations. This document describes key evaluation fundamentals such as types of evaluations and their purposes, things to consider in selecting an evaluator, and activities involved with evaluation implementation. https://evalhub.workforcegps.org/resources/2018/09/07/19/58/WIOA-Evaluation-Toolkit
- **RESEA Evaluation and Evidence Resources:** This page provides links to RESEA guidance and EvalTA products. https://rc.workforcegps.org/resources/2019/07/30/17/32/RESEA_Evaluation_Evidence_Resources

- **RESEA Evaluation Technical Assistance Webinars:** This webinar series introduces states to key evaluation concepts and prepares them to evaluate their RESEA interventions.
  - ***"How Does Evaluation Help My State and Where Do We Start?":*** This webinar describes initial steps states can take to plan evaluations of their RESEA interventions. https://www.workforcegps.org/events/2019/04/17/16/43/Evaluating-RESEA-How-Does-it-Help-My-State-and-Where-Do-We-Start
  - ***"Which Evaluation Designs Are Right for My State?":*** This webinar introduces states to a range of evaluation design types and describes what kinds of information can be learned from each type. https://www.workforcegps.org/events/2019/05/07/13/07/What-Evaluation-Designs-Are-Right-For-My-State
  - ***"What Evaluation Details Do I Need to Plan For and How Long Will It Take?":*** This webinar prepares states to plan evaluation activities from start to finish. Attendees will explore common activities at each stage of the evaluation. https://www.workforcegps.org/events/2019/05/30/13/54/What-Evaluation-Details-Do-I-Need-for-a-Plan-and-How-Long-Will-It-Take
  - ***"Procuring and Selecting an Evaluator":*** This webinar provides an overview of procurement activities and elements to support states in selecting qualified, independent evaluators to conduct high quality evaluations. https://www.workforcegps.org/events/2019/05/30/14/24/Procuring-and-Selecting-an-Independent-Evaluator
  - ***"Assessing Data for Your RESEA Evaluation":*** This session prepares attendees to discuss and assess options to use administrative data to support evaluations. https://www.workforcegps.org/events/2019/07/23/13/26/Assessing-Data-and-De-Mystifying-Random-Assignment-for-your-RESEA-Evaluation-Office-Hours-Session-1
  - ***"De-Mystifying Random Assignment Designs for RESEA":*** This session aims to increase states' familiarity and confidence with random assignment evaluation designs. https://www.workforcegps.org/events/2019/07/23/13/26/Assessing-Data-and-De-Mystifying-Random-Assignment-for-your-RESEA-Evaluation-Office-Hours-Session-1
  - ***"Using the Clearinghouse for Labor Evaluation and Research (CLEAR) - A Demonstration":*** This session includes a demonstration of the CLEAR website functionalities and explains how CLEAR may be used to find evidence relevant to states' RESEA programs. https://www.workforcegps.org/events/2019/06/05/13/28/Using-the-Clearinghouse-for-Labor-Evaluation-and-Research-CLEAR-A-Demonstration
- **Online Training:** WorkforceGPS provides a range of self-paced training modules designed to improve the public workforce systems' capacity to serve customers and improve program performance. To access trainings, visit: https://www.workforcegps.org/online-training.
- **Tools**: Among the Evaluation and Research Hub's evaluation planning tools, we recommend:
  - ***Evaluation Readiness Assessment Tool:*** This tool will help states understand their readiness to conduct rigorous evaluations. https://evalhub.workforcegps.org/resources/2018/09/07/19/45/Evaluation-Readiness-Assessment-Tool
  - ***Evaluation Design Assessment Tool:*** This tool will help states assess the evaluability of their proposed intervention and understand operational considerations. https://evalhub.workforcegps.org/resources/2018/09/07/19/53/Evaluation-Design-Assessment-Tool

DOL Chief Evaluation Office

DOL's Chief Evaluation Office (DOL CEO) manages, and implements the Department of Labor's (DOL's) evaluation program (https://www.dol.gov/asp/evaluation/). In its work, DOL CEO has drafted guiding policies and compiled resources on a range of evaluation issues, including:

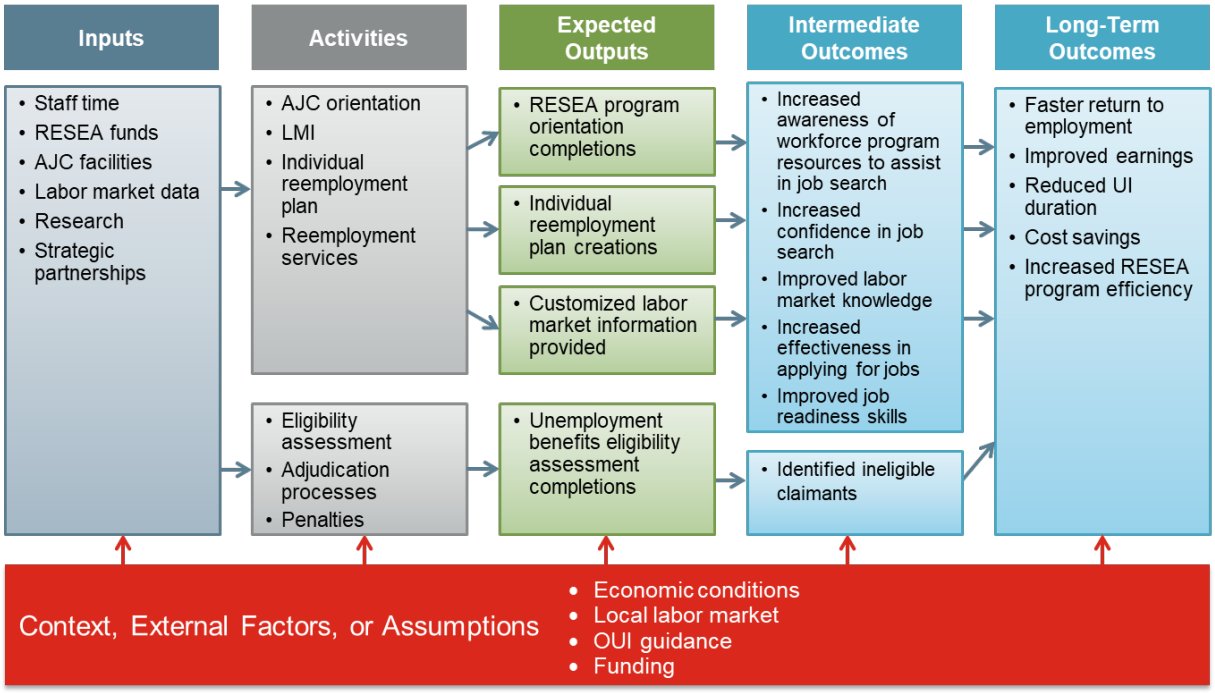
- **Labor Evaluation Policy:** This evaluation policy statement presents key principles that govern DOL's planning, conduct, and use of program evaluations. https://www.dol.gov/asp/evaluation/evaluationpolicy.htm
- **Public Use Data:** "Public use data" are files prepared by data suppliers with the intent of making them open to the public. On this site, states can find data files compiled through previous evaluations, surveys, programs, and more. https://www.dol.gov/asp/evaluation/PublicUseData.htm
- **Completed Reports:** This page contains completed reports from DOL-funded programs and evaluations. https://www.dol.gov/asp/evaluation/CompletedStudies.htm

Appendix D. Logic Model Template

Logic Model Example

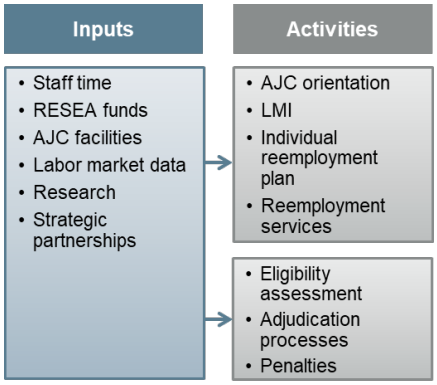
Below is a sample logic model—a graphical representation of a generic RESEA intervention. A logic model provides a clear visual representation of how what you put into your RESEA program (e.g., staff time, RESEA funds, and other resources) combined with the activities that are a part of your regular program operations (e.g., eligibility assessments, provision of labor market information, creation of individual service plans, and referrals to services) are hypothesized to lead to changes in claimants' intermediate (short-term) and long-term outcomes. The logic model is a useful tool in making decisions about the direction of the evaluation.

Exhibit D-1. A Sample RESEA Logic Model



Logic Model Elements

Together, a RESEA program's the inputs and activities serve as the ingredients for achieving the desired outcomes and describe the program in action.



Inputs consist of the resources that go into operating the program or intervention. For an RESEA intervention, inputs may include RESEA grant funds, staff time, and AJC resources.

Activities most commonly refers to the RESEA services that are delivered to claimants as part of the RESEA program or intervention in an effort to generate the desired outcomes. Major RESEA program components, as described in your RESEA application to DOL, belong in the activities category.

Outputs are products developed, deliverables completed, or milestones accomplished when claimants engage with RESEA program inputs and activities. In other words, program outputs occur when the inputs and activities accomplish the intended objectives.

Expected Outputs

- RESEA program orientation completions
- Individual reemployment plan creations
- Customized labor market information provided
- Unemployment benefits eligibility assessment completions

The logic model hypothesizes that these outputs lead to outcomes for the RESEA participants and for the RESEA program. For your logic model, you may find it helpful to include specific targets for each output. For instance, your program might aim to have 70% of claimants complete the RESEA program orientation and have 90% of claimants who attend an RESEA meeting create individualized reemployment plans.

Intermediate Outcomes

- Increased awareness of workforce program resources to assist in job search
- Increased confidence in job search
- Improved labor market knowledge
- Increased effectiveness in applying for jobs
- Improved job readiness skills

- Identified ineligible claimants

Intermediate outcomes are the expected program service delivery milestones or goals achieved that your logic model hypothesizes to lead to long-term participant long-term outcomes. Note that sometimes outputs and intermediate outcomes of a workforce program are the same. For example, the determination that an individual is ineligible for the intervention can be both an output and outcome.

Long-term outcomes are the changes in behavior, attitude, aptitude/skill, and knowledge for staff, participants, or workforce system that the logic model expects to result from the outputs and short-term outcomes. Most important for RESEA interventions, expected long-term outcomes include increases and improvements in employment, earnings, and reduction in weeks of receipt of UI benefits for the claimants receiving the RESEA interventions. All outcomes that the program is expected to affect should be included in the logic model.

