COLLABORATION FOR EFFECTIVE EDUCATOR DEVELOPMENT, ACCOUNTABILITY, AND REFORM CENTER

FINAL CENTER EVALUATION PLAN

PREPARED FOR
Division of Research to Practice
Office of Special Education Programs
U.S. Department of Education

PREPARED BY
Collaboration for Effective Educator Development, Accountability, and Reform Center

July 12, 2013
# CONTENTS

## INTRODUCTION ........................................................................................................... 1
  CENTER MISSION, GOALS, AND ACTIVITIES ..................................................................... 1
  CONCEPTUAL APPROACH .................................................................................................. 5

## FORMATIVE EVALUATION DESIGN .............................................................................. 9
  FORMATIVE EVALUATION APPROACH ............................................................................. 9
  FORMATIVE EVALUATION QUESTIONS ............................................................................ 10
  FORMATIVE EVALUATION SOURCES AND COHORTS .......................................................... 13
  FORMATIVE DATA COLLECTION METHODOLOGIES .......................................................... 13
  FORMATIVE DATA ANALYSIS APPROACHES ................................................................... 18
  FORMATIVE PERFORMANCE MEASURES ......................................................................... 19

## SUMMATIVE EVALUATION DESIGN ............................................................................ 28
  SUMMATIVE EVALUATION APPROACH ........................................................................... 28
  SUMMATIVE EVALUATION QUESTIONS ........................................................................... 29

## PLANS AND ANTICIPATED TIMELINE FOR PREPARING CENTER EVALUATION REPORTS .................................................................................................................. 34

## REFERENCES .................................................................................................................. 35

### LIST OF EXHIBITS

Exhibit 1: Collaboration for Effective Educator Development, Accountability, and Reform Center Logic Model .......................................................................................................................... 7

Exhibit 2: Data Sources by Service Area, Year, and Method of Data Collection .......... 16

Exhibit 3. Formative Performance Measures by Activity and Data Source ............... 20

Exhibit 4. Summative Evaluation Questions ................................................................ 31
INTRODUCTION

This report describes the evaluation plan for major activities being conducted by the Collaboration for Effective Educator Development, Accountability, and Reform Center (the Center). The Center’s evaluation focuses on both processes and outcomes. The evaluation addresses the processes through which the Center’s integrated program of products and services are being planned and implemented and the extent to which this program promotes its intended outcomes. The findings from this evaluation serve formative and summative purposes. Formatively, the evaluation will summarize opinions about the quality, relevance, and usefulness of Center products and services. These findings will be used toward the ongoing improvement of the Center. Thus, formative findings will be emphasized in the first four years of the Center. In the fifth year, the evaluation will be oriented toward a summative purpose—namely, to understand the value or worth of the Center in relation to its mission and goals.

The evaluation plan outlines the key components of the ongoing evaluation, including the conceptual approach, the evaluation design, and the methods and procedures being used to evaluate Center products and services throughout its five years of operation (January 1, 2013–December 31, 2017). This report serves as a planning document for conducting the annual evaluation of the Center over its five years of operation. Please note, however, that the evaluation team expects to modify these plans—as necessary and appropriate—in response to annual findings and recommendations derived from the ongoing evaluation activities.

This report begins with an introduction that presents the Center's mission, goals, and activities, followed by the Center's conceptual approach. The formative and summative evaluation designs are described separately, including the data collection sources and methodologies. The final section discusses the anticipated timeline for preparing Center evaluation reports.

CENTER MISSION, GOALS, AND ACTIVITIES

The Center is funded by the U.S. Department of Education, Office of Special Education Programs (OSEP) under a grant awarded to the University of Florida. Key personnel at the University of Florida, directed by Mary Brownell, in conjunction with the Center’s partners, are conducting the ongoing work of the Center and its major activities. The University of Florida will lead the evaluation process, with the American Institutes for Research leading the formative evaluation and the University of Texas at Austin leading the summative evaluation. The products and services developed by the Center are anticipated to have an impact at the national, state, and local levels.

The Center is funded under OSEP’s Technical Assistance and Dissemination (TA&D) program and has developed a mission statement consistent with OSEP’s goals for this program. By building and sustaining an aligned, coherent, and systemic approach to educator preparation, the Center will fulfill its mission to promote, support, and reinforce teacher and leader effectiveness across the career continuum and ensure that students with disabilities achieve college and career readiness standards in inclusive settings. In order to achieve this mission, the Center has three overarching goals, as specified in the cooperative agreement:
• **Goal 1:** State education agencies (SEAs) review and reform certification or licensure standards, in collaboration with teacher and leader preparation programs.

• **Goal 2:** Institutions of higher education (IHEs) and local education agencies (LEAs) restructure and improve teacher and leader preparation programs, in collaboration with SEAs.

• **Goal 3:** SEAs, IHEs, and LEAs develop data systems that will be used to inform special education teacher preparation programs improvement.

In order to achieve these overarching goals, the Center will achieve three specific objectives and successfully conduct activities for each objective, as described below.

**Objective 1: Knowledge Development**

To identify policy and practices that support SEAs’ evaluation and reform of educator licensure standards, facilitate the integration of evidence-based practices (EBP) in the content of teacher and leader professional learning systems, and describe how K-12 SWD outcome data may be used validly in the evaluation of professional learning systems.

To achieve Objective 1, the Center will successfully conduct the following five activities:

• **Activity 1.1.** Conduct and disseminate an analysis of state policies related to teacher and leader licensure standards, program accreditation, and program evaluation (with attention to the use of K-12 student data).

• **Activity 1.2.** Develop and disseminate (nine content and three pedagogy) innovation configurations (as tools for use in integrating EBP into the content of teacher and leader preparation).

• **Activity 1.3.** Develop and disseminate (five) literature syntheses and related reform rubrics on restructuring and improving teacher preparation, restructuring and improving leader preparation, teacher preparation program evaluation (including the use of K-12 student data), leader preparation program evaluation (including the use of K-12 student data) and systems change and reform.

• **Activity 1.4.** Identify and disseminate effective strategies for restructuring preparation, fostering collaboration, improving training, integrating EBP into program content, and evaluating program outcomes, based on analysis of successful 325T projects.

• **Activity 1.5.** Develop model needs assessments for SEAs, on the one hand, and IHEs, LEAs, and non-profits, on the other, and TA plans for restructuring and redesigning teacher and leader professional learning systems, for use in intensive states.

**Objective 2: Technical Assistance**
To offer a comprehensive program of universal, targeted, and intensive Technical Assistance (TA), as requested and appropriate, to key stakeholders in states and localities across the nation. To achieve Objective 2, the Center will achieve three sub-objectives (each with discrete activities), as described below.

**Objective 2.A.: Universal TA**

To increase awareness for key stakeholders in SEAs, IHEs, LEAs, and nonprofits in revising standards, reforming teacher and leader professional learning systems, and refining program evaluation processes that use K-12 SWD outcome data.

To achieve Objective 2.A., the center will successfully conduct the following two activities:

- **Activity 2.A.1.** Develop and maintain a fully accessible website, linked to TACC, with an online library to archive content and pedagogy ICs, research syntheses, reform rubrics, briefs, training manuals, online newsletters, webinars, podcasts, ask the expert videos, and other TA tools and products on the Center’s website.

- **Activity 2.A.2.** Develop and maintain interactive website features, including keyword search function, online chats, and discussion boards.

**Objective 2.B.: Targeted TA**

To improve readiness of select states, IHEs, LEAs, and nonprofits in revising standards, reforming teacher and leader professional learning systems, and refining program evaluation processes that use K-12 SWD outcome data.

To achieve Objective 2.B., the Center will successfully conduct the following three activities:

- **Activity 2.B.1.** Respond to TA requests from SEAs, IHEs, and professional organizations by identifying existing Center resources and consultative help for assistance in revising standards; and reforming, restructuring, or evaluating professional learning systems.

