Implementation Evaluation Matrix

**Purpose:** This resource was designed by the National Center for Systemic Improvement (NCSI) to provide states with a sample approach and tool to plan and track measures of State Systemic Improvement Plan (SSIP) implementation. Early and ongoing evaluation of implementation supports continuous improvement by helping states identify solutions that foster expected progress toward intended outcomes, including the State-identified Measurable Result (SiMR). States should feel free to adapt this template or develop their own to meet state needs. This resource will assist states in addressing the SSIP requirements laid out in the State Performance Plan/Annual Performance Report (SPP/APR) Part B and Part C Indicator Measurement Tables[[1]](#footnote-1) and the SSIP Phase II OSEP Guidance and Review Tool,[[2]](#footnote-2) which call for the evaluation of implementation as well as outcomes.

**Intended Audience**: This tool is intended for state teams focused on SSIP planning, implementation, and evaluation. The team completing this matrix should include or seek input from stakeholders who can inform or support activity implementation, implementation data collection, data analysis, and the use and dissemination of findings to support continuous improvement. NCSI technical assistance facilitators and technical assistance providers can support these efforts, including the use of this matrix.

# Instructions

The matrix will support states in answering evaluation questions related to SSIP implementation. If you have not yet clearly defined your intended activities and evaluation questions, see the Sample SSIP Action Plan Template,[[3]](#footnote-3) a resource for developing a Phase II SSIP improvement and evaluation plan. The template will help you develop plans for evaluating and reporting on SSIP implementation. It also assists with planning for outcome measurement (for more information, see the Evaluation of Intended Outcomes section, beginning on page 5). For additional resources to support SSIP evaluation planning, see *State Systemic Improvement Plan (SSIP) Evaluation Planning: A Companion Resource to OSEP’s SSIP Phase II Guidance Tool.[[4]](#footnote-4)* For a broader overview of Phase II, see the *SSIP Phase II Process Guide*.[[5]](#footnote-5)

This matrix will help you specify data collection plans and evaluate the implementation of each prioritized activity. It is important to remember when completing the matrix that although an implementation activity could occur at a higher level of the system (e.g., the state), the data for evaluation might come from a lower level of the system (e.g., districts or schools).

The following sections provide an explanation of the matrix template and three partial examples. A blank matrix template is provided at the end of this document. To navigate between document sections, please use the following links:

Contents

[Instructions 1](#_Toc444874637)

[Partial Template With Explanations 3](#_Toc444874638)

[Example 1: State-Level Activity for K–12 6](#_Toc444874639)

[Example 2: School-Level Activity 8](#_Toc444874640)

[Example 3: Early Intervention Program-Level Activity 10](#_Toc444874641)

[Blank Template 13](#_Toc444874642)

## Partial Template with Explanations

Below is an abbreviated version of the template that lays out the following three steps for evaluating implementation activities:

Step 1: State Your Evaluation Questions

Step 2: Develop Your Plan for Evaluating Implementation and Tracking Implementation Progress

Step 3: Identify Next Steps+

Within each step, explanations are provided in italicized text and footnotes.

**Step 1: State Your Evaluation Questions**

To guide planning, please list the evaluation questions you seek to answer with the implementation information you will collect.

*For guidance on writing evaluation questions, see the previously described resources or Step 4 of* A Guide to SSIP Evaluation Planning,[[6]](#footnote-6) *developed by the IDEA Data Center.*

**Step 2: Develop Your Plan for Evaluating Implementation and Tracking Implementation Progress**

*Within the matrix on the next page, the gray row explains the column headers and the next row is a blank template for evaluating a single activity. In a full template, use one row per activity, adding or deleting rows as needed. The matrix may be further customized to fit state needs. For example, states will develop their own scoring criteria using a rating scale. This supports the comparison of implementation of activities that yield different kinds of data. See the footnotes in the template for additional information and examples of possible modifications.*

| **Activity to Evaluate** | **Data Collection Plan** | **Evaluation of Activity Implementation** |
| --- | --- | --- |
| **SSIP Activity** | **Level of System** | **Sources/Methodology** | **Schedule** | **Scoring Criteria** | **Data/Score** | **Notes** |
| Implemented activity from logic model or action plan | Agency or system level at which activity is implemented[[7]](#footnote-7)  | Describe data source/measurement tool, collection and analysis methods, and parties responsible. | Specify data collection and analysis schedule.[[8]](#footnote-8) | Specify criteria for scoring/rating implementation.[[9]](#footnote-9)  | Provide data used to determine score. Mark score. | Additional information on implementation or rationale for score[[10]](#footnote-10) |
|  | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/ measurement tool:****Data collection and analysis methods:****Parties responsible:** | **Collection schedule:****Analysis schedule:** | 0 = 1 =2 =3 =  | **Data**:[[11]](#footnote-11)**Date**:[[12]](#footnote-12)**Score:**[ ]  0[ ]  1[ ]  2[ ]  3 |  |