Long-Term Outcomes

- Faster return to employment
- Improved earnings
- Reduced UI duration
- Cost savings
- Increased RESEA program efficiency

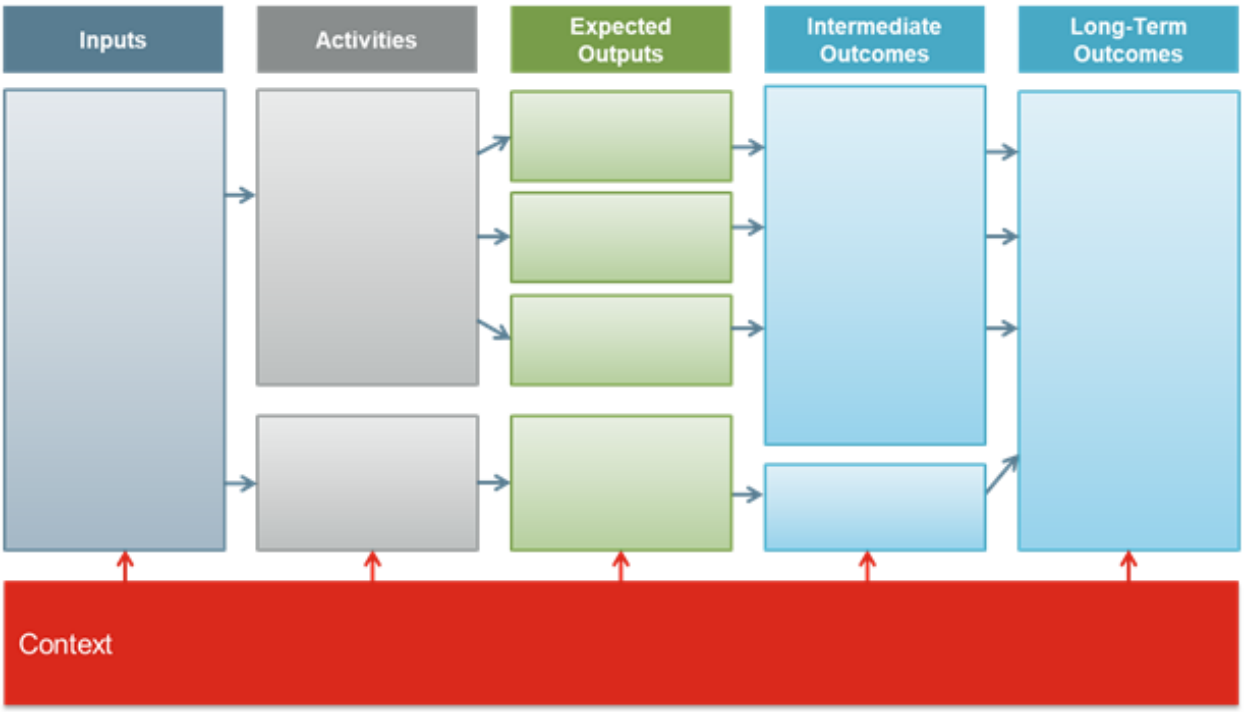
Context or Assumptions. A logic model may also include a space for assumptions made as well as any external factors that may bear on the intermediate- and long-term outcomes. The assumptions and external factors in a logic model provide context for the program's evaluation. In the sample logic model above, there are no assumptions outlined for the RESEA program, but there are some external factors listed: economic conditions, local labor market, Office of Unemployment Insurance guidance, and funding.

Relationships. Logic models illustrate the program's inputs linked to its specific activities, the activities linked to specific outputs, and the outputs linked to specific outcomes. Rather than just a long list of program components without logical paths, the arrows in a logic model show which inputs and activities are expected to affect which outputs and outcomes. A clear understanding on the part of the program and its evaluator of the expected effects allows for the creation of a more defined evaluation plan. When needed, logic model also provides an opportunity both to refine the study and re-articulate the expected results based on the reality of program operations.

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)



Logic Model Creation and Refinement: This graphical template will enable you to identify the essential elements. Once you develop an initial logic model, you should refine and assess its comprehensiveness, or modify it to meet your program's evaluation needs. The self-assessment questions in Exhibit G.2, allow you to further refine or modify components or elements of the logic model as you address the gaps identified in the responses to these questions.

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 that provides 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. The narrative also articulates the relationships between program elements and the intermediate- and long-term outcomes those elements are expected to affect.

Exhibit D-2. Self-Assessment Logic Model Questions

Self-Assessment of RESEA Program Logic Models: Discussion Questions
☐ Does the model include critical inputs required for the implementation of the service activities? (e.g., accessible technology and other resources, American Job Center facilities, program partner services, staffing, case managers, and partners with training services)
☐ Does the model include all of the current activities provided to RESEA claimants as a part of the intervention being tested? Is there an existing or expected sequence of claimant activities that follows a logical path or pattern?
☐ Does the model include all of the "immediate and short-term outputs" of the program? (For example, measureable milestones that are necessary but not sufficient conditions for achieving outcomes such as full participation in RESEA meetings and creation of Reemployment Plans.)
☐ Does the model include all of the hypothesized immediate changes and/or outcomes expected for participants, across all relevant domains?
☐ Are these immediate changes and/or outcomes an assumed result of specific services?
☐ Does the intervention's underlying theory of the change identify expected participant outcomes for particular RESEA services?
☐ Does the logic model suggest links between immediate, short-- and longer-term outcomes?
☐ Are the longer-term claimant outcomes likely to be measurable in the life of the evaluation?
☐ As a complete visual and narrative text, does the logic model tell a clear and complete story about your RESEA intervention's service strategy in the study?
☐ If the logic model assumes a theory of change, how does the hypothesis to lead to moderate and long-term outcomes?
☐ If using a visual representation, does supplementary narrative text provide a clear and complete story?
☐ Are there assumptions about external conditions or other external factors that could affect the successful implementation of the RESEA intervention?
☐ Are these identified external conditions or other external factors shown in the model?

Source: Modified from the Evaluation Toolkit for Prospective Workforce Innovation Fund Grantees (2014, May), developed by Abt Associates.

Appendix E. Data Availability for Evaluations of RESEA Programs: A Tool to Identify Data Sources

To successfully implement an evaluation of your state's RESEA program, an evaluator will almost certainly need to identify and draw upon your state's UI and RESEA administrative data sources. The information the evaluator seeks to develop an analysis plan may be housed or managed in different data systems. This data availability tool is available to help you find and obtain the data needed to successfully complete an evaluation. You may want to speak to different state workforce agency partners and staff, including RESEA program staff, UI claimant database administrator, labor market information or performance management team, and any other key data or IT staff to determine where different data may be stored, how to access the data, how data are recorded, and to address any potential challenges. This tool may help to facilitate conversations about data sources and availability.

This tool is divided into six sections based on the types of data an evaluator will need to access:

1. UI claimant characteristics and events associated with their UI application;
2. UI payment(s) made to claimants with approved claims;
3. Reemployment services tracking;
4. RESEA eligibility and assignment;
5. RESEA appointment; and
6. UI sanctions/suspensions/reasons for non-payment data.

Each section describes the types of data sources that may be needed and lists relevant data elements from each source. The data elements listed are the most likely needed to complete the RESEA evaluation. However, each evaluation is unique. Each evaluation's specific research questions and evaluation design type will determine what data will be needed. It is possible that access to all the data elements included in the tool may not be needed to successfully conduct any given state's RESEA evaluation. As you complete the six sections of this tool, you may identify data elements that are not available through your state's administrative data system. In these instances, you may want to discuss other possible options for obtaining the data with other state program administrator(s) and the evaluator.

Within each category, a table is provided to help you document where the data are stored and how the data are stored (i.e. are the data stored as text, numerical, date, etc.?). Each table lists likely data elements that may be incorporated into your RESEA evaluation. Each list encompasses a general set of data items that are commonly needed. Your state's interests in a narrower or wider evaluation scope of research questions may require different data. You may add, delete, or move the data items to address the analysis plan, as needed. Blank rows are provided in the table for any additional data elements.

The tool also includes space to document information about potential sources for the data elements needed. Information to gather about each database includes:

- Name(s) of the database(s);
- Agency or office that manages this database;
- Point of contact and email address or phone number;
- Operating system and software requirements needed to access the data; and
- Additional notes about the database, such as access requirements or security restrictions

It may be helpful to print out sections of this tool and use it to guide conversations with the appropriate points of contact for each data system. Another approach to guide conversations may be to send a copy to the database administrator(s) to fill out in advance of those conversations. After you collect this information, you will want to share it with the evaluator. In order to help them develop plans to obtain required data, the evaluator will likely ask you to provide an overview of the available data systems and connect them to key points of contact. Information about the data systems will help the evaluator better determine the appropriateness of each data source for the RESEA evaluation and how to arrange access to it.

1. UI Claimant Data

The evaluator will most likely need information about the characteristics of UI claimants and events associated with each claimant's application for UI. These events include the dates when they applied, the results of the application (approved, denied, withdrawn, etc.), and, if approved for benefits, the weekly benefit amount, date of exit, and the reason for exiting (employment, exhausted benefits, etc.). Evaluators may use claimant demographics and characteristic data for many different purposes. For instance, the evaluator may use the data to conduct descriptive analyses about claimants who engage with RESEA services or to make sure that claimants assigned to treatment and comparison groups in an impact study share similar characteristics.

As each claimant submits an application for UI, the data from that application and the processing of that application is stored somewhere.

The data are likely to be stored in a "master" UI claimant database. This is typically a system that has a record for each individual along with unique events associated with that claimant. While it is likely that all the claimant data are stored in a single database, there is a possibility that information might be stored in multiple databases.

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements*:*

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Date of application		
Demographics characteristics		
Prior employment		
Prior earnings		
Disposition (whether approved, denied, withdrawn, etc.)		
Date of disposition		
Approved weekly amount (if approved)		
Maximum number of weeks (if approved)		
Date exited UI		
Reason for exit (employment, exhausted benefits, disqualified, etc.)		
Employment status (after exiting UI)		
Earnings (after exiting UI)		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

2. UI Payment Data

The evaluator may need information about the UI payments that are made to each claimant included in the evaluation. This includes the following information for each claimant for each week: whether a claim was filed, whether the claim was paid, what the amount of the payment was and when it was made if the payment was approved, and if the payment was not approved, why it was not approved.

You will want to identify the database where this information is stored. It is possible that this information is stored in the "master" UI claimant database identified in the "UI Claimant Data" section.

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements:

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Claimed filed?		
Claim status (approved, declined)		
Date of each payment (if approved)		
Amount of each payment (if approved)		
Reason for Decline (if not approved)		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

3. Reemployment Services Tracking Data

Claimants may be referred to reemployment services as part of their RESEA meeting. The evaluator will need to understand the range of reemployment services that are available to and utilized by claimants included in the evaluation. In particular, the evaluator will likely need to document the services that each claimant is referred to, when they were scheduled to receive a service, and whether they appeared for the services at the appointed times. Therefore, you will need to identify how and where these referrals are tracked in the data systems.

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements:

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Type of service (e.g., career counseling, job search, job placement service, skills training, adult education, etc.)		
Service provider		
Scheduled date		
Received service (Yes/No)?		
Date completed		
Did claimant complete all required activities?		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

4. RESEA Eligibility and Assignment Data

When a claimant is approved for UI benefits, they may be assigned to various programs. For example, they may be assigned to RESEA or WPRS, depending on their "profiling" score. If you are conducting an impact evaluation of a component of a program, different participants may be assigned to different versions of the program (e.g., either a version that includes or one that excludes that component). You will want to ensure the evaluator has access to these assignments so that they can appropriately track activities and analyze outcomes by study group (e.g., intervention and comparison groups). You will want to identify the database that records the eligibility for these programs and records the assignment of claimants to them.

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements:

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Eligible for RESEA?		
Profiling score		
Assignment (e.g., WPRS, RESEA, Intervention group, Comparison Group, etc.)		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

5. RESEA Appointment Data

Claimants selected for RESEA are typically required to appear for a meeting with a RESEA program staff member to review the claimant's ongoing eligibility, receive labor market information, develop or review an employment plan, and obtain a referral for reemployment services. The evaluator will most likely want this information to document claimant's receipt of RESEA services.

You will want to identify the database(s) that stores information about all scheduled RESEA appointments, regardless of whether the appointment was kept or not. This information may include when the appointments were scheduled, whether the claimant appeared, and the content of the meeting (e.g., eligibility review, employment plan review, etc.).

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements:

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Location or site of appointment		
Scheduled date		
Canceled or re-scheduled?		
Did claimant attend (y/n)?		
Date of attendance		
Content of meeting		
Did claimant complete all required activities?		
Outcome of eligibility review (Was claimant disqualified? For what?)		
Was claimant referred to reemployment services? Which one?		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

6. UI Sanctions/Suspensions/Reasons for Non-payment Data

The evaluators may need data on compliance/noncompliance issues for each person in the study. Non-compliance with RESEA program requirements—starting with, but not limited to, attendance at the required initial RESEA meeting—sometimes starts a formal non-compliance process. This process is sometimes called a non-monetary determination and usually consists of multiple formal steps. Depending on the evaluation design plan for the study, the evaluator may want to document when and why the steps occurred for each person and what the outcomes were for each step. If the process resulted in UI sanctions or suspensions, the evaluator will want to know when they started and ended, as well as why they ended.