- **Activity 2.B.2.** Respond to TA requests from SEAs and IHEs seeking intensive TA.

- **Activity 2.B.3.** Continue targeted TA supports for 325T projects by coordinating with NCIPP.

**Objective 2.C.: Intensive TA**

To improve capacity across selected state level teams (SEA, IHE, LEAs) in revising standards, reforming teacher and leader professional learning systems, and refining program evaluation processes that use K-12 SWD outcome data.
To achieve Objective 2.C., eight intensive TA activities will be provided to SEAs and IHEs, as listed below.¹

**SEA TA & D Activities**

- **Activity 2.C.1.** Recruit and select at least five states annually for intensive TA, garner the required commitments from applicant SEAs and IHEs, and negotiate memoranda of understanding.

- **Activity 2.C.2.** Assist states in establishing the SEA leadership team.

- **Activity 2.C.3.** Assist SEA leadership team in conducting a needs assessment and developing comprehensive TA plans for revising teacher and leader standards, reforming teacher and leader preparation, and improving program evaluation.

- **Activity 2.C.4.** Assist the SEA leadership team in developing an evaluation plan to assess improvements in teacher and leader program content, pedagogy, and evaluation.

- **Activity 2.C.5.** Develop and maintain a password-protected, web-based system for SEAs and IHEs participating in intensive TA to share information and resources.

**IHE TA & D Activities**

- **Activity 2.C.6.** Assist SEA leadership team in identifying leadership teams at each (of three) participating IHEs.

- **Activity 2.C.7.** Assist IHE leadership teams in developing TA plans for reforming teacher and leader preparation and evaluating program outcomes.

- **Activity 2.C.8.** Develop customized tools and resources, provide expert consultants, and offer need-based training in onsite and online formats consistent with SEA and IHE TA plans in intensive states.

**Objective 3: Leadership and Coordination**

To increase awareness of, and establish the Center as a recognized source for information and resources about, standards revision, preparation reform, and program evaluation development by (a) establishing partnerships with professional organizations and other key stakeholders and (b) developing and supporting communities of stakeholders involved in related revision, reform, and development.

To achieve Objective 3, the Center will successfully conduct the following four activities:

- **Activity 3.1.** Establish collaborative partnerships (among partner organizations).

¹ Please note that Activities 2.C.5 and 2.C.8 are cross-cutting activities that involve both SEAs and IHEs.
**Activity 3.2.** Consult with stakeholders (including senior advisors and the project advisory board).

**Activity 3.3.** Coordinate with other OSEP-funded Centers and TACC on product development and dissemination and maintain communication and coordination with OSEP Project Officers.

**Activity 3.4.** Develop and sustain networked improvement communities (NICs) among stakeholder groups involved in intensive TA.

The following section describes the Center's conceptual approach to evaluating its progress toward conducting these activities and accomplishing its goals.

**Conceptual Approach**

The Center uses a logic model to frame its conceptual approach and help guide its evaluation of Center products and services. The mental health and social services fields have used logic models extensively to conceptualize large-scale evaluations of newly introduced programs (Hatry & Kopczynski, 1997). In general, logic models attempt to provide planning and implementation structures to evaluation efforts by outlining the hypothetical causal relationships among resource inputs, implementation processes, and intended outcomes (Orwin & Goldman, 1996). OSEP used a logic model (Schiller, McInerney, & Kane, 1996) to depict causal relationships among Parts B, C, and D of the Individuals with Disabilities Education Act (IDEA). The logic model for IDEA became the basis for developing Part D performance indicators, as required by the Government Performance and Results Act (DeStefano et al., 2004). Logic models have also been used to guide the evaluations of other IDEA Part D national TA centers, such as the National Center to Inform Policy and Practice in Special Education Professional Development (2009), the Elementary and Middle School Technical Assistance Center (Mesmer, Hamilton, & McInerney, 1998), the Access Center: Improving Outcomes for All Students K–8 (Davis et al., 2003), the National Center on Student Progress Monitoring (Kim & McInerney, 2004), and the National Center on Response to Intervention (Poirier, Brown, Rose, Shami, & McInerney, 2008).

Exhibit 1 displays a schematic illustrating the logic model for the evaluation of the Center. Overall, the logic model illustrates the potential causal relationships among the Center’s planned activities and outcomes. The left side of the schematic lists the Center’s three goals. The Inputs column describes the resources the Center will draw upon to reach its goals. The Outputs column lists the Center’s Activities, Participants, and Products. The Activities are indexed to Center Objectives as described and numbered above. Each activity is associated with certain groups of Participants, which in turn are associated with specific Products. These Outputs will support the Center’s intended Outcomes, shown in the right column. Different short-term outcomes are expected from each main activity, and these short-term outcomes will lead to the intermediate outcomes and, in turn, the long-term outcomes. With respect to intensive TA participants, short-
term outcomes follow from the planning and training in the first year of participation, intermediate outcomes follow from the implementation of plans in the second year of participation, and long-term outcomes follow from efforts to sustain and scale up the program in the subsequent two to three years. Thus, we can observe the full scope of the logic model for the first two cohorts of participants (i.e., those that start in the first two years of the Center). The formative evaluation provides feedback on the Center’s outputs, short-term outcomes, and intermediate outcomes. The findings from the formative evaluation will be used to assess specific performance measures (described on page 18). The summative evaluation addresses the intermediate and long-term outcomes.
Exhibit 1. Collaboration for Effective Educator Development, Accountability, and Reform Center Logic Model

Major Components of CEEDAR Logic Model

**GOALS**

Goal 1: SEAs review and reform certification or licensure standards, in collaboration with teacher and leader preparation programs.

Goal 2: IHEs and LEAs restructure and improve teacher and leader preparation programs, in collaboration with SEAs.

Goal 3: SEAs, IHEs, and LEAs develop data systems that will be used to inform special education teacher preparation programs improvement.

**INPUTS**

- Cooperative Agreement
- Key Personnel at UF, AIR, and KU
- Organizational Partners
- Senior Advisors and Consultants
- Staff
- UF, AIR, and KU institutional resources Space, furniture, utilities, technology, and tech support

**OUTPUTS**

- Activities
- Participants
- Products

1. 1-5. Knowledge Production

2. Technical Assistance

2A.1.2 Universal

2B.1.3 Targeted

2C.1.8 Intensive

3.1-4 Leadership and Communication

**OUTCOMES**

- Short-Term
- Intermediate
- Long-Term

- SEAs, LEAs, and IHEs increase awareness of CEEDAR TA tools and how to use them

- SEAs, LEAs, and IHEs increase knowledge of how to initiate and sustain successful preparation systems reform

- SEAs, LEAs, and IHEs increase knowledge of how to initiate and sustain successful preparation systems reform

- SEAs make systemic changes in teacher and leadership personnel development

- SEAs align licensure, accreditation, and evaluation policy

- States build their capacity to scale up TE and LE reform

- SEAs achieve EBP content in TE and LE programs

- SEAs improve TE and LE pedagogy

- Graduates of reformed TE programs implement EBP with fidelity

- Graduates of reformed LE programs create responsive school contexts

- SWDs achieve college and career ready standards

**Note:** Products of Leadership and Communication activities lead to the top-most short term outcome (‘SEAs, LEAs, and IHEs increase awareness of CEEDAR TA tools and how to use them’).
Products from SEA and IHE Activities (fig. 1)

Products

Products from SEA Activities
- Policy self-assessment, TA plan and other products resulting from state leadership team activities
- State leadership team workshops and other TA products on conducting licensure, accreditation, and evaluation policy analysis
- Reformed state licensure & certification standards
- Improved SEA teacher and leader evaluation systems that use SWD outcome data

Products from IHE Activities
- TE and LE needs assessments, TA plan, and other products resulting from state leadership team activities
- IHE faculty workshops and other TA products on conducting content and pedagogy IC analyses
- IHE faculty workshops and other TA products on implementing and teaching EBPs
- Reformed IHE courses & syllabi
- Improved IHE teacher and leader evaluation systems that use SWD outcome data
**FORMATIVE EVALUATION DESIGN**

The Center’s formative evaluation will collect and analyze quantitative and qualitative data that are directly relevant to OSEP’s program performance measures for its TA&D program (i.e., quality, relevance, and use). The Center will use these data to foster the Center’s continuous improvement. This section discusses the overall formative evaluation approach, evaluation questions by activity, evaluation sources and cohorts, and methods for data collection and analysis. The section concludes with a discussion of how the evaluation findings will be used to create program and project performance measures.