**Step 3: Identify Next Steps**

Based on the scores and notes above, identify next steps to address concerns with specific activities or broader implementation plans.[[13]](#footnote-13)

*Step 3 may help you identify strategies for strengthening implementation of a specific activity. For example, if Step 2 revealed low implementation for a certain activity, you may disaggregate your data by site and learn that some sites have strong implementation while others are implementing the activity only partially or not at all. In Step 3, you can identify next steps for digging deeper into why implementation varies across sites. Are there differences in the infrastructure, resources, or supports they received for implementation? How can you address these differences?*

*This step also may help you identify the need to add an activity that you had not previously considered. Weak implementation of one activity may reveal the need for an earlier action or resource that facilitates implementation of that activity. Alternatively, if you have strong implementation but weak outcomes, you may need to add to or change your strategies. This highlights the importance of connecting implementation and short-term outcome data.*

## **Example 1: State-Level Activity for K–12**

The following example lays out evaluation criteria for a state-level activity using a 0–2 scale where:

0 = little or no implementation

1 = partial implementation

2 = full implementation

The state would apply this scale to all activities but set different criteria or anchors as needed. The Evaluation Questions and Next Steps sections have been completed only in relation to the example activity shown in the matrix (thus, only one evaluation question is listed).

**Evaluation Questions**

To guide planning, please list the evaluation questions you seek to answer with the implementation information you will collect.

* Did the state develop criteria for school district improvement plans?

| **Activity to Evaluate** | **Data Collection Plan** | **Evaluation of Activity Implementation** |
| --- | --- | --- |
| **SSIP Activity** | **Level of System** | **Sources/Methodology** | **Schedule** | **Scoring Criteria** | **Data/Score** | **Notes** |
| Define criteria for school district improvement plans | [x]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/measurement tool:**Artifacts of criteria development (meeting notes) and completion (approved criteria guidance document)**Data collection and analysis methods:**Artifacts submitted (collection) and reviewed for evidence of criteria completion (analysis; see Scoring Criteria)**Parties responsible:**Criteria development task force creates and submits artifacts to Bob Jones for analysis | **Collection schedule:**Artifacts submitted as produced**Analysis schedule:** Monthly review of development artifacts until criteria are finalized | 0 = No development artifacts have been submitted.1 = Artifacts reveal evidence of criteria development (e.g., planning meeting, draft criteria).2 = Criteria have been finalized and approved for distribution to school districts. | **Data**: Latest meeting notes included list of 10 potential criteria with 3 defined.**Date**: 9/31/2015**Score:**[ ]  0[x]  1[ ]  2 | The team had hoped to have all criteria defined by this point but had to cancel a planned meeting. |

**Next Steps:** Based on the scores and notes above, identify next steps to address concerns with specific activities or broader implementation plans.

* Schedule an additional meeting to compensate for the missed meeting.
* Task force members should divide up remaining criteria and draft definitions for review at the next meeting.

## Example 2: School-Level Activity

The following example lays out evaluation criteria for a school-level activity using a 0–3 scale where:

0 = little or no implementation

1 = some implementation

2 = moderate implementation

3 = strong implementation

**Evaluation Questions**

To guide planning, please list the evaluation questions you seek to answer with the implementation information you will collect.

* To what extent did pilot schools implement universal literacy screening in Grades K–3?

| **Activity to Evaluate** | **Data Collection Plan** | **Evaluation of Activity Implementation** |
| --- | --- | --- |
| **SSIP Activity** | **Level of System** | **Sources/Methodology** | **Schedule** | **Scoring Criteria** | **Data/Score** | **Notes** |
| Pilot schools will implement universal literacy screening in Grades K–3. | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[x]  School[ ]  Provider[ ]  Other (describe below): | **Data source/measurement tool:**Screening score summaries by school and grade**Data collection and analysis methods:**Summaries submitted via state data system portal (collection); state verifies all relevant schools and grades submitted (analysis)**Parties responsible:**School data managers (collection); Lily Snyder (analysis) | **Collection schedule:**Three times per year, within one month of fall, winter, and spring screening dates**Analysis schedule:** Initial review one month after last school’s screening date; follow-up as needed | 0 = 0–29% of pilot schools1 = 30–69% of schools2 = 70–90% of schools3 = 91–100% of schools | **Data**: 75% of pilot schools**Date**: 10/28/2016**Score:**[ ]  0[ ]  1[x]  2[ ]  3 | All schools from one district reported that their staff were not yet adequately trained in the new screening tool and that no fall screening occurred. Other late schools reported that the screening was completed but they were facing delays in data entry. |

**Next Steps:** Based on the scores and notes above, identify next steps to address concerns with specific activities or broader implementation plans.