Use the tables below to document information about the relevant data elements and data systems.

Relevant Data Elements:

Data Element	Name of Database (Where is the data stored?)	Notes about Data Element (e.g., type of data field)
Date of suspension or termination		
Reason for suspension or termination		
Date payments resumed (if ever)		
Reason payments resumed		

Information about Relevant Database(s):

Name of Database	Agency or Office that Manages Database	Point of Contact	Operating System & Software	Additional Notes

Appendix F. Evaluation Timeline Overview and Worksheet

This document provides important information on key considerations that Reemployment Services and Eligibility Assessment (RESEA) program directors and other state workforce agency leaders responsible for RESEA evaluations should think about while planning a timeline for their state's evaluation(s).

Why Is It Important to Create a Strong Evaluation Timeline for Your Evaluation?

Evaluation timelines serve many purposes. First, creating an evaluation timeline is essential for your agency to **create a realistic plan for the evaluation.** This timeline should take into account all the steps you or your evaluator will need to complete during the evaluation—from planning and designing the evaluation to communicating your evaluation findings**.** Your timeline will help set expectations for the timeframe needed to collect and analyze the data needed for the evaluation. Developing a clear, mutually understood timeline for data collection and analysis will help keep your evaluation moving forward as efficiently as possible. Failing to create a realistic and comprehensive timeline can lead to miscommunications between you and your evaluator, a failure to collect and analyze data on time, and other potentially expensive setbacks that can undermine the success of your evaluation or its ability to reliably provide information about your program.

The timeline can also help your team **build support for and cooperation around** the evaluation. Evaluations are often multi-year efforts that require coordination among multiple partners including your RESEA team, staff from various other parts of your state workforce agency, senior agency leadership, and potentially an evaluator from outside your agency. Building an evaluation timeline will help you ensure that key evaluation activities are aligned with intervention activities such that both the evaluation and intervention can operate smoothly. Developing a timeline helps you and your partners arrive at a common understanding for when activities need to occur and deadlines need to be met. By clearly communicating milestones, deadlines, contingencies, and next steps, the timeline can become a tool for building an informed and engaged stakeholder base that supports your evaluation efforts.

Finally, a timeline will help your evaluation **stay on schedule and on budget**.

What Information Belongs in the Evaluation Timeline?

Your evaluation timeline will vary depending on the type of evaluation you are conducting and the research design you select. For example, *prospective* impact evaluations, which enroll participants over a set time period and then follow them for a period, will take longer to plan for, launch, and complete than will a *retrospective* impact evaluation, which uses existing data that has previously been collected.[60] Here are some estimates of time from initial planning through final report for various evaluation types:

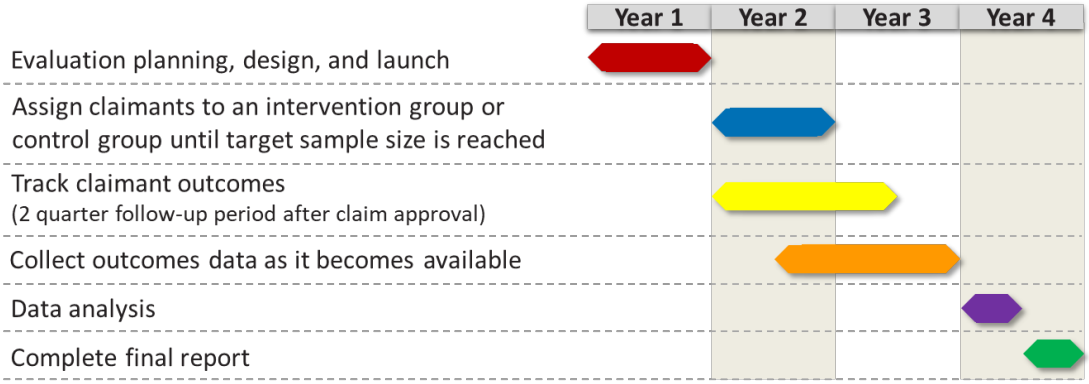
- Implementation study: 12-18 months
- Outcomes study: 12–24 months
- Retrospective quasi-experimental impact study: 18–24 months

[60] For more information on this, please see, "What Evaluation Details Do I need to for a Plan and How Long Will it Take?" by going to https://rc.workforcegps.org/resources/2019/07/30/17/32/RESEA_Evaluation_Evidence_Resources

- Prospective impact study (random assignment or quasi-experimental designs): 4 years or more

Figure 1 (below) provides an illustration of a timeline for an example impact study. The primary reason that prospective studies take longer than retrospective ones is that new data must be collected. If you are developing and testing a new intervention for a prospective study, more launch time may be required for that as well. As shown in the illustration, the impact evaluation spans multiple years to allow enough time to build the sample size and track outcomes for claimants included in the study. Additionally, it is possible that evaluation steps may take place concurrently.

Figure 1. Illustrative Example of Impact Study Timeline



The table below lists six broad **phases of a program evaluation** (planning, designing, launching, collecting data, analyzing data, and communicating findings), with each phase broken into tasks, and each task into its associated activities. The table also provides estimates for how long each type of task typically takes under optimal conditions. Your team can start to build a timeline around these tasks as a starting point and your evaluator may add or remove evaluation activities to your timeline based on the specifics of your evaluation.

You can use this table to take note of the reminders and resources in column 1 ("Task"), and then fill in the Start Date and End Date for each task. The estimates should be regarded as a starting point and tailored to your state's situation as necessary. Once you have obtained the services of an evaluator, it is important to work with that evaluator to create as accurate a timeline as possible, knowing you may have to modify it as your evaluation proceeds. Tasks are listed in the order in which they will typically occur, though some tasks may occur simultaneously. For example, while the evaluation is collecting data, the evaluator also may be analyzing preliminary data and preparing interim reports.

Regardless, the table below will help you think through the steps required for whatever type of evaluation you plan to conduct, and in turn, how long that evaluation is likely to take to complete. The worksheet that follows will help you with step-by-step planning.

Exhibit F-1. Evaluation Phases, Tasks, and Activities

Task	Activities Associated with This Task	Who Is Responsible	Time It Takes	Start Date	End Date
Phase 1 : Planning the Evaluation					
Define your evaluation *All members of your team should be on the same page about what intervention is being tested and what questions the evaluation will answer. You can use the RESEA EvalTA webinars "How Does Evaluation Help My State and Where Do We Start?" and "What Evaluations Are Right for My State?" to begin to think through this task.*	• Identify which intervention(s) your state will test. If your state is interested in testing an intervention or set of interventions that is not currently part of your RESEA program, develop the intervention and train staff to implement the intervention. • Determine research questions, evaluation parameters, and main outcomes of interest. • Identify possible evaluation designs.	• State RESEA agency staff	<1–6 months		
Select an evaluator *State workforce agencies that have a pre-established relationship with a research center or have internal staff who can run an evaluation will not have to advertise for an evaluator. However, staff still will have to spend time creating a detailed scope of work for that evaluator.*	• Assess internal evaluation capabilities and determine the extent to which additional external expertise/resources are necessary to conduct the evaluation design of interest. ○ If you decide to use state's *internal evaluation staff*, establish a firewall between the RESEA program and evaluation staff. It is important for evaluation staff to be independent from program. ○ Or if your state agency has an existing *evaluation partnership* with a third party, determine what steps are needed to include your RESEA evaluation in that scope of work. ○ Or if you determine that you need to hire *an independent evaluator*, you will: ▪ write and publicize the RFP, ▪ review applications, ▪ select an evaluator, ▪ write and negotiate a contract, ▪ approve the selected evaluator and finalize contract.	▪ State RESEA agency staff	1–12 months		

Task	Activities Associated with This Task	Who Is Responsible	Time It Takes	Start Date	End Date
Conduct an evaluability assessment[61] (for impact evaluations) *If you are considering conducting an impact evaluation, an evaluability assessment can help you to assess the whether an intervention is ready to be tested using an impact Your team can use the Evaluation Design Assessment Tool available at the URL below to conduct an evaluability assessment: https://evalhub.workforcegps.org/resources/2018/09/07/19/53/Evaluation-Design-Assessment-Tool*	• Have your evaluator conduct an evaluability assessment. • If the assessment identified areas of program design or implementation that should be strengthened before an evaluation occurs, address these areas prior to the start of the evaluation.	• Evaluator	2–3 months		
Phase 2 : Designing Your Evaluation					
Refine evaluation plans *At this stage in the process, it is important to use the evaluator's expertise to refine your original research questions and preliminary evaluation plans. The evaluator may need several weeks to understand the RESEA intervention being tested, your preliminary evaluation plans, and make revisions as needed.*	• Discuss evaluation goals with evaluator. • Refine research questions of interest. • Define intervention being tested. • Finalize evaluation design. • Define measureable outcomes. • Determine data sources. • Begin to formalize a data sharing agreement and create a plan for accessing data securely (if needed).	• Evaluator • State RESEA agency staff • State data staff	2–6 months		
Submit required package or application to Institutional Review Board (IRB) *Check with your IRB early on to determine how often it meets and when it reviews submissions. Your evaluator should submit evaluation plans to the IRB and account for the time the IRB needs for review.*	• Work with your evaluator, determine which IRB will be used and gather its requirements • Determine human subjects protection needs. • Develop data security and human subjects protection plan. • Prepare and submit a package or application required by the IRB. • Respond to questions from IRB and revise application (as needed).	• Evaluator, in consultation with state RESEA agency staff and U.S. Dept. of Labor	1–2 months		

[61] An evaluability assessment is conducted to determine the extent to which an evaluation of an intervention is feasible. Conducting an evaluability assessment will help states ensure that the intervention to be tested is well-defined and will confirm whether data and other operational resources needed for the evaluation are available.

Task	Activities Associated with This Task	Who Is Responsible	Time It Takes	Start Date	End Date
Write an evaluation design report (EDR) *It is important to not have an elaborate design report slow down the entire evaluation. However, an EDR should include at minimum the following:* • *Purpose, scope, and research questions;* • *Evaluation design type, methods, and procedures;* • *Data collection and analysis plans;* • *Reporting and dissemination; and* • *Evaluation timeline.*	• Have your evaluator create an evaluation design report. • Review the EDR and ask clarifying questions (as needed). • Revise the EDR (as needed)	• Evaluator • State RESEA agency staff	2–4 months		
Phase 3: Launching the Evaluation					
Determine data requirements and modify systems (as needed) *Many of the activities associated with this task should be started early in the evaluation process. It's particularly important to begin formalizing data sharing agreements as soon as possible. Consult with your state workforce agency's legal, contracts, and IT staff on what the process is and how long it typically takes. The timeline to formalize data sharing agreements can vary, from as little as a few weeks to a year!*	• Develop data collection instruments (as needed) • Align data forms and data entry practices across sites. • Finalize data sharing agreements / plan for accessing data. • Make necessary modifications to existing data systems (as needed). • Establish random assignment process (if applicable).	• Evaluator • State RESEA agency staff • State data and IT staff	2–4 months		
Train agency staff *Because you may need to take many staff schedules into account, start conversations around scheduling staff training several weeks before your targeted training date.*	• Develop data collection protocols (written materials). • Train agency staff on evaluation procedures.	Evaluator	2–4 weeks		