**FORMATIVE EVALUATION APPROACH**

The formative evaluation focuses on Center formative processes and short-term and intermediate outcomes.

**Center processes**, the outputs described in the logic model are the combined efforts of Center staff and partner schools and organizations to produce products and deliver services. The Center is concerned, first, with producing products and delivering services that are high quality and relevant to the needs of the intended audiences. The Center also is concerned with whether the intended audiences and other stakeholders use the products and services and find them helpful for improving educational policy and practice. Finally, the Center strives to be cost efficient across its activities. The evaluation of Center processes will use several data sources:

- Customer surveys (annual and postevent)
- Follow-up interviews with a sample of survey respondents and other stakeholders
- Extant data to indicate Center outputs and level of use of products, including website analytics
- Extant data on Center expenditures

**Formative center outcomes** are the changes observed for the direct users of products or recipients of services. In general, the evaluation will assess the extent to which the Center has effectively built capacity to support and implement reformed education preparation systems by observing changes in the following:

- *Awareness* (e.g., practitioners learn where to find Center TA tools and how to use them)
- *Knowledge and skills* (e.g., increased knowledge of how to initiate and sustain successful preparation systems reform)
- *Application of EBP to personnel preparation* (e.g., SEAs begin to make systematic changes in standards for teacher and leader licensure and certification; and IHEs begin to make systematic changes in teacher and leader preparation programs, including course syllabi and curricula)

---

2 The summative evaluation will address changes observed for indirect participants such as graduates of reformed preparation programs and the students they serve.
For example, in states receiving intensive TA, a Memorandum of Understanding (MOU) between the recipient and the Center will guide the creation of a work plan and a timeline with milestones for completing the work. These products will assist in evaluating the recipient’s capacity to reform educator preparation programs. As another example, annual surveys of the Center’s consumers can be used to monitor the extent to which Web-based products are helping policymakers, administrators, educators, and other appropriate stakeholders reform preparation programs and implement EBP with fidelity.

The formative evaluation is grounded in the conceptual linking of the outputs of center activities to short-term outcomes (e.g., within the first year of intensive TA participation) and intermediate outcomes (e.g., the second year of participation), as depicted in the logic model.

Formative outcomes will be evaluated using several approaches, including—

- Self-reported survey ratings by Center consumers indicating that their attitudes, knowledge, and behaviors toward preparation systems reform have changed after receiving and using Center products and services
- Indicators of change in educator practice as measured by capacity rubrics (based on the work plans and timelines created for intensive TA states)

**Formative Evaluation Questions**

This section summarizes the formative evaluation questions for each objective and its associated activities, as described and numbered above. These questions address Center processes and anticipated short-term outcomes. Additional questions related to intensive TA also address intermediate and long-term outcomes, with formative data being used in the summative evaluation design (see later sections). The evaluation questions related to cost efficiency and long-term outcomes cut across all activities and are presented separately.

**Objective 1: Knowledge Development.**

1. To what extent did a panel of reviewers consider knowledge development products to be of high quality?
2. To what extent did a panel of reviewers consider knowledge development products to be of high relevance?
3. To what extent did a panel of reviewers consider knowledge development products to be of high usefulness?
4. To what extent did consumers perceive knowledge development products created by the Center (e.g. ICs, literature syntheses, and other products and reports) to be of high quality?
5. To what extent did consumers of knowledge development products perceive the products to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?
6. To what extent did consumers perceive knowledge development products to be useful?
7. To what extent did knowledge development and other center products enhance consumer knowledge of, awareness of, and application of EBP to effective teacher and leader preparation programs?

**Objective 2: Technical Assistance.**

- **Objective 2.A. Universal TA**
  1. To what extent did consumers of universal TA perceive the Center’s universally available online products and other TA materials to be of high quality?
  2. To what extent did consumers of universal TA perceive the Center’s website and online products, and other TA materials to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?
  3. To what extent did consumers of universal TA perceive the Center’s universally available online products and other TA materials to be useful?
  4. To what extent did the Center’s universal TA online products and other TA materials support teacher and leader preparation reform efforts?

- **Objective 2.B. Targeted TA**
  1. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be of high quality?
  2. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?
  3. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be useful?
  4. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be of high quality for preparing applications for intensive TA, when appropriate?
  5. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be relevant to preparing applications for intensive TA, when appropriate?
  6. To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be useful in preparing applications for intensive TA, when appropriate?
  7. To what extent did the Center’s targeted TA resources and tools support teacher and leader preparation reform efforts?
• **Objective 2.C. Intensive TA**

1. To what extent was the MOU and its resulting blueprint implemented according to timeline in each intensive state?
2. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be of high quality?
3. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?
4. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be useful?
5. To what extent are participating IHE preparation programs’ practices effective in building knowledge and skills of teacher and leader candidates?
6. To what extent have the Center’s TA services increased intensive states’ capacity to scale up and sustain reform?
7. To what extent do SEAs align licensure, accreditation, and evaluation policy?
8. To what extent do IHEs infuse EBP content in TE and LE program?
9. To what extent do IHEs align pedagogy with EBP?
10. To what extent do graduates of reformed TE and LE programs implement EBP?
11. To what extent have SEAs and IHEs begun designing, piloting, and/or implementing improved teacher and leader evaluation systems that use SWD outcome data?

**Objective 3: Leadership and Coordination.**

1. To what extent did the Center establish collaborative partnerships and consult with professional organizations and individuals interested in personnel preparation reform?
2. To what extent did NIC participants perceive the Center’s products and services to be of high quality?
3. To what extent did NIC participants perceive the Center’s products and services to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?
4. To what extent did NIC participants perceive the Center’s products and services to be useful?

---

3 Evaluation questions 8-12 of Objective 2.C. address our progress towards achieving the Center's three goals, and thus address intermediate and long-term outcomes. Data will be reviewed formatively to improve Center services, as well as summatively as part of the Center's final report.
5. To what extent did the Center coordinate with other OSEP-funded Centers and the TACC on product development and dissemination and maintain communication and coordination with OSEP Project Officers?

Two evaluation questions cut across objectives:

1. What is the cost per unit of TA provided by the Center?
2. To what extent has the Center accomplished its planned work scope in a timely manner?

**Formative Evaluation Sources and Cohorts**

The formative evaluation will address these evaluation questions with a series of data collection efforts targeting the recipients of the products and activities of each Center Activity:

- **Objective 1, Activities 1.1-5:** developers and consumers of Knowledge Development products
- **Objective 2:**
  - Activities 2.A.1-2: consumers of universal TA products
  - Activities 2.B.1-3: recipients of targeted TA services and products
  - Activities 2.C.1-8: recipients of intensive TA services and products
- **Objective 3, Activities 3.1-3.4:** participants in Leadership and Coordination partnerships and activities

Progress toward intermediate and long-term outcomes will be depicted using a cohort-based design (e.g., Baltes, Cornelius, & Nesselroade, 1979), where each cohort of intensive TA participants will be tracked from its time of program entry through the last year of Center operations. This will permit us to observe whether the early cohorts (i.e., those joining in years 1 and 2) sustain and scale up implementation, and whether ratings of quality, relevance, and usefulness show improvement from early to late cohorts.