* Communicate with districts in which no schools are screening regarding training rollout and identify needed supports.
* Investigate data entry procedures and staffing at late schools.

## Example 3: Early Intervention Program-Level Activity

In the following example, the state chose to customize the template so that the scoring criteria are defined in Step 1 for each evaluation question. Note that the state opted to use different scoring scales for different activities, so scores are not comparable across all activities (e.g., a 2 would indicate full implementation on a 0–2 scale but not on a 0–3 scale). The Step 2 matrix has reduced columns focused on tracking evaluation scores over time.

**Step 1: Evaluation Questions and Implementation Scoring Criteria**

To assist in reporting for Phase III, list the evaluation questions that will be answered with the information collected and define the implementation scoring criteria.

* In how many programs included in the Tele-Health Use Plan was Tele-Health technology deployed?

Scoring criteria: 0 = No deployment; 1 = deployment in at least half of the programs specified in the plan; 2 = deployment in all programs specified in the plan

* Did direct service providers in the programs where the technology was deployed attend the training?

Scoring criteria: 0 = No direct service providers; 1 = up to half of the direct service providers; 2 = more than half of the direct service providers; 3 = 90% of the direct service providers

**Step 2: Implementation Evaluation Tracking Record**

| **SSIP Improvement Activity** | **Data Collection Plan** | **Evaluation** |
| --- | --- | --- |
| **Data Source/Methodology and Schedule*** Data source
* Data collection process
* Data analysis methods
* Parties responsible & schedule
 | **Scoring Criteria and Score*** Date score assigned
* Data
* Score
 | **Notes**Additional information on implementation or rationale for score |
| Deploy Tele-Health technology and train providers to use it | **Data source**: Documentation related to deployment and training, Tele-Health Use Plan, training logs**Data collection process**: Posting in KanbanFlow program management software**Data analysis methods**: Review documents and complete yes/no checklist based on the evaluation questions**Parties responsible**: Tele-Health workgroup lead**Schedule**: Collect data and score after 9/15/2016 and 11/15/2016 | **Evaluation Question 1 Date:** 9/16/16**Data**: 56% of programs**Score:** [ ]  0 [x]  1 [ ]  2**Evaluation Question 2 Date:** 9/16/16**Data:** 42% of providers**Score**: [ ]  0 [x]  1 [ ]  2 [ ]  3**Evaluation Question 1 Date:** 11/22/16**Data:** 100% of programs**Score**: [ ]  0 [ ]  1 [x]  2**Evaluation Question 2 Date:** 11/23/16**Data**: 80% of providers**Score**: [ ]  0 [ ]  1 [x]  2 [ ]  3 | 9/16/16: Deployment has proceeded according to the Tele-Health Use Plan schedule (half of programs are scheduled to have the technology by 9/15/30); however, the trainers have reported delays in scheduling trainings in those programs because the direct service providers have had to attend to a substantial increase in referrals over the past 3 months due to the success of the SSIP work on Child Find.11/22/16: Deployment has been completed according to the plan schedule.11/23/16: The “in milieu” trainings helped catch up and the remainder scheduled will allow us to reach 90% in the next few weeks. |

**Step 3: Next Steps**

9/16/16:

* Continue Tele-Health deployment as planned/scheduled
* Discuss with program directors and trainers the possibility of fitting the training into the ongoing work of direct service providers with enrolled families. With the permission of a few families, the trainers could train a service provider while they are delivering actual services, as opposed to a stand-alone training that does not involve a real-life service activity. It is possible that a few more service providers also could observe and receive the “in milieu” training that way.

11/22/16 and 11/23/16:

* Continue “in milieu” training to reach 90% criterion, estimated to be achieved by 12/15/16.