Task	Activities Associated with This Task	Who Is Responsible	Time It Takes	Start Date	End Date
Phase 4 : Collecting Data					
Build sufficient sample and collect data (if applicable) *The duration of this task will vary immensely depending on the size of your eligible claimant population, your preferred evaluation design, and your outcomes of interest. For some studies, building sufficient sample will take just a few months, whereas for other studies it may take a few years!*	• Begin assigning claimants to research groups (for *prospective* studies). • Conduct baseline data collection (if applicable). • Notify evaluation participants of assignment status (if applicable) and any additional steps necessary to take part in the program (e.g., orientation, subsequent services, filling out paperwork). • Collect administrative data (e.g., state employment services records, claim applications, Unemployment Insurance wage records). • Conduct survey (if applicable). • Conduct interviews, focus groups, and/or observations (if applicable).	• Evaluator • State RESEA agency staff	Varies depending on evaluation design and sample size needed		
Monitor data and evaluation procedures (*prospective* studies only) *It's important to communicate regularly with your evaluator throughout the evaluation. You can ask your evaluator to provide updates through written progress reports and/or conference calls. You can help your evaluator assess data quality and completeness by routinely providing your evaluator with extracts of program data.*	• Actively monitor evaluation activities. • Provide data extracts to evaluator to check for completeness. • Ensure staff are providing services consistent with participants' assigned status. • Troubleshoot and resolve data issues.	• Evaluator • State RESEA agency staff	Ongoing, throughout entire data collection phase		
Phase 5: Analyzing the Data					
Prepare data for analysis *The time this takes your evaluator will depend on the complexity and integrity of the data as well as the analytic methods used.*	• Clean and organize primary data. • Conduct robustness checks. • Prepare descriptive analysis. • Prepare fieldwork data for analysis.	• Evaluator	2–3 months		
Conduct analysis *To accurately estimate the time needed to conduct the analysis, your evaluator will need to consider your evaluation's outcomes of interest, preferred follow-up period, number of records being analyzed, and complexity of the study design.*	• Conduct analysis according to evaluation design plan.	• Evaluator	2–3 months		

Task	Activities Associated with This Task	Who Is Responsible	Time It Takes	Start Date	End Date
Phase 6: Communicating Findings					
Write interim report(s) *Make sure to communicate with your evaluator in the beginning of the evaluation regarding what deliverables you will require and whether they include one or more interim reports.*	• Write and revise (as appropriate) report. • Review report and ask clarifying questions (as needed).	• Evaluator • State RESEA agency staff	2–4 months		
Write and submit final report *Make sure to leave enough revision time for the final report. Note that each deliverable may require several rounds of comment and revision.*	• Write and revise (as appropriate) report that includes findings from all relevant analyses—outcome/impact, cost, implementation, etc. • Review report and ask clarifying questions (as needed).	• Evaluator • State RESEA agency staff	2–4 months		
Prepare public use data set (if possible) *This may or may not be feasible, but the Department of Labor prefers that evaluations conducted with federal funds generate public use data sets for other researchers to use, to help generate further relevant knowledge.*	• Check data for quality and comprehensiveness. • Remove all PII (e.g., names, Social Security numbers, addresses). • Write comprehensive codebook (i.e., a guide describing each variable in data set).	• Evaluator	2–6 weeks		

Appendix G. Designing & Launching a Random Assignment Evaluation of RESEA Programs

The Random Assignment (RA) Study Checklist outlines steps Reemployment Services and Eligibility Assessment (RESEA) program directors and their evaluators should consider when designing and implementing a RA impact evaluation. An independent evaluator will likely be responsible for carrying out many of the steps included in this checklist. However, in order to develop appropriate evaluation plans and to oversee an evaluator's work, it is important for state RESEA leadership to understand the steps involved in designing and launching a high quality RA impact evaluation.

RESEA program directors can use the checklist as a quick reference guide to ensure that each step has been completed. More specific information on what each step entails is provided in the "Detailed Steps for Planning Your Evaluation" section that immediately follows the checklist.

1. Define Your Evaluation's Focus and Select Your Evaluator

- ❑ Determine which intervention(s) will be tested with the RA impact evaluation as well as in which locations and over what time period the intervention(s) will be tested
- ❑ Select an independent evaluator with the appropriate skill set to design and launch your planned evaluation

2. Plan the Random Assignment Impact Evaluation

Work with your evaluator to:

- ❑ Describe the role of the evaluator to staff and facilitate evaluator's tasks
- ❑ Document the counterfactual condition (i.e., what the intervention being tested will be compared to)
- ❑ Determine outcomes of interest and baseline data requirements
- ❑ Identify appropriate data sources and data elements
- ❑ Identify data access requirements and formalize a data sharing agreement (if needed)
- ❑ Assess suitability of existing data for evaluation purposes
- ❑ Determine the appropriate sample size, ratio, and duration for random assignment

3. Protect the Privacy and Rights of Human Subjects

- ❑ Have your evaluator obtain Institutional Review Board (IRB) approval[62]

4. Design the Random Assignment Impact Evaluation

Work with your evaluator to:

- ❑ Determine the point of random assignment
- ❑ Develop evaluation procedures and protocols
- ❑ Design a randomization algorithm
- ❑ Develop a plan for monitoring service receipt
- ❑ Create a plan for documenting changes to the RESEA program over time
- ❑ Develop a plan for data analysis and reporting
- ❑ Initiate random assignment of claimants and monitor the process

---

[62] Some state agencies have an internal Institutional Review Board that must approve of state-supported studies. Most research firms and all universities do also. If your state agency has its own Institutional Review Board, they may require your evaluator to submit an IRB package to their IRB, in addition to the one that your evaluator may submit to its preferred IRB.

Detailed Steps for Planning Your Evaluation

An Overview: What is a Random Assignment Evaluation?

Random assignment evaluations are impact evaluations that provide scientific evidence of how effective an intervention is.[63] Random assignment works similar to a lottery or flipping a coin. Under a traditional two-armed random assignment design, all individuals who are eligible for the intervention will be assigned at random to one of two groups:

- an "intervention" group,[64] where the individual is able to receive the intervention being studied; or
- a "control" group, where the individual is not able to receive intervention being studied.

The group of claimants who will be randomly assigned will vary depending on whether the random assignment evaluation is testing the whole RESEA program or a specific intervention or component of the program. For an evaluation of the whole RESEA program, each Unemployment Insurance (UI) claimant who is eligible to be selected for RESEA (any claimant who is not in an exempt group) would be randomly assigned to:

- an intervention group that receives the RESEA program services; or
- a control group that does not receive any RESEA program services.

For an evaluation of a component of RESEA, the UI claimants who are selected for RESEA by whichever means the states use would be randomly assigned to:

- an intervention group that receives a version of the RESEA program with the component being tested; or
- a control group that receives a version of the RESEA program without the component being tested.

A claimant's selection into the intervention group is determined completely by chance. Characteristics of the claimant, such as age, gender, race, earnings history, and weekly benefit amount, do not factor into which group the claimant is assigned to. Because claimants are assigned randomly, the intervention and control groups are similar on observable and unobservable characteristics on average over time. As a result, the groups will be systematically different due only to the intervention being tested. Therefore, the evaluation can confidently conclude that any differences in outcomes between the two groups were caused by the intervention rather than anything about the characteristics of the claimants in each group.

Steps in a Random Assignment Evaluation

This document provides state Reemployment Services and Eligibility Assessment (RESEA) program directors a detailed explanation of steps involved in implementing a rigorous random assignment (RA) evaluation. Although many of the steps will be completed by or in collaboration with an independent evaluator, RESEA program directors should be knowledgeable about each step to make sure the evaluation will meet the needs of their RESEA program.

---

[63] Random assignment studies are also referred to as "experimental" evaluations or "randomized controlled trials" (RCTs).

[64] The intervention group may also be referred to as the "treatment" group.

This document is a companion to the "Designing & Launching a Random Assignment Study of RESEA Programs: A Random Assignment Study Checklist." It is a tool to help think through designing and implementing an RA impact evaluation for your state's RESEA program. You may also want to consult other sources of information prepared to support your efforts, including the webinar "Demystifying Random Assignment."

Define your Evaluation's Focus and Select Your Evaluator

Determine Which Intervention(s) Will Be Tested Using Random Assignment

A key first step in designing an impact evaluation is determining what specifically your state is interested in learning about your RESEA program. Your team may be interested in testing the impact of your entire RESEA program on claimant outcomes, such as UI duration, employment, and earnings. In other words, your state would learn, on average, how do outcomes differ for claimants selected for RESEA relative to what their outcomes would be if they had not been selected for RESEA? Alternatively, your state may be interested in learning what difference specific components of your RESEA program (e.g., a second RESEA meeting, intensive case management) makes on claimant outcomes. For example, a state that currently offers only one RESEA meeting, might explore how claimant outcomes would vary if they were to offer two or more RESEA meetings to claimants.

As part of determining what intervention will be tested, your team should start to define the parameters of the evaluation and craft initial research questions that will later be refined in collaboration with the evaluator. This involves identifying the following aspects of the evaluation: timeframe, program sites or geographic area covered, and target population for the evaluation. These elements will form the basis for the evaluation's design and the work the evaluator will do.

You should work with your team to have a sense of what intervention you would like to test and whether an evaluation of the intervention will be feasible before you choose an evaluator. Then, once an evaluator has been selected, your team will work with the evaluator to refine the evaluation design. For more on this topic, see the RESEA EvalTA webinar "Evaluating RESEA – How Does it Help My State and Where Do We Start?"

Select an Independent Evaluator

When selecting an evaluator, consider the individual's or organization's qualifications and experience conducting RA impact evaluations. You should consider their experience implementing and monitoring RA impact evaluations, as well as conducting rigorous analysis and reporting. Your state can tap the expertise of an in-house unit that specializes in program evaluation. If your state makes that choice, you should ensure that the evaluator is not and will not be involved in RESEA program development or administration. Usually that unit and individual employees would not report directly to the state RESEA Director or the state UI Director. If your state does not have much or any experience conducting RA impact evaluations, finding a highly qualified external evaluator will be important.

States will need to develop a Statement of Work (SOW) for the evaluator that describes which evaluation tasks, responsibilities, and deliverables the evaluator will be responsible for carrying out. The SOW should also address practical considerations, such as budget available, period of performance and timelines for the evaluation. Accurately defining the scope of work for your evaluation, complete with an appropriate timeline and budget, will help ensure that you and your evaluator are on the same page about critical aspects of the evaluation. To learn more about this, see the RESEA Evaluation Technical Assistance (EvalTA) webinar "Procuring and Selecting an Independent Evaluator."

Plan the Random Assignment Impact Evaluation

In order to design and implement an RA impact evaluation, RESEA program directors will need to work with their evaluator to:

Describe the Role of the Evaluator to Staff, and Facilitate the Evaluator's Efforts
Strong, experienced evaluators will provide appropriate direction for the overall evaluation and help troubleshoot issues that may arise during its course. Evaluators should provide input into, if not fully develop, the evaluation timeline, the RA design, training materials on study procedures for program staff, and a data analysis plan. Evaluators lead data analysis and reporting as well. (See the Evaluation Timeline brief for more on that topic.) In order to ensure that the evaluator can be most effective in their work, state RESEA leaders should inform their staff of the evaluator's role and how the evaluator will be engaging with staff over the course of the evaluation. RESEA leadership should also strongly encourage staff to facilitate evaluation efforts to the extent possible.

Document the Counterfactual Condition
It will be important to understand what services the control group is receiving, because those conditions represent what the intervention group would experience absent the intervention, called the ***counterfactual***. The evaluator must documenting the counterfactual, including identifying what other similar services are available in the community to claimants in the study. Doing this will help your state understand how different the experience of the intervention group members will be from the experience of the control group members, called the ***service contrast***. Understanding the service contrast is necessary.

Determine Outcomes of Interest and Baseline Data Requirements
Work with your evaluator to decide which outcomes the evaluation will use to estimate impact of the intervention, and how your evaluation will measure those outcomes. Likely outcomes of interest will include UI duration (i.e., number of weeks claimed and/or paid), UI benefit (e.g., benefit amount received), employment, and earnings. Your state may also want your evaluator to analyze more intermediate outcomes, such as attendance at the RESEA meeting, response to non-attendance, and services received.