**Formative Data Collection Methodologies**

The Center’s evaluation will use multiple data collection methods and sources to provide a comprehensive picture of performance for each activity, in terms of processes and outcomes. This section identifies the sources of evaluation data and describes data collection methods, as well as analytical procedures. Exhibit 2 displays the planned data collection activities by Center activities, sources, and year. The evaluation will use a multifaceted approach to collect evaluation data.

**Annual Surveys**

Annual surveys will be administered starting in year 1 to participants receiving intensive (or targeted) TA; users of universal TA resources; and participants in NICs. The survey will include forced-choice and multiple-choice items, and will address the following topics: Center resources or services accessed, level of satisfaction with these resources or services, and perceived level of
impact of the TA on teacher development and program improvement. The surveys will be administered via a Web link embedded in a personalized email and will take 10 to 15 minutes to complete. The Center will enlist the help of key partners (e.g., the state special education director) to encourage survey recipients to respond. Surveys will be psychometrically evaluated for reliability and validity using an item-response theory approach (Wright & Masters, 1982). Surveys will be analyzed descriptively and longitudinally by cohort.

**Postevent Surveys**
Postevent surveys will be administered in years 1–3 to collect immediate feedback on the quality, relevance, and usefulness of specific training and/or TA events (e.g., train-the-trainer sessions for SEA and LEA staff). These surveys will be used to make ongoing improvements to TA services. These surveys will be administered via paper and pencil to participants in face-to-face TA events, and via the Internet to participants in webinars. Approximately eight post event surveys will be administered each year (two per quarter). To provide frequent formative feedback to Center leadership, these findings will be reported quarterly.

**Stakeholder Interviews**
The Center will conduct semi-structured telephone interviews on an annual basis with a sample of participants receiving intensive TA. This sample will be stratified based on stakeholder group (SEA, IHE, LEA) and role (e.g., administrator or faculty member). The sample of users will be followed as a panel in years 1–4 of the Center. It is anticipated that this panel will consist of 15 individuals each from Cohorts 1 and 2; each panel will be followed for 3 years, to include the 2 years of intensive TA and the following year. This will provide in-depth feedback on different phases: planning, implementation, and scale-up (see logic model). These interviews will focus primarily on the outcomes aligned to each phase. Interviews will be recorded and transcribed prior to qualitative analysis.

**Capacity Rubric**
The Center will use a capacity rubric, developed from the MOU and work plan for each intensive TA state, to determine the impact of intensive TA on the SEA’s and IHE’s capacity to support innovations in personnel preparation and classroom practice. The rubric will include a review of SEA and IHE documents, as well as interviews with staff on respective leadership teams. This rubric will be deployed annually. The rubric will articulate the milestones for each participating organization (e.g., establishing policies for licensure, revising curricula of training programs, coordinating a network of trainers, monitoring of outcomes, etc.). Two internal evaluators will examine artifacts and documents to determine the ratings for each milestone.

**Tracking Logs**
Center staff responsible for each intensive TA state will be asked to complete logs to record the TA services provided to each state (e.g. meetings and workshops).

**Extant Data Sources**
The evaluation will capitalize on extant data generated from Center activities.
- Website usage data.
• Workshop attendance records
• Artifacts of Objective 1, Activities 1.1-5 products and activities (e.g. ICs)
• Artifacts of Objective 2, Activities 2.A.1-C.8 products and activities, including universal (archived webinar), targeted (workshop speaker or consultant referral), and intensive (workshop)
• Artifacts of Objective 3, Activities 3.1-4 activities and products (e.g. Center announcement in newsletter of national association partnerships)
• Other work products of the Center
## Exhibit 2: Data Sources by Service Area, Year, and Method of Data Collection

<table>
<thead>
<tr>
<th>Objectives / Activities</th>
<th>Data collection</th>
<th>Surveys and Interviews</th>
<th>Extant data and other instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective 1, Activities 1.1-5: Knowledge Development</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1. Develop and disseminate analyses of state policies.</td>
<td>Years 1-5</td>
<td>Web consumers, stakeholders developing analyses</td>
<td>Website analytics, analysis report feedback from peer reviewers</td>
</tr>
<tr>
<td>1.2. Develop and disseminate ICs</td>
<td>Years 1-5</td>
<td>Web consumers, stakeholders developing ICs</td>
<td>Website analytics, ICs, feedback from peer reviewers</td>
</tr>
<tr>
<td>1.3. Develop and disseminate literature syntheses</td>
<td>Years 1-5</td>
<td>Web consumers, stakeholders developing syntheses</td>
<td>Website analytics, syntheses feedback from peer reviewers</td>
</tr>
<tr>
<td>1.4. Identify and disseminate effective strategies used by successful 325T projects</td>
<td>Years 1-5</td>
<td>Web consumers, 325T grantees</td>
<td>Website analytics, feedback from 325T grantees, relevant 325T reports</td>
</tr>
<tr>
<td>1.5. Develop model needs assessments and TA plans</td>
<td>Years 1-5</td>
<td>Web consumers, stakeholders developing assessments and plans</td>
<td>Website analytics, model assessments and plans, feedback from peer reviewers</td>
</tr>
<tr>
<td><strong>Objective 2.A., Activities 2.A.1-2: Universal TA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.A.1. Develop and maintain online website library</td>
<td>Years 1-5</td>
<td>Web consumers</td>
<td>Website analytics</td>
</tr>
<tr>
<td>2.A.2. Develop and maintain interactive website features</td>
<td>Years 1-5</td>
<td>Web consumers</td>
<td>Website analytics</td>
</tr>
<tr>
<td><strong>Objective 2.B., Activities 2.B.1-3: Targeted TA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.B.1. Respond to targeted TA requests from SEAs, IHEs, and professional organizations for assistance</td>
<td>Years 1-5</td>
<td>Targeted TA recipients</td>
<td>Records of requests and responses</td>
</tr>
<tr>
<td>2.B.2. Respond to targeted TA requests from SEAs and IHEs seeking intensive TA</td>
<td>Years 1-5</td>
<td>Targeted TA recipients</td>
<td>Records of requests and responses</td>
</tr>
<tr>
<td>2.B.3. Continue NCIPP targeted supports for 325T projects</td>
<td>Years 1-5</td>
<td>Targeted TA recipients</td>
<td>Records of requests and responses</td>
</tr>
<tr>
<td><strong>Objective 2.C., Activities 2.C.1-8: Intensive TA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.C.1. Recruit and select states for intensive TA</td>
<td>Years 1-4</td>
<td></td>
<td>National call, artifacts of application and selection</td>
</tr>
<tr>
<td>Objective 3, Activities 3.1-4: Leadership and Coordination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1. Establish collaborative partnerships</td>
<td>Years 1-5</td>
<td>Partner organization staff</td>
<td>Artifacts of partnerships</td>
</tr>
<tr>
<td>3.2. Consult with stakeholders</td>
<td>Years 1-5</td>
<td>Stakeholders</td>
<td>Artifacts of partnerships</td>
</tr>
<tr>
<td>3.3. Coordinate with other OSEP-funded Centers and TACC and maintain communication and coordination with OSEP Project Officers</td>
<td>Years 1-5</td>
<td>N/A</td>
<td>Artifacts of coordination and communication</td>
</tr>
<tr>
<td>3.4. Develop and sustain NICs</td>
<td>Years 1-5</td>
<td>NIC participants</td>
<td>Meeting artifacts, artifacts of partnerships, website analytics</td>
</tr>
</tbody>
</table>
FORMATIVE DATA ANALYSIS APPROACHES

This section describes the methodology for analyzing quantitative and qualitative data. The American Institutes for Research will analyze data using appropriate quantitative statistical and qualitative ethnographic techniques and will present the results in an annual evaluation report.