# Blank Template

**Evaluation Questions**

To guide planning, please list the evaluation questions you seek to answer with the implementation information you will collect.

| **Activity to Evaluate** | **Data Collection Plan** | **Evaluation of Activity Implementation** |
| --- | --- | --- |
| **SSIP Activity** | **Level of System** | **Sources/Methodology** | **Schedule** | **Scoring Criteria** | **Data/Score** | **Notes** |
| Implemented activity from logic model or action plan | Agency or system level at which activity is implemented | Describe data source/measurement tool, collection and analysis methods, and parties responsible. | Specify data collection and analysis schedule. | Specify criteria for scoring/rating implementation.  | Provide data used to determine score. Mark score. | Additional information on implementation or rationale for score |
|  | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/ measurement tool:****Data collection and analysis methods:****Parties responsible:** | **Collection schedule:****Analysis schedule:** | 0 = 1 =2 =3 = | **Data**:**Date**:**Score:**[ ]  0[ ]  1[ ]  2[ ]  3 |  |
|  | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/ measurement tool:****Data collection and analysis methods:****Parties responsible:** | **Collection schedule:****Analysis schedule:** | 0 = 1 =2 =3 = | **Data**:**Date**:**Score:**[ ]  0[ ]  1[ ]  2[ ]  3 |  |
|  | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/ measurement tool:****Data collection and analysis methods:****Parties responsible:** | **Collection schedule:****Analysis schedule:** | 0 = 1 =2 =3 = | **Data**:**Date**:**Score:**[ ]  0[ ]  1[ ]  2[ ]  3 |  |
|  | [ ]  State[ ]  Regional (e.g., professional development providers)[ ]  District[ ]  School[ ]  Provider[ ]  Other (describe below): | **Data source/ measurement tool:****Data collection and analysis methods:****Parties responsible:** | **Collection schedule:****Analysis schedule:** | 0 = 1 =2 =3 = | **Data**:**Date**:**Score:**[ ]  0[ ]  1[ ]  2[ ]  3 |  |

**Next Steps:** Based on the scores and notes above, identify next steps to address concerns with specific activities or broader implementation plans.

1. See Part B Indicator 17 and Part C Indicator 11 in the measurement tables available from <https://osep.grads360.org/#program/spp-apr-resources> [↑](#footnote-ref-1)
2. Available for Part B and Part C from <https://osep.grads360.org/#program/ssip> [↑](#footnote-ref-2)
3. Available from <https://osep.grads360.org/#program/ssip> [↑](#footnote-ref-3)
4. Available at <http://ncsi.wested.org/wp-content/uploads/2016/02/State-Systemic-Improvement-Plan-SSIP-Evaluation-Planning_-A-Companion-Resource-to-OSEP%E2%80%99s-SSIP-Phase-II-Guidance-Tool.pdf> [↑](#footnote-ref-4)
5. Available at <http://ectacenter.org/topics/ssip/ssip_phase2.asp> [↑](#footnote-ref-5)
6. Available from <https://ideadata.org/resource-library/5697cca3140ba0ca5c8b4599/> [↑](#footnote-ref-6)
7. The provided checklist can be removed or customized to reflect your state’s agencies. [↑](#footnote-ref-7)
8. Align your implementation data collection to your evaluation design. How are you examining change over time for the outcomes anticipated for this activity? When and how often will you want information on the implementation of this activity? [↑](#footnote-ref-8)
9. A rating scale will support easier interpretation of raw data and support comparisons across activities that are evaluated using different types of data. Each score or rating will be associated with a range of raw data. Although a three-point 0–2 scale is provided, states may choose to use a different scale, such as 0–3. Examples are provided in the next section. We recommend that states use the same scale for all activities for easier interpretation of scores. Appropriate criteria will vary depending on the activity and state context. Involve your stakeholders in a discussion of what level of implementation is needed to achieve the highest score, considering that perfect or 100% implementation of a given activity may not always be feasible or necessary to make progress toward your goals. [↑](#footnote-ref-9)
10. If your implementation score is not where you expected it to be at this point in time, consider possible reasons. Is implementation of this activity dependent upon completion of another activity? If the activity is being implemented across multiple sites, does the extent of implementation vary across sites? [↑](#footnote-ref-10)
11. The actual number or percentage completed, per the scoring criteria. For example, 10 training modules developed, or 86% of teachers trained. [↑](#footnote-ref-11)
12. Provide the date that the data being used to evaluate the activity were collected. As states reevaluate an activity, they should keep records of previous data and scores to examine progress over time. [↑](#footnote-ref-12)
13. When small teams are completing or reviewing the matrix, they may use this space to identify additional stakeholders who can help them interpret findings and create plans to address concerns. This section can be expanded and refined based on their input. [↑](#footnote-ref-13)