Your evaluator will also want to include baseline data on claimant or claim characteristics in the analysis. Baseline data refers to information collected about claimants in the study before they received any of the services that are being evaluated. The baseline data provides a snapshot of the claimant's initial circumstances. Baseline data might include study participants' information at the time they apply for the initial UI claim (claimant demographics) and their previous claim history (including number of weeks claimed and benefit amounts) to include in the analysis. The evaluator will also want information about individuals' randomization status (including reasons for exemption from randomization). If your team is interested in learning more about how impacts vary by certain claim or claimant characteristics, you should work with your evaluator to determine what additional data should be collected and analyzed.

Identify Appropriate Data Elements and Data Sources
Having identified the outcomes of interest and the baseline data requirements, your evaluator will need to identify the specific data elements needed and where the data will come from (the data source). Most likely the data needed are already being recorded in your state administrative data systems, such as the Employment Services System, UI wage, or UI claims data.

Identify Data Access Requirements and Formalize a Data Sharing Agreement (if needed)
Having identified the data elements to be collected and sources for the data, your evaluator will need to determine how to access the necessary data. This consideration is particularly important

for evaluations that will rely on administrative data sources, as access to these data sources are typically more restrictive. Your evaluator may need to enter into a formal data sharing agreement with the third party organization or state agency that maintains the data. Data sharing agreements commonly include information about: specific data to be shared; rules associated with storing, sharing, and using those data; name and information for a data custodian at both parties to the agreement; time period covered by the agreement; and any requirements for destruction off shared data upon completion of the evaluation. Your evaluator may need your help in determining who to contact to access each data source, but your evaluator will be responsible for setting up and executing the data sharing agreement with the appropriate parties.

Assess Suitability of Existing Data for Evaluation Purposes

If the evaluation will rely on data from your state's administrative data system, your evaluator should discuss the quality and completeness of the data with knowledgeable staff. Such staff may include IT staff, program administrators or lead staff, state labor market information directors and other staff responsible for entering and maintaining the data. Then, your evaluator will need samples of the data to verify the quality of the data. Your state's data management staff can produce test files containing data that your evaluator can use to run various analyses. Based on the initial data quality findings, you and your evaluator may need to develop additional monitoring plans to troubleshoot any identified issues. Such an assessment will help your evaluation team determine whether the administrative data already being collected are sufficiently detailed and complete to allow the evaluation to fully describe claimants' experiences while receiving UI. This step is vital because your evaluation will only be as strong as the quality of your data. For more on this topic, see the RESEA EvalTA webinar, "Assessing Data for Your RESEA Evaluation."

Determine Appropriate Sample Size, Ratio, and Duration for Random Assignment

The number of claimants assigned to the intervention and control groups (combined, known as ***sample size***) needs to be large enough to detect impacts of a certain size to be relevant or expected. The smaller the impact to be detected, the larger the sample must be. The minimum sample sizes needed to detect RESEA impacts on employment and UI duration outcomes are likely larger than you would expect—that is, several to many thousands of claimants. This is especially true if your evaluator plans to use an unequal random assignment ratio. The random assignment ratio refers to the percentage of claimants in the study who were assigned to the intervention group versus the comparison group. Using a 1:1 RA ratio (i.e., one claimant is assigned to the intervention group for each claimant assigned to the control group) will give you the greatest statistical power—thus minimizing the size of the sample you will need. Studies that use unequal RA ratios (e.g., 2:1) will need larger sample sizes in order to obtain the same statistical power as a study using a 1:1 ratio. When determining which RA ratio to use, your evaluator will need to consider the annual number of UI claimants eligible for RESEA in your state for whole RESEA program evaluation or the annual number of UI claimants selected for RESEA for a component evaluation. Your evaluator also will need to consider the available funds to serve them.

It will be important to make sure *prior to beginning the evaluation* that you and the evaluator have a realistic understanding of how large a sample the evaluation will need to detect the tested intervention's impact on the evaluation's outcomes of interest. Statistical methods are available to calculate minimum sample size, depending on the intervention and outcomes you are evaluating. The RESEA EvalTA team can help you determine the sample size required for your intervention and selected outcomes by contacting us at RESEA@abtassoc.com. An experienced evaluator will know how to perform these calculations.

Protect the Privacy and Rights of Human Subjects

Have Evaluator Obtain Institutional Review Board Approval
It is vital to ensure that your evaluation appropriately protects its human subjects. Federally funded research and evaluation involving human subjects must comply with federal and state laws and regulations governing their ethical treatment, privacy, and other rights. Compliance begins by having your evaluator submit the study design to an Institutional Review Board for review. The role of the IRB is to ensure the evaluation has appropriate protections in place *before* the study begins. Often an IRB will exempt a study from a full IRB review or from informed consent provisions, but that is a decision for the IRB to make. The evaluator needs to get an official IRB approval (or exemption).

Most universities and some experienced research organizations have an internal IRB. Some states also have their own IRBs. In some cases, IRB approval may be required by both your state agency's IRB (if you have one) and the IRB in the evaluator's institution or organization.

Have your evaluator reach out to the IRB(s) ahead of submitting an application in order to understand the IRB's procedures, required materials, meeting schedule, and timeframe for submitting materials for review. Firm knowledge of the IRB process will help avoid delays with implementing the evaluation, as the evaluation will not be able to proceed without IRB approval or exemption. For more on this topic, see the RESEA EvalTA webinar, "What Evaluation Details Do I Need to Plan and How Long Will it Take."

Design the Random Assignment Impact Evaluation

Your evaluator will be primarily responsible for much of what occurs in designing and launching the random assignment evaluation. But your RESEA program staffs' continued input and cooperation throughout this phase will be needed for the success of the study.

Determine the Point of Random Assignment
Work with your evaluator to determine where to conduct random assignment into your existing process for selecting RESEA claimants. Typically random assignment will occur as part of the selection of claimants for RESEA. In a study of a whole program, random assignment will be part of the mechanism for determining which non-exempt claimants are selected for RESEA. In a study of a particular component of your RESEA program, selection of claimants for RESEA will occur first, after which random assignment will be done to determine whether the RESEA claimant receives a version of the program that includes the component or excludes the component. Random assignment must also occur *before* any RESEA services are received.

Develop Evaluation Procedures and Protocols
Work with your evaluator to develop the following materials:

- A Random Assignment Training Guide: This guide will serve as a reference for staff on all study procedures as they have to do to implement the study. The training guide should include an overview of the evaluation and step-by-step instructions for performing study intake, conducting random assignment, and ensuring human subjects protections and data security.
- A Data Security Plan: This plan should document how sensitive, personally identifiable information will be collected, transmitted, and stored; who will have access to the data; and a plan for data destruction after the evaluation has concluded.

Formal evaluations typically must inform people about their participation in a study, including random assignment, and how their data will be used. It is possible that the data use descriptions

that your state already provided to individuals when they file their UI claims will be sufficient to cover informed consent requirements. The IRB will review the informed consent procedures planned. Generally, evaluations of RESEA program will be able to apply to the IRB for waivers of informed consent because the RA process is part of regular operations. However, this will depend on your state's laws. If your evaluator is unable to get an informed consent waiver, your evaluator will need to develop:

- Informed consent forms for claimants: Informed consent forms will discuss the purpose of the study, risk and benefits of participation, the voluntary nature of participation, among others details of being a participant in the study.
- Scripts for staff during the study eligibility and enrollment process: These scripts will help staff explain the evaluation, the consent form, the baseline information form, and the results of random assignment to potential study participants.

Design a Randomization Algorithm

Work with your evaluator to determine how random assignment will be implemented. Good automated options include developing and inserting a randomization algorithm into your existing scheduling system or developing an external database or web-based platform to conduct random assignment. Once programmed, the computer algorithm can be easily monitored and difficult to tamper with, making it easier to ensure that the process is working as it should. Using a computer algorithm also reduces burden for program staff who might be involved in the process by depersonalizing the random assignment process and being incorporated into existing processes. If using a computer algorithm is not feasible for your state, manual random assignment is possible (e.g., staff using a table of random numbers). But manual random assignment is much more difficult to implement successfully and is not recommended. There are benefits and drawbacks to each approach, including user friendliness and compliance. You should work with your evaluator to determine which option is best for your state and the evaluation.

Develop a Plan for Monitoring Service Receipt

As part of the evaluation, your evaluator may need to monitor what services intervention and control group members are receiving to ensure that the intervention is being implemented as planned. Monitoring allows you to identify and correct implementation issues, so that your findings really do capture the impact of the intervention that you intended to evaluate. This may be particularly important for states planning to test a component of their RESEA program, which is likely to be more complex to implement. The monitoring plan will likely involve examining data on random assignment and service receipt.

Create a Plan for Documenting Changes to the RESEA Program Over Time

States may need to make changes to the RESEA intervention prior to the evaluation start. Changes may include tailoring the RESEA intervention to include certain program elements for the intervention group and other elements for the control group. However, once the evaluation is implemented and random assignment has begun, from a research perspective, it is best to minimize any changes to the RESEA program model for the duration of the evaluation. But for practical reasons, this may not always be possible. In preparation for the analysis and reporting phase, states will need to work with their evaluators to record any changes to the intervention, RESEA program, or policies that occur after the study begins.

Develop a Plan for Data Analysis and Reporting

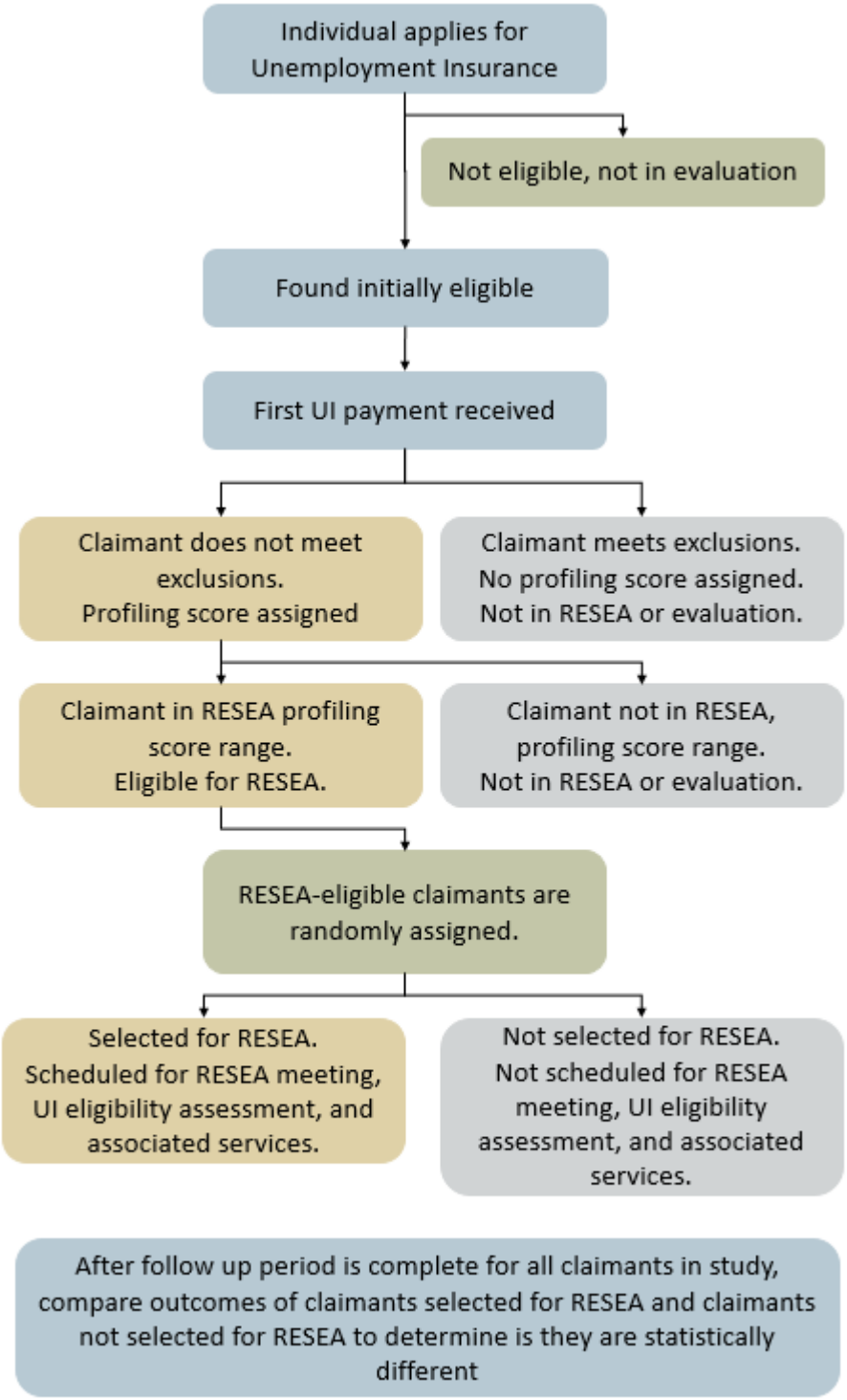
In a random assignment study, impacts are estimated by analyzing differences in the average outcomes of the intervention group versus the control group. Analyses may involve making technical decisions in areas such as control variables to use, weighting (if complex random

assignment probabilities are used), and approaches to supplemental analyses such as subgroup analyses, mediational analyses (analyses of how intermediate outcomes lead to final outcomes), or cost-benefit analyses. Your evaluator should determine up front what the set of analyses is that will be conducted, what the approach to those analyses will be, and how results will be reported. The evaluator should communicate these decisions to you through an analysis plan.