Evaluation of Surveys

To ensure that the evaluation findings and recommendations are based on valid and reliable data, a psychometric validation will be conducted for the Web-based surveys. The Rasch model for ordered categories (Andrich, 1978; Rasch, 1980; Wright & Masters, 1982) will be the primary method used for the item analyses. A psychometric validation allows an evaluator to create valid scale scores on latent traits (e.g., professional development quality) by evaluating all the measurement properties of the instrument for construct and content validity. These scale scores—which are based on multiple items that fit together from a theoretical perspective—give a quantitative view of frequency and intensity of an individual’s responses.

Quantitative Analysis

Prior to conducting any analyses of quantitative data, evaluation staff will clean data and recode certain variables to make the analyses and interpretation easier. Descriptive statistics will provide a snapshot of the data, enabling the researchers to determine which inferential analyses may be most appropriate, given the distribution of different variables or fluctuations in the data. The evaluation team will also conduct analyses to explore relationships between respondents’ characteristics and their response patterns. For example, it is possible to observe if fidelity of implementation is lower for teachers in certain districts or certain types of schools. The reason for such relationships can be further probed through interviews or other exploratory analyses. Data will also be analyzed longitudinally by cohort to determine how outcomes develop across several years of experience with a product or service. The longitudinal analysis will employ descriptive approaches to identify trends over time.

Qualitative Analysis

Data from annual stakeholder interviews will be transcribed prior to analysis. The evaluation team will analyze qualitative data to explore and identify themes in concerns, experiences, and interests in issues related to improving teacher and leader effectiveness. These findings will be used to provide explanations for patterns observed in quantitative data. For example, if satisfaction with a particular product was low among certain stakeholders, qualitative data can be used to understand why that product is not meeting the needs of that stakeholder group.

Review of the Implemented Scope of Work

The evaluation will document progress made by the Center in implementing its work plan in each year. The evaluation team will review the following:

- The approved Center work plan
- Copies of deliverables, including research-based reports, implementation materials, product reports, and other resources produced by the Center; TA logs; and newsletters, e-mail updates, and media clippings
Based on these data sources, the evaluation team will conduct a detailed review, completed with the assistance of Center staff, of progress made on activities and deliverables in each year. The status of planned activities will be categorized as follows: completed (all activities completed, including ongoing and recurring activities such as product reviews); in development (started but not yet completed); delayed (postponed); or not started. The scope review will document the reasons for deliverables that are behind schedule (e.g., a change in priorities endorsed by OSEP).

**Formative Performance Measures**

The evaluation will collect data to address specific project performance measures. These measures will address the processes and outcomes related to key Center activities. These performance measures will reflect the evaluation questions and draw on the data sources and methods described in this evaluation plan. The specific targets for project performance measures will be finalized in consultation with OSEP. Exhibit 3 summarizes these proposed measures in alignment with objectives, evaluation questions, and data sources. As indicated in the table, certain project performance measures align with each of the OSEP program performance measures for the TA&D program. The program performance measures are listed below:

- **Program Performance Measure 1**: The percentage of products and services deemed to be of high quality by an independent review panel of qualified experts or individuals with appropriate expertise to review the substantive content of the products and services.

- **Program Performance Measure 2**: The percentage of products and services deemed to be of high relevance to teacher and leader effectiveness by an independent review panel of qualified members of the target audiences of the technical assistance and diseminations.

- **Program Performance Measure 3**: The percentage of products and services deemed to be of high usefulness by target audiences to improve teacher and leader effectiveness.

- **Program Performance Measure 4**: The federal cost per unit of technical assistance provided by the TA&D program.

- **Program Performance Measure 5**: The percentage of projects that achieve proposed milestones within the approved timeline.

---

4 In its continuation grant, the Center will report project data that is relevant to each of the program performance measures.
### Exhibit 3. Formative Performance Measures by Activity and Data Source

<table>
<thead>
<tr>
<th>Evaluation Questions</th>
<th>Project Performance Measures</th>
<th>Data Sources/Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective 1, Activities 1.1-5: Knowledge Development</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. To what extent did a panel of reviewers consider knowledge development products to be of high quality?</td>
<td><em>Addressing Program Performance Measure 1:</em> By December, 2013, a panel of expert reviewers will rate the quality of all knowledge development products (five research syntheses/reform rubrics, one state-by-state policy analysis, one 325T best practice review, twelve innovation configurations) with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>Survey of reviewers</td>
</tr>
<tr>
<td>2. To what extent did a panel of reviewers consider knowledge development products to be of high relevance?</td>
<td><em>Addressing Program Performance Measure 2:</em> By December, 2013, a panel of expert reviewers will rate the relevance of all knowledge development products (five research syntheses/reform rubrics, one state-by-state policy analysis, one 325T best practice review, twelve innovation configurations) with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>Survey of reviewers</td>
</tr>
<tr>
<td>3. To what extent did a panel of reviewers consider knowledge development products to be of high usefulness?</td>
<td><em>Addressing Program Performance Measure 3:</em> By December, 2013, a panel of expert reviewers will rate the usefulness of all knowledge development products (five research syntheses/reform rubrics, one state-by-state policy analysis, one 325T best practice review, twelve innovation configurations) with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>Survey of reviewers</td>
</tr>
<tr>
<td>4. To what extent did consumers perceive Knowledge Development products created by the Center (e.g. ICs, literature syntheses, and other products and reports) to be of high quality?</td>
<td><em>Addressing Program Performance Measure 1:</em> At the end of Years 1–5, a sample of website users representing the target audiences will rate the quality of the Center’s products (i.e., rigorous, based on up-to-date research, and clearly useful) with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>Web-consumer survey</td>
</tr>
</tbody>
</table>

---

5. Each product will be reviewed by three experts prior to dissemination. Reviewers will be asked to rate multiple parameters of product quality.

6. For all Likert-type scales on surveys, higher ratings represent more positive ratings. For ratings of quality, relevance, and usefulness, a possible scale is 1 = poor, 2 = fair, 3 = good, and 4 = excellent. A mean rating of 3.2 indicates that respondents rate the product or service to be at least “good,” on average, and requires that at least some respondents rate the product or service as “excellent.” For example, this target is met for the expert panel review if two reviewers rate the product with a 3 and one reviewer rates the product with a 4, resulting in a mean of 3.33.

7. Survey items for reviewers will be incorporated into the review form.
<table>
<thead>
<tr>
<th>Quality?</th>
<th>communicated) with a mean of at least 3.2 using a four-point Likert-type scale for items on a Web-consumer survey.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. To what extent did consumers of Knowledge Development products perceive the products to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?</td>
<td><strong>Addressing Program Performance Measure 2:</strong> At the end of Years 1–5, a sample of website users representing the target audiences will rate the relevance of the Center’s products to efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale for items on a Web-consumer survey.</td>
</tr>
<tr>
<td>6. To what extent did consumers perceive Knowledge Development products to be useful?</td>
<td><strong>Addressing Program Performance Measure 3:</strong> At the end of Years 1–5, a sample of website users representing the target audiences will rate the usefulness of the Center’s products to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale for items on a Web-consumer survey.</td>
</tr>
<tr>
<td>7. To what extent did Knowledge Development and other center products enhance consumer knowledge of, awareness of, and application of EBP to effective teacher and leader preparation programs?</td>
<td>At the end of Years 1–5, a sample of website users representing the target audiences will rate the degree of impact of the Center’s products on their knowledge of, awareness of, and application of EBP to effective teacher and leader preparation programs with a mean of at least 2 using a four-point Likert-type scale on a Web-consumer survey.</td>
</tr>
</tbody>
</table>

**Objective 2.A., Activities 2.A.1-2: Universal TA**

<table>
<thead>
<tr>
<th>1. To what extent did consumers of universal TA perceive the Center’s universally available online products and other TA materials to be of high quality?</th>
<th><strong>Addressing Program Performance Measure 1:</strong> At the end of Years 1–5, a sample of universal TA recipients will rate the quality of the Center’s products (i.e., rigorous, based on up-to-date research, and clearly communicated) with a mean of at least 3.2 using a four-point Likert-type scale for items on a Web-consumer survey.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. To what extent did consumers of universal TA perceive the Center’s website and online</td>
<td><strong>Addressing Program Performance Measure 2:</strong> At the end of Years 1–5, a sample of universal TA recipients will</td>
</tr>
</tbody>
</table>

---

For measures of impact, a possible Likert-type scale could be 1 = no impact, 2 = some impact, 3 = strong impact, and 4 = very strong impact. A mean of 2 would indicate that the recipients, on average, reported some impact of universal TA.
products, and other TA materials to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?

| Addressing Program Performance Measure 1: | At the end of Years 1–5, a sample of targeted TA recipients will rate the quality of the Center’s universal products and other targeted supports for efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale. | Survey |

| Objective 2.B., Activities 2.B.1-3: Targeted TA |

| Addressing Program Performance Measure 2: | At the end of Years 1–5, a sample of targeted TA recipients will rate the relevance of the Center’s universal products and other targeted supports for efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale. | Survey |

| Addressing Program Performance Measure 3: | At the end of Years 1–5, a sample of targeted TA recipients will rate the usefulness of the Center’s universal products and other targeted supports for efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale. | Survey |
4. **To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be of high quality for preparing applications for intensive TA, when appropriate?**

**Addressing Program Performance Measure 1:**
At the end of Years 1–4, a sample of targeted TA recipients, representing states that were not selected for Intensive TA and wish to reapply, will rate the quality of the Center’s products in preparing applications for intensive TA with a mean of at least 3.2 using a four-point Likert-type scale.

| Survey |

5. **To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be relevant to preparing applications for intensive TA, when appropriate?**

**Addressing Program Performance Measure 2:**
At the end of Years 1–4, a sample of targeted TA recipients, representing states that were not selected for Intensive TA and wish to reapply, will rate the relevance of the Center’s products in preparing applications for intensive TA with a mean of at least 3.2 using a four-point Likert-type scale.

| Survey |

6. **To what extent did recipients of targeted TA perceive the Center’s targeted supports and other services to be useful in preparing applications for intensive TA, when appropriate?**

**Addressing Program Performance Measure 3:**
At the end of Years 1–4, a sample of targeted TA recipients, representing states that were not selected for Intensive TA and wish to reapply, will rate the usefulness of the Center’s products in preparing applications for intensive TA with a mean of at least 3.2 using a four-point Likert-type scale.

| Survey |

7. **To what extent did the Center’s targeted TA resources and tools support teacher and leader preparation reform efforts?**

At the end of Years 1–5, a sample of targeted TA recipients will rate the degree of impact of the TA (including the Center’s universal products and other targeted supports) on their ability support teacher and leader preparation reform efforts with a mean of at least 3.2 using a four-point Likert-type scale for items on the TA recipient survey, and during interviews.

| Survey, interview |

---

9 The measures for questions 4–6 are only tracked through Year 4 as that is the last year in which states may apply for intensive TA.

10 As described in Footnote 8, a possible Likert-type scale for impact items is 1 = no impact, 2 = some impact, 3 = strong impact, and 4 = very strong impact. Thus, a mean of 3.2 for impact measures suggests that respondents felt the products and services had at least a “strong” impact (rating of 3), on average, and requires that at least some recipients report a “very strong” impact (rating of 4).
**Objective 2.C., Activities 2.C.1-8: Intensive TA**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent was the MOU and its resulting blueprint implemented according to timeline in each intensive state?</td>
<td>At the end of Years 1-5, recipients and Center staff will achieve 80 percent of annual blueprint milestones.</td>
<td>Extant data, blueprints and capacity rubrics derived from intensive state MOUs</td>
</tr>
<tr>
<td>2. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be of high quality?</td>
<td><em>Addressing Program Performance Measure 1:</em> At the end of Years 1-5, a sample of intensive TA recipients from state leadership teams will rate the quality of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>TA participant and consultant surveys and interviews</td>
</tr>
<tr>
<td>3. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?</td>
<td><em>Addressing Program Performance Measure 2:</em> At the end of Years 1–5, a sample of intensive TA recipients from state leadership teams will rate the relevance of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>TA participant and consultant surveys and interviews</td>
</tr>
<tr>
<td>4. To what extent did recipients of intensive TA perceive the Center’s intensive supports and other services to be useful?</td>
<td><em>Addressing Program Performance Measure 3:</em> At the end of Years 1–5, a sample of intensive TA recipients from state leadership teams will rate the usefulness of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>TA participant and consultant surveys and interviews</td>
</tr>
<tr>
<td>5. To what extent have the Center’s TA services increased participating IHEs’ capacity to build the knowledge and skills of teacher and leader candidates?</td>
<td>At the end of Years 1–5, a sample of intensive TA recipients from state leadership teams will rate the degree of impact of the TA on their ability to build candidates’ knowledge and skill with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>TA participant and consultant surveys and interviews</td>
</tr>
<tr>
<td>6. To what extent have the Center’s TA services increased participating IHEs’ capacity to build the knowledge and skills of teacher and leader candidates?</td>
<td>At the end of Years 4–5, a sample of intensive TA recipients from state leadership teams will rate the degree of impact of the TA on their ability to build candidates’ knowledge and skill with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>TA participant and consultant surveys and interviews</td>
</tr>
</tbody>
</table>

---

As previously mentioned, Evaluation Questions 2.C. 7-11 serve both formative and summative evaluation purposes. The Summative Evaluation Design includes further detail on the analysis of questions 8-10.
<table>
<thead>
<tr>
<th>Question</th>
<th>Description</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>services increased intensive states’ capacity to scale up and sustain reform?</td>
<td>State leadership teams formed in Years 1 and 2 (the second cohort will only be surveyed at the end of Year 5) will rate the degree of impact of the TA on their ability to scale up and sustain reform with a mean of at least 3.2 using a four-point Likert-type scale.</td>
<td>Consultant surveys and interviews</td>
</tr>
<tr>
<td>7. To what extent do SEAs align licensure, accreditation, and evaluation policy?</td>
<td>At the end of Year 5, 100% of participating SEAs will have submitted a plan for aligning licensure, program approval, and evaluation policy.</td>
<td>Capacity rubric (including review of documents reflecting relevant plans and policies), annual surveys, stakeholder interviews</td>
</tr>
<tr>
<td>8. To what extent do IHEs infuse EBP content in TE and LE program?</td>
<td>At the end of Year 5, 100% of participating IHEs will have submitted appropriately revised course syllabi reflecting infusion of EBP content in a TE or LE program.</td>
<td>Capacity rubric (including review of syllabi and other relevant program documents), annual surveys, stakeholder interviews</td>
</tr>
<tr>
<td>9. To what extent do IHEs align pedagogy with EBP?</td>
<td>At the end of Year 5, 100% of participating IHEs will have submitted appropriately revised course syllabi reflecting EBP-aligned pedagogy.</td>
<td>Capacity rubric (including review of syllabi and other relevant program documents), annual surveys, stakeholder interviews</td>
</tr>
<tr>
<td>10. To what extent do graduates of reformed TE and LE programs implement EBP?</td>
<td>By the end of Year 5, the majority of observed teacher graduates from sampled IHEs will demonstrate EBP-aligned practice.</td>
<td>Capacity rubric, annual surveys, stakeholder interviews, teacher knowledge survey, teacher practice observation</td>
</tr>
</tbody>
</table>

12 The number of syllabi to be revised will vary across participating IHEs. Innovation Configurations will be used to review syllabi and identify those in need of revision to reflect EBP content and EBP-aligned pedagogy. The TA blueprint for each state will specify the syllabi to be revised by each IHE and the timeline for completing such revisions.