Initiate Random Assignment of Claimants and Monitor the Process

With the random assignment process established and RESEA program staff trained, RESEA program staff will begin to conduct study intake and random assignment of claimants. The start date should be determined in collaboration with your evaluator. The evaluator should be readily available to answer questions and troubleshoot issues that may come up. The evaluator should also continue to monitor the random assignment process to make sure that it is being carried out as intended.

Appendix H. Sample Random Assignment Flowchart



Appendix I. Evaluation Design Report Template

This document aims to familiarize Reemployment Services and Eligibility Assessment (RESEA) program directors with the purpose and format of an Evaluation Design Report (EDR). An independent evaluator will be responsible for writing the EDR. However, it will be beneficial for RESEA program leadership to consider the different types of information included in the EDR in order to anticipate questions for which evaluators may require the input of RESEA program staff, as well as to ensure that the EDR aligns with the goals of the evaluation.

What Is an Evaluation Design Report (EDR)?

An evaluation design report (EDR) is typically a narrative-style written document that describes all aspects of the pending evaluation, including plans for data collection, monitoring, analysis, and reporting. Generally, an EDR is 30 to 70 pages long for a major evaluation, but it could be shorter, depending on the study and the evaluator. Written by the evaluation team, an EDR not only documents the planned technical approach, but also guides the RESEA program staff and evaluator in understanding the complete series of evaluation activities. That is, the EDR provides a roadmap to what those activities are, how they will be implemented, when they will start and conclude, and who will be involved in executing each activity.

EDRs typically include information on the following:

- Study purpose and scope.
- The intervention being tested, including a program-specific logic model.
- Research questions that are measurable, relevant to the program, and time-bound.
- Evaluation method(s), including the study design type; methods for creating a treatment and comparison group (if applicable); and proposed outcome measures and other data needed that are valid, appropriate, and measurable.
- Data collection plans, including data elements of interest, their associated data sources, and data collection period.
- Data analysis plan and methods, including suitable controls for mitigating any threats or risks to successful interpretation of findings and overcoming any limitations.
- Timeline and milestones for program services and evaluation activities, as well as information on how evaluation activities and timeline dovetail with intervention activities and timeline.
- Plan for reporting details to convey the evaluation's progress, results, and findings.

The evaluator is responsible for writing the EDR, but the RESEA program staff will need to be available to answer questions about the program and to review the draft EDR. The evaluator may have questions about the RESEA program's design, staff roles, and other program aspects that have implications for the evaluation. The RESEA program staff should review the planned evaluation design to ensure that it meets the evaluation's scope. They also should understand how various evaluation activities link to or coordinate with program operations. It is important for the evaluator and RESEA program staff to have a mutual understanding and agreement on the evaluation design prior to the start of the evaluation.

How to Use the Evaluation Design Report Template

Given the critical importance of the EDR, the RESEA EvalTA team has prepared an EDR template (in the next section) for evaluators charged with preparing or overseeing the preparation of the report. Within each section, the template briefly explains what information is to be included, questions that should be addressed, and links to helpful resources, where available. The topics discussed in this document, if fully articulated in the EDR, will help to ensure the rigor of evaluation. However, the organization of the EDR does not need to conform to the format of this template. The template should be viewed as a flexible set of guidelines to be customized to the type of evaluation planned for your program.

The exhibit below maps out the EDR template sections by study design type. Note that your state may wish to conduct several evaluation types at once; for example, combining an impact study and an implementation study. If so, your EDR will need to include sections for all relevant study types.

Exhibit I-1. EDR Sections, by Study Design Type

Section	Evaluation Design Type			
	Implementation Study	Outcomes Study	Impact Study	
			RCT	QED
Introduction				
Intervention Background (incl. logic model and target population)	X	X	X	X
Literature Review and Evidence Base	X	X	X	X
Evaluation Purpose and Research Questions	X	X	X	X
Evaluation Design				
Research Questions	X	X	X	X
Outcome Measures		X	X	X
Study Sample		X	X	X
Treatment and Comparison Group Description			X	X
Data Sources and Collection				
Administrative Data	X	X	X	X
Survey Data	X	X	X	X
Interviews	X			
Document Review	X			
Analysis and Interpretation				
Analytic Methods	X	X	X	X
Model Specification			X	X
Baseline Equivalence Testing			X	X
Minimum Detectable Impacts (MDI)			X	X
Multiple Comparisons Problem			X	X
Subgroup Analyses		X	X	X
Risks and Challenges				
Threats to Validity		X	X	X
Human Subjects Protections	X	X	X	X
Data Security	X	X	X	X
Logistics				
Reporting	X	X	X	X
Schedule	X	X	X	X

The RESEA EvalTA team is available to answer questions or have conversations with states about their EDR, as needed. To contact the RESEA EvalTA team, email RESEA@abtassoc.com.

Evaluation Design Report (EDR) Template

1. Introduction

The introduction section provides an overview of the intervention that is being studied, the existing evidence base for the intervention, and the evaluation's scope and purpose.

Intervention Background

Describe the purpose and scope of the RESEA ***intervention*** that will be the focus of the evaluation. The evaluation may examine the entire RESEA intervention or a particular element of the intervention. The EDR should provide details about how the intervention functions and whom it seeks to serve (the target population). Although the RESEA program staff and evaluator are familiar with the RESEA intervention, providing a summary serves as a critical foundation for (1) clarifying the intervention activities; (2) identifying specific aspects of the intervention that will be evaluated; and (3) establishing clear research questions.

This section should also include a ***logic model***, which is a detailed description of the intervention as it is planned to operate. The logic model may be presented both as a flowchart or other graphic representation and as narrative description. Logic models convey information about the intervention's ***inputs***, ***activities***, ***outputs***, and ***outcomes***. It also accounts for ***external conditions*** (e.g., economic and political changes, other programs available in the community) that may affect the intervention's intended short- and long-term outcomes. The logic model also articulates the intervention's underlying ***theory of change***; that is, the hypothesized causal relationships among the intervention activities, outputs, and outcomes. Together, the logic model and theory of change help to establish the rationale for the planned evaluation's design and methods.

Questions to Answer

- What does the **interven**tion intend to do or change? And for whom?
- What are the main components of the intervention? And what aspects are being evaluated?
- What are the critical inputs required for implementing the intervention activities?
- What are all of the intervention activities and services provided to claimants?
- What are the expected immediate, short-, and long-term outcomes for claimants? And how are they linked?
- Does the logic model tell a clear, integrated, and complete story about the intervention?
- What, if any, assumptions have been made about external factors that could affect implementation of the intervention and its impacts?

Resources

- 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
- B. Lawton, P.R. Brandon, L. Cicchinelli, & 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: https://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=404

Literature Review and Evidence Base

The EDR should provide a discussion and analysis of the evidence base for the RESEA intervention being evaluated. Clarifying what is already known will help the evaluator, RESEA program staff, and other stakeholders understand why the chosen research questions are important, what the evaluation might expect to find, and what might explain those findings. Evaluators should review, synthesize, and summarize research on and evaluations of relevant interventions. In addition to the study findings, the evaluator should also pay careful attention to the study design and methodology; for example, types of data sources, data collection and analysis methods, and rigor of the study design.

The U.S. Department of Labor's Clearinghouse for Labor Evaluation and Research (CLEAR) is a valuable resource for RESEA-relevant evidence. CLEAR provides reviews and summaries of research on a range of employment and labor topics as well as assessing the study quality of causal impact studies.[65]

Questions to Answer

- What are relevant past interventions?
- How were relevant interventions studied?
- What were the study findings, and are they reliable?
- What gaps, if any, exist in the knowledge base?
- How will the planned evaluation enhance the broader workforce system or contribute to the workforce evaluation literature?

Resources

- **Clearinghouse for Labor Evaluation and Research**: CLEAR reviews studies in a variety of labor-related topic areas. https://clear.dol.gov/
- **CLEAR's Reemployment topic area**: This topic area focuses on studies of interventions designed to promote faster reemployment of unemployment insurance (UI) claimants. https://clear.dol.gov/topic-area/reemployment

Evaluation Purpose and Research Questions

The EDR should provide a statement establishing the purpose and scope of the evaluation. This statement should clearly indicate which aspect(s) of the RESEA intervention the evaluation will study, what outcomes the intervention aims to change and over what time period, and who comprises the target population.

The evaluator should also specify the ***research questions*** that the evaluation aims to answer about the RESEA intervention. Clear, meaningful research questions are a critical part of any evaluation. The research questions should align with the intervention's logic model and reflect the evaluation's purpose and scope. In turn, the research questions will drive the evaluation's study design and methods. To that end, the research questions should be specific and measurable (i.e., identify specific elements or outcomes to examine); answerable; discrete and limited in number; and rooted in program knowledge and realistic expectations.

The research questions may be grouped by question type: implementation, outcome, or impact. ***Implementation questions*** ask how the intervention is operating (e.g., "How do claimants flow

[65] CLEAR provides study quality ratings, which it calls *causal evidence ratings*, only for casual impact research studies. Causal studies include experimental and non-experimental research studies (e.g., randomized controlled trials/RCTs and quasi-experimental designs/QED) that attempt to estimate the impact of a program, policy, or intervention on outcomes.

through the steps of our RESEA program?"). ***Outcome questions*** ask about the extent to which the intervention is meeting its goals (e.g., "What percentage of claimants appear for their first scheduled meeting?"). ***Impact questions*** aim to understand whether the intervention *causes* changes in claimant outcomes (e.g., "Does our RESEA intervention improve claimants' employment and earning outcomes compared to what it would be without the intervention?"). For impact questions, the evaluator should also discuss any underlying hypotheses. The question types will depend on the type of study being conducted.

Questions to Answer

- What aspects of the intervention will the evaluation examine? What is the evaluation's time frame?
- What outcomes does the evaluation plan to measure to demonstrate changes or successes? What is the time frame for tracking outcomes, including the start point (e.g., claim application date or receipt of services) and duration?
- Whom does the intervention seek to serve?
- Are the research questions logical and linked to the specific intervention?
- Do the research questions meet the evaluation's purpose and scope and reflect the rigor of the study design?
- For impact questions, are there hypotheses about expected outcome changes that are related to the intervention?
- For outcomes and impact studies, do the research questions identify the outcome(s) of interest and outcome measure(s)? Is the time frame specified for measuring the outcome(s) and, if applicable, for the follow-up time period?