13 The data for documenting implementation will be collected by observing instruction in classrooms of graduates from sampled IHEs. These data will be available no sooner than year 4 of the project, upon graduation and placement of the year 1 cohort of pre-service teachers. The number of graduates observed and the timing of those observations will depend on the duration of the preparation program and the timing of the sampled IHE’s revision of the relevant preparation program(s).
11. To what extent have SEAs and IHEs begun designing, piloting, and/or implementing improved teacher and leader evaluation systems that use SWD outcome data?

<table>
<thead>
<tr>
<th>At the end of Year 5, 100% of SEAs and IHEs will have submitted plans for developing, piloting, and implementing teacher and leader evaluation systems that use SWD outcome data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity rubric, annual surveys, stakeholder interviews</td>
</tr>
</tbody>
</table>

**Objective 3, Activities 3.1-4: Leadership and Coordination**

1. To what extent did the Center establish collaborative partnerships and consult with professional organizations and individuals interested in personnel preparation reform?

<table>
<thead>
<tr>
<th>At the end of Years 1-5, the Center will have formed or continued subcontractor agreements with all relevant, funded partners.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artifacts of partnerships</td>
</tr>
</tbody>
</table>

2. To what extent did NIC participants perceive the Center’s products and services to be of high quality?

<table>
<thead>
<tr>
<th>At the end of Years 1–5, a sample of NIC participants will rate the quality of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys</td>
</tr>
</tbody>
</table>

3. To what extent did NIC participants perceive the Center’s products and services to be highly relevant to their own work (as well as national and state initiatives) to help improve teacher and leader effectiveness?

<table>
<thead>
<tr>
<th>At the end of Years 1–5, a sample of NIC participants will rate the relevance of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys</td>
</tr>
</tbody>
</table>

4. To what extent did NIC participants perceive the Center’s products and services to be useful?

<table>
<thead>
<tr>
<th>At the end of Years 1–5, a sample of NIC participants will rate the usefulness of the Center’s products and efforts to improve teacher and leader effectiveness with a mean of at least 3.2 using a four-point Likert-type scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys</td>
</tr>
</tbody>
</table>

5. To what extent did the Center coordinate with other relevant OSEP-funded Centers and the TACC on product development and dissemination and maintain communication and coordination with OSEP Project Officers?

<table>
<thead>
<tr>
<th>At the end of Years 1–5, a sample of relevant Centers (non-funded partners) and the TACC will rate the effectiveness of the Center’s coordination efforts with a mean of at least 3.2 using a four-point Likert-type scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys</td>
</tr>
</tbody>
</table>

**Cross-Objective Measures**

1. What is the cost per unit of TA provided by

<table>
<thead>
<tr>
<th>Addressing Program Performance Measure 4:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost data</td>
</tr>
<tr>
<td>Question</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>To what extent has the Center accomplished its planned work scope?</td>
</tr>
</tbody>
</table>

**Addressing Program Performance Measure 5:**
At the end of years 1-5, Center staff will achieve or demonstrate progress toward achieving 100% of work plan milestones.
SUMMATIVE EVALUATION DESIGN

The summative evaluation will focus on intermediate and long-term outcomes in Exhibit 1 using extant data, including data from the formative evaluation, and data provided by the Center investigators and staff.

SUMMATIVE EVALUATION APPROACH

Summative evaluation is often characterized in terms of causality (“Did the program cause changes in the outcomes?”); however, causality can be difficult to demonstrate. A logic model suggests a series of if-then relationships, and for purposes of program planning and management, it is an exceedingly useful tool. It provides a “common sense” perspective on the apparent probabilities associated with the occurrence of \( y \) (i.e., an outcome) having done \( x \) (the program in this case), thus highlighting important relationships in a given data set. It is tempting to assume causality when describing two covarying variables (\( x \) causes \( y \)); however, demonstrating such in a formal sense depends on the quality of the available counterfactual (comparison group or some estimate of “what would have happened to the treated group if treatment had not been applied”), which depends on the design of one’s evaluation study. A strong counterfactual represents a more internally valid design that supports robust claims about a program’s causal relationship with an output or outcome. Randomly assigning units (whether student, teachers, schools, districts, or states) to a treated or untreated group generally yields the best estimate of “what might otherwise have been.” Other designs are possible (and more plausible), and while less rigorous, they can approximate, to differing degrees, the benefits of randomized approaches (i.e., they have high internal validity). Naturally-occurring contrasts (Bryk, 2010; Fagan, 1990) are attractive in situations where budget is limited or when the program under study may evolve over time, whether in response to formative evaluation findings or due to other influences. Although naturally-occurring contrasts can be specified prospectively, they sometimes become apparent only after a program’s onset. Either case requires a well-specified plan for collecting potentially salient data. Bryk’s (2010) systematic naturalistic inquiry provides a useful framework for such an undertaking.

Systematic naturalistic inquiry focuses on prospectively gathering and assimilating evidence about an existing program or initiative to improve ongoing practice, estimate program impact, and represent the complexity of implementing new practices in complex social settings. It combines multilevel modeling, longitudinal design, and a value-added perspective in organizing, collecting, and iteratively analyzing program, person, and context data and identifying important inter-relationships across these domains. In this case, baseline data collected at different levels of the model and from different cohorts of students and teachers will provide the basis for constructing counterfactuals. We will collect pre-implementation data in project years 1 and 2 at the teacher and student levels within selected IHE/SEA/LEA triads. We will incorporate formative data collected from IHE, SEA, and LEA sources. Teacher data and student data will be collected from a sample of schools (discussed in more detail below) within LEAs served by participating IHEs in cohort 1 intensive states.

Systematic naturalistic inquiry is well suited for initiatives implemented at scale, implemented across multiple and diverse sites, and implemented in phases over time. It facilitates a “learning as you go” approach to ongoing implementation and program development. It also provides a basis for evaluating the continuum or “cascade” of effects. In this case, we will be able to address changes from the SEA level, through the IHE level, to the classroom level, including
changes in student-level outcomes. While the likelihood of detecting student-level changes is remote (or at least demonstrating a causal link between such changes and the Center’s activities is remote) given the level of implementation, systematic naturalistic inquiry provides a framework for systematically cataloguing information at all levels of the model for the purpose of detecting meaningful patterns. Although systematic naturalistic inquiry is not specific on data analysis, it accommodates empirical data types (including systematically collected and codeable qualitative data), and we anticipate applying hierarchical crossed value-added effects modeling (ref) during analysis.

Qualitative data will be combined with the quantitative data to create a comprehensive dataset for capturing naturally-occurring contrasts. Best practice in combining qualitative and quantitative data analysis generally involves an integrative analytic strategy, either a concurrent mixed methods approach (Jang et al., 1989) where qualitative and quantitative strands of data are analyzed independently or a data consolidation/data merging technique that involves creating a single data set for analysis from diverse types of data (Caracelli & Greene, 1993; Greene, Caracelli, & Graham, 1989). Because the choice of technique is often determined (or at least influenced) by the type and quality of available data, we would like to defer the decision on this point. We tend towards a concurrent methods approach, because it more reliably represents the “flavor” of qualitative data; converting narrative data types to binary forms often compromises the features that make these data useful in the first place. At the same time, consolidation can be practical when the qualitative data are relatively categorical in nature. A hybrid approach can also be useful (Jang, McDougall, Pollon, Herbert, & Russell, 2008), where data sources and data types are mixed iteratively and at different points in the analysis process, highlighting themes that may be embedded in the data.