2. Evaluation Design

The evaluation design and methods section provides an overview of the study design type(s) selected for the evaluation, as well as the methods the evaluators will use. This section should also provide justification for the approach selected and explain its appropriateness for the evaluation's purpose and research questions.

There are three major types of evaluation studies:

- **Implementation studies** aim to document how an intervention operates and what activities and services claimants engage with. Implementation findings can be used to determine whether a RESEA intervention is being carried out in a manner consistent with its goals and design. Implementation studies can stand alone or be combined with outcomes or impact studies. In combination, the implementation study helps contextualize and explain outcomes or impact findings.
- **Outcomes studies** aim to understand whether a RESEA intervention meets its targets. Outcomes studies generally analyze observed characteristics of claimants following their participation in the program and assess those characteristics against program goals.
- **Impact studies** aim to determine *whether* the RESEA intervention was the cause of a change observed in claimant outcomes, and if so, *how much* change did the intervention cause. Impact studies allow evaluators to determine how specific claimant outcomes *with* the intervention differ from what outcomes would have been *without* the intervention. Two common types of impact studies are (1) experimental studies, or ***randomized controlled trials (RCTs)***; and (2) ***quasi-experimental design (QED)*** studies.

The EDR should discuss the subsections that follow for each evaluation type, if the evaluation will conduct more than one study type. For example, if the evaluator plans to conduct both an implementation study and an impact study, the design and methods should be discussed for both study types.

Questions to Answer

- Is the evaluation design an impact study or an outcomes study? And specifically, what type of impact or outcomes study (i.e., an RCT, a QED, a pre-post design, etc.)?
- Why is the study design appropriate (i.e., will it answer the evaluation research questions)?
- For impact studies, how are the treatment and comparison groups defined?
- For outcomes and impact studies, what are the outcomes of interest being examined? How are the outcomes defined?

Resources

- **"How Does Evaluation Help My State and Where Do We Start?":** This webinar describes initial steps states can take to plan evaluations of their RESEA interventions. https://www.workforcegps.org/events/2019/04/17/16/43/Evaluating-RESEA-How-Does-it-Help-My-State-and-Where-Do-We-Start
- "**Which Evaluation Designs Are Right for My State?":** This webinar introduces states to a range of evaluation design types and describes what kinds of information can be learned from each type. https://www.workforcegps.org/events/2019/05/07/13/07/What-Evaluation-Designs-Are-Right-For-My-State
- **"De-Mystifying Random Assignment Designs for RESEA":** This webinar aims to increase states' familiarity and confidence with random assignment evaluation designs. https://mahernet.adobeconnect.com/_a14339732/p9uumcab4ro3/
- **"Evaluations Using Your Existing Administrative Data: Quasi-Experimental Designs":** This webcast provides states with an overview of QEDs and when QEDs may be appropriate to use. https://mahernet.adobeconnect.com/_a14339732/pubd2hfg1ksi/
- **"Implementation Studies":** This webcast discusses how to use implementation studies to gain insights into how a RESEA intervention is operating and how these insights can support an impact evaluation and intervention improvement. https://mahernet.adobeconnect.com/_a14339732/pxdxnu8sqy9w/

Research Questions

The research question(s) for each of the evaluation's study types should be specified or reiterated, if previously included in the introduction, in this section. Including the research questions at this point in the EDR will help to make it explicit how they are linked to the selection and specifications of the evaluation design as well as to the individual outcomes or impact measures.

Outcome Measures (for Outcomes and Impact Studies)

The EDR should describe the main outcomes of interest for the evaluation and specify how the outcomes will be measured. Each outcome measure should be aligned with at least one research question of interest, the intervention logic model, and the type of data collected and when it is collected.

Thinking through how the outcomes will be measured will help to ensure that the appropriate data will be collected. For example, for an evaluation designed to examine intervention impacts on the outcome of *earnings*, the outcome measure might be *claimant earnings at 18 months after second RESEA meeting,* where requiring that claimants attend a second one-on-one meeting is the

intervention. For this outcome measure, the evaluator would need to collect information on claimants' earnings when they first entered the study (at "baseline") and then at 18 months post-intervention.

If there are multiple outcomes of interest, the evaluator should indicate which outcomes are ***confirmatory*** (indicating whether the program is on the right track) and which are ***exploratory***. These are sometimes called *primary* and *secondary*.

Questions to Answer

- What are the outcomes of interest? Which are confirmatory and which exploratory?
- How will the outcomes of interest be measured? Are the proposed measures valid, appropriate, and reliable?

Study Sample (for Outcomes and Impact Studies)

The EDR should describe the study sample, including information about target sample size and whether the full population of claimants served or only a subset of the population will be included in the evaluation. If only a subset of claimants will be included, the evaluator should provide a rationale for the sampling scheme and any exclusion or inclusion criteria that will be used to select the sample.

Treatment and Comparison Group Descriptions (for Impact Studies Only)

For impact studies, the evaluator will need to describe the treatment group and comparison group conditions as well as the process that will be used to create the two groups. That process will vary depending on the specific type of impact study planned (i.e., RCT or QED). In a properly designed and implemented impact study, the experiences of the comparison group will reflect what the experiences of the treatment group *would have been* without the intervention.

For RCTs, the EDR will need to provide a detailed, step-by-step description of the process that will be used to randomly assign claimants to the treatment group or control group. The description of that random assignment process should include:

- Participant consent procedures;
- Point of random assignment—that is, *when* during the RESEA program random assignment will occur;
- Mechanics of random assignment—that is, *how* claimants will be assigned to one group or the other (e.g., computer algorithm, manually using a table of random numbers);
- Random assignment ratio—that is, one claimant assigned to the treatment group for every claimant assigned to the comparison group (1:1) or some other ratio?

The EDR should also describe *how* data about the claimants at baseline (i.e., the point at which claimants are enrolled in the study but before they receive any services) will be collected, and the *type* of data collected. Baseline data typically include information about claimants' demographics and education and employment history, important for measuring program impacts and for determining baseline equivalence between the treatment and control groups.

The evaluator should discuss plans for minimizing crossover (i.e., when a comparison group member accesses the intervention) and for monitoring the random assignment process. Monitoring includes:

- Ensuring staff follow the random assignment procedures and do not attempt to circumvent them.
- Tracking claimants in the study to ensure random assignment integrity and fidelity.
- Documenting program services provided and accessed by treatment group members.

The EDR should describe the *counterfactual*—that is, the conditions that claimants would have experienced in the absence of the intervention.

For QEDs, the EDR should clearly define the characteristics of the treatment group members, including the eligibility criteria. If applicable, the procedures for study intake should be described. The EDR should document RESEA program activities, and participation of treatment group members in the intervention.

The EDR should describe the comparison group, including the source of the comparison group members, and demonstrate how those members are as similar as possible to the treatment group members. The EDR should also articulate the source of data for the comparison group, what type of data will be collected and when, and to what degree those data are comparable to what will be collected for the treatment group.

The evaluator should describe a clear plan for tracking both groups and strategies for ensuring there is no overlap between the treatment and comparison groups.

Questions to Answer

- What are the treatment group and comparison group conditions?
- How will the treatment and comparison groups be formed? Is the source(s) of the comparison group random assignment or other selection procedures?
  - In RCTs, how will the evaluator ensure that random assignment has been faithfully executed?
  - In QEDs, from what population(s) will the comparison group sample(s) be drawn? Is the comparison group similar to the treatment group?

Resources

- **"De-Mystifying Random Assignment Designs for RESEA":** This session aims to increase states' familiarity and confidence with random assignment evaluation designs. https://mahernet.adobeconnect.com/_a14339732/p9uumcab4ro3/
- **"Evaluations Using Your Existing Administrative Data: Quasi-Experimental Designs":** This webcast provides states with an overview of QEDs and when QEDs may be appropriate to use. https://mahernet.adobeconnect.com/_a14339732/pubd2hfg1ksi

3. Data Sources and Collection

In the EDR, the evaluator will need to identify the data required for measuring the outcome(s) of interest, the source of the data, and a plan for collecting the data. Doing so will help ensure that the evaluator will collect data that is reliable, accurate, and complete for answering the evaluation's research questions.

Four data sources may be useful to an RESEA evaluation: administrative data, survey data, interviews, and document review. Evaluations may rely on one or more data sources. The EDR should describe in detail each data source that will be used, including:

- ***Type(s)*** of data to be collected from the data source, and the extent to which that data will cover the population to be included in the analysis.
- Potential ***challenges*** with collecting or accessing the data. For example, if the data will be coming from a third party or another state agency, the evaluator will need to describe the plan to gain access to the data.
- ***Time period*** for data collection. It is critical for the evaluator to determine whether the desired data can be collected at the appropriate time. For impact studies, data typically needs to be collected at baseline (i.e., when they first enter the study) and at follow-up (i.e., a specified point after the claimant enters the study for which outcomes data will be collected). The evaluator should specify a plan to ensure that follow-up data will be collected for all study participants—claimants in both the treatment group and the comparison group—for a uniform time period.

Points to cover in the EDR with respect to each kind of data source are discussed below.

Questions to Answer

- Is there a data source for each outcome measure?
- What methods will be used to collect data from each of the data sources?
  - What is the time frame for data collection?
  - How frequently will data be collected?
- For impact studies:
  - What is the source for baseline data? What type of baseline data will be collected? Is the baseline data appropriate and adequate for measuring changes in outcomes?
  - How will data be consistently collected across treatment and comparison groups?

Resources

- **"Assessing Data for Your RESEA Evaluation":** This webinar prepares attendees to discuss and assess options to use administrative data to support evaluations. https://www.workforcegps.org/events/2019/07/23/13/26/Assessing-Data-and-De-Mystifying-Random-Assignment-for-your-RESEA-Evaluation-Office-Hours-Session-1

Administrative Data

Administrative data refers to the information that is routinely collected as a part of the regular administration of program or intervention activities. State agencies, such as employment services, Unemployment Insurance, and employment security agencies, typically record information about their clients and the administration of their programs, which can facilitate RESEA evaluation.

In the EDR, the evaluator should describe the administrative data sources, including details on the entity collecting the data and how data are collected. The evaluator should also describe the samples and larger population from which data are collected, as well as the outcome measures that will come from these data. If applicable, the evaluator should discuss the procedures necessary for obtaining consent to access the data and the timeline for doing so. For impact studies, the evaluator should describe plans for obtaining comparable administrative data for the treatment group and the comparison group.

Survey Data

A survey is helpful for collecting information from key informants on indicators that may not be available from administrative sources. Surveys might be used to collect detailed information about claimants' satisfaction or experiences with the intervention, or RESEA program staff experiences implementing the intervention.

If surveys are planned, the evaluator should describe the purpose of the survey (e.g., which research question(s) the survey responses will help answer). The evaluator should detail the plan for administering the survey, including timing and frequency, sampling plan (if the survey will not be administered to the entire population of study participants), types of questions to be asked, and administration mode (e.g., web, phone, in-person). The EDR should also discuss the anticipated response rates and strategies for ensuring rates are sufficiently high.

Interviews

Interviews are particularly helpful for gathering information about program implementation. As such, interviews are commonly used for implementation studies. When conducted in conjunction with an outcomes or impact study, interview information could be valuable for interpreting and contextualizing the study findings.

In the EDR, the evaluator should describe the plan for conducting the interviews. This description should include information about who will be interviewed (e.g., program staff, claimants), format of the interview (e.g., in person, by phone, virtually), types of questions to be asked, anticipated number and length of interviews, and the timing of the interviews. The evaluator should consider including sample interview guides or other supporting materials in an appendix to the EDR.

Document Review

Program documents (e.g., intervention manual, program memos) are a valuable source of information about RESEA program design and operation. If document review is planned, the EDR should describe the types of documents to be obtained and reviewed, how, and by whom.

4. Analysis and Interpretation

Plans for data analysis and interpretation of results must be specified in the EDR. This means describing the analytic method(s) and rationale(s) for choosing each method.

Analytical Method

The analytic method must be aligned with and appropriate for the evaluation design type.

Implementation Study. Unlike impact and outcomes studies, which are quantitative, implementation studies typically rely heavily on analysis of qualitative data such as document reviews and transcripts from interviews or focus groups. The EDR should describe the analysis plan, including any use of software (e.g., NVivo). If the implementation study will be conducted in conjunction with an impact or outcomes study, the evaluator should discuss how the implementation results will inform the interpretations the impact or outcomes study findings.

Outcomes Study. Outcomes studies may involve multiple analysis methods. For example, an outcomes analysis might involve conducting statistical tests—such as basic *t*-tests, chi-square tests, and analysis of variance—in order to determine correlation or covariance. The evaluator should

describe each specific analytic method to be used and identify the outcome(s) associated with it. The evaluator should also discuss, if applicable, methods for determining statistically significant differences in outcomes for subgroups (e.g., by gender, by age).

Impact Study. RCT and QED impact studies typically involve specific types of statistical analysis for comparing outcomes between the treatment and comparison groups. The evaluator should provide a detailed description of the appropriate statistical modeling approach and model specifications, including hypotheses and equations. The narrative description should include a discussion of plans for determining baseline equivalence, calculating minimum detectible impacts (MDI), handling missing data or non-response bias, dealing with clustering, and addressing multi-site evaluations, if applicable. The evaluator should also discuss the inclusion or exclusion criteria to be used for selecting covariates (i.e., independent variables).

Questions to Answer

- How will the data be analyzed? Is the data analysis plan appropriate for the type of data collected, for the evaluation design, and for answering the research questions?
- For impact studies:
  - What statistical analysis methods will be used? Are model specifications provided and sufficiently described?
  - What are the decision rules for the inclusion and exclusion of covariates?
  - What is the plan for calculating effect size?
  - If applicable, how will the analysis address the issue of multiple comparisons?
- For subgroup analysis, is the subgroup defined and an appropriate sample size? What outcomes will be examined?

Model Specifications (for Impact Study Only)

The evaluator should provide the model specifications (e.g. "Greek models") for each of the planned statistical analyses. The model specifications should include terms representing the dependent variable (i.e., outcome variable), treatment indicator (i.e., whether the observation or sample member is part of the treatment or comparison group), and covariates. The estimate related with the treatment indicator is the estimate of program impact on the outcome of interest. In addition to the Greek model, the EDR should include a narrative description that explains the variables and key parameters.

Baseline Equivalence Testing (for Impact Study Only)

The EDR should describe the plan for baseline equivalence testing. This plan should include the methods, models, and variables to be used. Tests of baseline equivalence allow evaluators to confirm that the characteristics of the treatment group and comparison group were similar (or "equivalent") at baseline. If the groups are equivalent at baseline, then any differences in outcomes measured at follow-up can be reasonably attributed to the intervention. As such, it is important to demonstrate baseline equivalence when conducting impact studies.

Minimum Detectable Impacts (for Impact Study Only)

The expected power based on the sample size of the study should be provided. A power calculation provides an estimate, given a specific sample size and analysis design, of how likely it is that a program effect detected will be statistically significant (i.e., not due to chance). Power analysis involves determining the minimum detectable impacts (MDIs). The MDI is the smallest true impact that the study has a high probability of detecting. In other words, the MDI allows the evaluator to

know much effect the intervention must have on an outcome of interest (e.g., earnings, employment) for the impact of the intervention to be detected with a given sample size and specified probability of error.

Multiple Comparisons Problem (for Impact Study Only)

The EDR should discuss strategies for addressing the problem of "multiple comparisons" if the study will be estimating a large number of outcomes. The multiple comparisons problem refers to the likelihood that statistical significance may be detected by chance due to testing for many outcomes. Common strategies for addressing this potential problem include *p*-value adjustments (e.g., Benjamini-Hochberg correction, Bonferroni correction) and concentrating on one confirmatory outcome for each research question.

Subgroup Analyses (for Outcomes and Impact Studies Only)

If subgroup analyses are planned, the evaluator should provide a description of them, including the rationale for the analyses and subgroup selection. The description should identify the subgroup (e.g., women, compared with men), outcomes of interest for the subgroup analysis, and subgroup sample size.

For impact studies, the evaluator should discuss the power needed to detect subgroup impacts, to help ensure that the sample size will be adequate to provide robust findings. The EDR should include the statistical model that will be used for the subgroup impact analysis or describe how the model will be different from the model used for the overall analysis. Additionally, the evaluator should discuss plans for testing baseline equivalence for the subgroups.

5. Risks and Challenges

The EDR must include a discussion of the limitations of the study and how they may affect interpretation of the study findings. The primary source of study limitations are threats to validity, which can be due to the intervention's implementation or the design of the evaluation. The EDR should clearly describe each threat, including how the threat may influence study findings and any plans for mitigating the threat. Additionally, the evaluator should discuss procedures for protecting human subjects.

Threats to Validity

The two types of validity that evaluators need to be concerned about are internal validity and external validity.

Internal validity refers to the extent to which the changes in the outcomes of interest are attributable to the intervention. Common threats to internal validity include:

- Selection bias, which occurs when the characteristics of claimants selected to participate in the study have a direct effect on the outcomes being measured.
- Non-response bias, which occurs when claimants in the study do not respond to surveys in such a way that results in notable differences between those who complete the survey and those who do not complete the survey.
- History, which occurs when unknown interventions may be affecting the treatment group members during the time of the evaluation.

- Maturation, which occurs when the outcomes observed are the result of natural changes occurring over time rather the intervention.
- Attrition, which occurs when claimants leave the study.

External validity refers to the extent to which the study results can be applied (or "generalized") to the overall population.

In the EDR, the evaluator should clearly discuss all threats related to the study design and strategies for addressing them. The evaluator should also discuss any implications the threats have for the study findings, including how the threats might influence results and how the evaluator will interpret the results in light of the threats.

Questions to Answer

- What are potential internal and external threats to validity?
- What strategies will be implemented to mitigate the threats?
- What are the implications of the threats for the study findings and their interpretation?

Human Subjects Protections and Data Security

In the EDR, the evaluator should indicate that the evaluation will be reviewed and approved by an Institutional Review Board (IRB) before study activities start. Studies that involve collecting information from and about individual people are considered research involving human subjects. As evaluators will be conducting human subjects research, it is their responsibility to obtain IRB review and approval of the study research design and methods. As part of the IRB process, the evaluator will need to provide information about the study, the type of information to be collected, how the information will be used, and security protocols to be in place to protect the information.

Depending on the data to be collected, the evaluator may need to obtain informed consent from potential study participants. This would involve the evaluator informing an eligible claimant about the study, how the person's data will be used, and the risks and benefits of participating in the study; and the claimant then granting permission for the information to be collected and shared with the evaluation team. If informed consent is required, the EDR should describe the procedure for obtaining it.

Evaluations of RESEA programs will most likely entail collecting personally identifiable information (PII) about claimants (e.g., Social Security number, date of birth). The EDR should include the evaluator's plans and protocols for maintaining the security of all data collected, but especially claimants' PII.

6. Logistics

The EDR should provide an overview of the evaluation logistics related to how and when findings will be reported as well as the overall schedule of evaluation activities. Discussing these details in the EDR will help ensure that state RESEA program administrators and the evaluator are on the same page about evaluation timing.

Reporting

The EDR should discuss what topics will be covered in all planned reports, such as interim and final reports. Any publication or release plans should also be noted.

Schedule

The EDR should include a discussion of the timing of critical evaluation activities, such as implementation of data collection procedures and development and dissemination of reports. Including this information in the EDR will help to make sure that evaluation activities are aligned with intervention activities.

The EDR should include a detailed, year-by-year (even month-by-month) work plan for all planned evaluation activities. This should include the time period when evaluation activities such as data collection will take place. The evaluator should make sure to allow ample time in the schedule to draft and refine the evaluation design and develop research protocols and data analysis plans. Additionally, the schedule should include information about specific deliverables, such as any reports and briefings. The schedule should include information about when drafts and final versions of reports will be submitted, as well as the time frame for reviews by RESEA program administrators and the evaluator's revisions.

Resources

- **"What Evaluation Details Do I Need to Plan for and How Long Will It Take?":** This webinar prepares states to plan evaluation activities from start to finish. Attendees will explore common activities at each stage of the evaluation. https://www.workforcegps.org/events/2019/05/30/13/54/What-Evaluation-Details-Do-I-Need-for-a-Plan-and-How-Long-Will-It-Take

Appendix J. 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.

administrative data: Information that is regularly collected as a part of the regular administration of program or intervention activities

attrition: Loss of subjects from the study sample over the course of the evaluation. There may be many causes for participant 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: The group of study participants whose experiences and outcomes are compared to the **intervention group**'s. In an experimental study, the comparison group is exposed either to a different intervention or to no intervention. (A no-intervention comparison group is called a "control group"; a control group is created by random assignment.)

component: Unique services, benefits, or strategies that are bundled to comprise a state's RESEA program.

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).

evaluation: Broadly used to include the systematic and objective study of workforce programs, systems, strategies, services, activities, or interventions.

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.

human subject: A living individual about whom an investigator (whether professional or student) is conducting research: (1) Obtains information or biospecimens through intervention or interaction with the individual, and uses, studies, or analyzes the information or biospecimens; or (2) Obtains, uses, studies, or analyzes, or generates identifiable private information or identifiable biospecimens. (Source: Code of Federal Regulations 29CFR 21.102(e)(1).)

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 may inform future program development or replication.

impact study: An impact study estimates the changes in individual claimant outcomes that are solely attributable to a specific intervention, program, or policy rather than characteristics of the claimants themselves. Impact studies determine whether a program or policy has its intended impact—that is, whether the program or policy causes a difference in claimant outcomes for which it is designed to influence.

informed consent: Process of providing an individual and their legally authorized representative information (as required by federal and local laws) for making a voluntary and informed decision whether or not to participate in a study.

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.

intervention group: The group in a study that receives the intervention being tested (versus the **comparison group**). Also called the "treatment" group.

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.

long-term outcome: Change in behavior, attitudes, aptitudes, skills, and knowledge for staff, participants, environments, or larger systems; and significant change in employment and earnings, employment retention, and receipt of credentials for workforce programs.

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.

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

outcomes 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 outcome 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).

personally identifiable information: Information 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. PII includes names, Social Security numbers, birthdates, addresses, and other related contact information.

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 outcomes study where behavior before a program (or a subject's participation in it) began (pre-program) is compared to behavior at a point after the program was completed (post-program).

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

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

quasi-experimental design (QED) 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. QEDs attempt to approximate an experimental design by using a comparison group, but they do not use random assignment to create a control group of study members who are identical to the intervention group.

randomized controlled trial (RCT) study: A research design in which participants are randomly assigned by lottery to an intervention group that receives services or a control group that does not receive services (or to one of two or more intervention groups). The difference between the average outcome for the intervention 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 selected to represent all of the respective and potentially relevant factors to the study and its outcomes.

sample: A subset of a larger population 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.

statement of work: A document that provides a high-level description of the evaluation tasks, responsibilities, and deliverables that an evaluator will be responsible for carrying out.

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

theory of change: A theory of change is a way to explain your underlying understanding of the issue you are addressing—it clarifies why you are doing what you are doing. It is a description of a program that includes a clear identification of the intended population, as well as, the theoretical basis or description of the expected causal mechanisms by which the intervention should work. Theories of change are program model representations used to engage stakeholders about proposed research.

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.