We will work with project investigators and key stakeholders in developing a preliminary program model. We will iteratively consult with the stakeholder group to update models to reflect newly uncovered information. We will align our efforts with those of formative evaluators at AIR to achieve seamless coverage of the Center’s activities, outputs, and outcomes. Our purposes are to document changes in the intermediate and long-term outcomes and, to a lesser extent, understand the sources of change that may contribute to changing outcomes. This involves “looking inside the black box” of intervention, at its program activities, their levels of implementation, and the theories of change that underlie hypothetical assumptions. We will be external to the formative team and, pending the availability of evidence, agnostic on questions of quality. However, a non-collaborative approach, in terms of our relation to the formative team, would handicap our respective efforts and compromise the quality of the evaluation outcomes. Thus, we will provide annual reports of activities and findings to project investigators and participate in other planning and data-related meetings per request.

**SUMMATIVE EVALUATION QUESTIONS**

We propose to focus on outcomes related to CEEDAR’s intensive technical assistance efforts (see Exhibit 4). The decision to narrow the summative evaluation’s focus reflects the priority the Center puts on intensive TA. We emphasize intensive technical assistance rather than other program elements because we anticipate small effects (which are typical of diverse and large-scale evaluation settings, regardless the program) and the most intensive elements of the treatment are the most promising for detecting differences. Also, intensive technical assistance is more likely to generate the kind of program-related activity around which naturally occurring contrasts tend to occur. We further propose to select a sample of intensive states in which to
implement summative evaluation. Again, the purposes are to accommodate the available budget and to maximize the likelihood of finding effects (assuming such) in a very low-powered context (i.e., to minimize Type II errors). *The proposed focus is the extent to which the CEEDAR model, when implemented with reasonable fidelity, advantages teacher practice and improves student outcomes. The independent variable, for purposes of summative evaluation, is the teacher training models implemented by IHEs in intensive states. Center-provided TA represents a mechanism for maximizing implementation of the model, and the formative evaluation plan will capture the extent to which supports are provided and the extent to which they support implementation.*

As discussed in the formative evaluation design, certain questions relevant to the intermediate and long-term impacts of intensive TA will be addressed using formative data for continuous improvement of Center services (see questions 8-12 under Objective 2.C). The Center's final report will summatively evaluate the impact of TA on these outcomes.

The questions outlined in Exhibit 4 represent summative *focus questions* for evaluating intermediate and long-term outcomes associated with intensive technical assistance. For each, we describe likely data sources and the nature or source of the counterfactual or contrast. Focus questions provide a core around which data collection mechanisms can be organized according to the earlier described framework (Bryk, 2011). The evaluation questions may change, subtly or substantially, in direct response to the findings of the formative evaluation or indirectly as a result of changes to the independent variable, whether due to formative evaluation or other factors. Additional questions may also be possible assuming the cast of the data collection “net” is sufficiently wide. Accordingly, we frame questions at a fairly inclusive level but provide detail that we expect to be adequate for OSEP’s review and sufficient for establishing reliable and data collection protocols and mechanisms.

*We will implement the summative evaluation in one IHE in one Cohort 1 intensive state.* Data at the SEA, LEA, and IHE levels will be based on the ongoing data collection efforts of the formative evaluation team. To the extent possible, we will coordinate with the AIR team to align a subset of survey, interview, and rubric items with the aims of the summative evaluation. Teacher- and student-level data will be collected from a sample of schools served by IHEs that are partnered with Cohort 1 intensive states. We will develop protocols for collecting teacher knowledge and practice data in participating schools. We will collaborate with CEEDAR Center team and with the formative evaluation team to devise cost-efficient reliable methods of data collection within targeted schools. Our plan is based on these assumptions:

- Eligible states will have a reliable large-scale testing program and be willing to share student-level data (de-identified) with the summative evaluation team. These data will represent the student-level achievement outcome.
- Eligible states will partner with an IHE that has a history of serving nearby LEAs (e.g., 50% of the LEA’s newly hired teachers are from the partnering IHE). We expect to collect teacher-level data on newly practicing professionals who were trained by participating IHEs. We think the data collection proposal is feasible to the extent that a given LEA relies on the IHE for new hires. This will localize the source of “downstream” data, facilitating its collection in a timely manner and minimizing the related costs.
- Eligible states will partner with an IHE that allows the non-intrusive collection of teacher level data, including data on teacher knowledge and teacher practice. They will provide
student-level data clustered in teachers. They will provide salient demographics at the student-level (SWD, ELL, etc.).
### Exhibit 4. Summative Evaluation Questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Data Source(s)</th>
<th>Contrast(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.  To what extent do IHEs infuse EBP content in TE and LE program?</td>
<td>Capacity rubric; Annual surveys; Stakeholder interviews</td>
<td>We will compare within the selected IHE across time (years 2, 3, 4); we will compare trends across measures/data sources. This represents a weak counterfactual to the extent that one is willing to assume that observed changes in the content of TE and LE programs were due to systematic factors other than the CEEDAR model.</td>
</tr>
<tr>
<td>2.  To what extent do IHEs improve TE and LE pedagogy?</td>
<td>Annual surveys; Stakeholder interviews; Teacher knowledge survey; Teacher practice protocols</td>
<td>We will compare within the selected IHE across time (years 2, 3, 4), and we will compare trends across measures/data sources. Also, using data from the measures of teacher knowledge and teacher practice, we will compare late pre-service teacher trainees (juniors/seniors) in project year 2 with early (freshmen/sophomores) pre-service teacher trainees in project year 2 within the participating IHE; we will follow the early pre-service teacher trainees through project year 4. The expectation is between-group differences in project year 2 (early v. late trainees) will be smaller than within-group differences in project year 4 (early trainees across project years 2, 3, 4) on measures of knowledge and practices aligned with EBP. We would argue that this represents a moderately robust counterfactual. Differences due to CEEDAR would be evident to the extent that early pre-service trainees make meaningful 2-year gains and that these gains represent outcomes in project year 4 above and beyond the project year-1 status of the group of late pre-service trainees.</td>
</tr>
<tr>
<td>3.  To what extent do graduates of reformed TE and LE programs implement EBP?</td>
<td>Capacity rubric; Annual surveys; Stakeholder interviews; Teacher knowledge survey; Teacher practice protocols</td>
<td>We will compare within the selected schools across time (years 2, 3, 4), and we will compare trends across measures/data sources. Also, using...</td>
</tr>
<tr>
<td>4. To what extent do students of graduates of reformed TE and LE programs achieve at levels higher than what might have otherwise been?</td>
<td>Capacity rubric; Annual surveys; Stakeholder interviews; Teacher knowledge survey; Teacher practice protocols; Student data</td>
<td>We will model trends in student achievement within LEA for the 7 year period beginning 3 years prior to the CEEDAR implementation. We will compare student trends within schools for teachers who were and were not part of the IHE reform using covariates at the student, teacher, school, LEA, and IHE levels to statistically control for confounds. We believe this represents a moderately robust to robust contrast, depending on data quality and on the prevalence and availability of salient covariates.</td>
</tr>
</tbody>
</table>

Note: Details on the expected data points (student achievement and covariate data) are available from the independent evaluator upon request.
**Plans and Anticipated Timeline for Preparing Center Evaluation Reports**

University of Florida will submit annual evaluation reports in the form of grant performance reports for continuation funding. The first report will confirm the evaluation plan and logic model and summarize any available data. Subsequent reports will provide data regarding project measures, including available data on immediate and short-term outcomes.

The final evaluation report will synthesize all findings and document evidence of impact, particularly related to long-term outcomes and effectiveness of teacher and leader preparation programs. The University of Florida will use summative findings to derive *lessons learned* about knowledge development, TA, and dissemination experiences across the Center’s five-year operation. These lessons will be summarized and offered to OSEP and the field to help guide the operation of other OSEP-supported national TA centers in the future. The University of Florida expects to submit this final evaluation report within 90 days after the end of the Center’s operation.
REFERENCES


