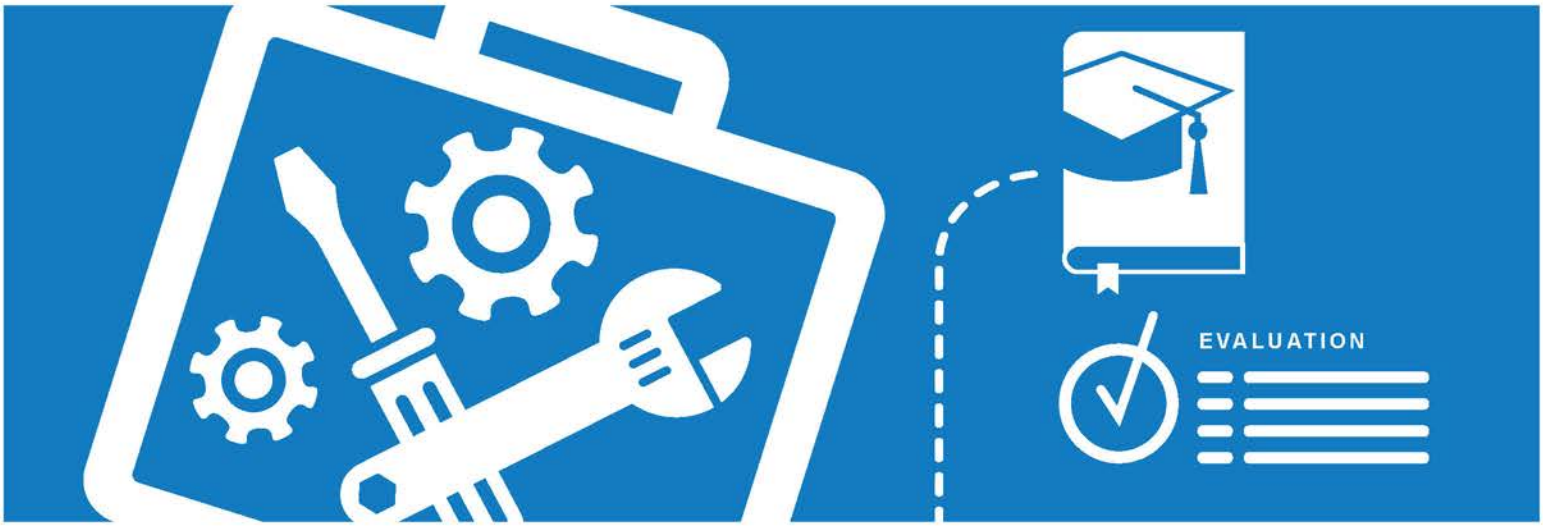


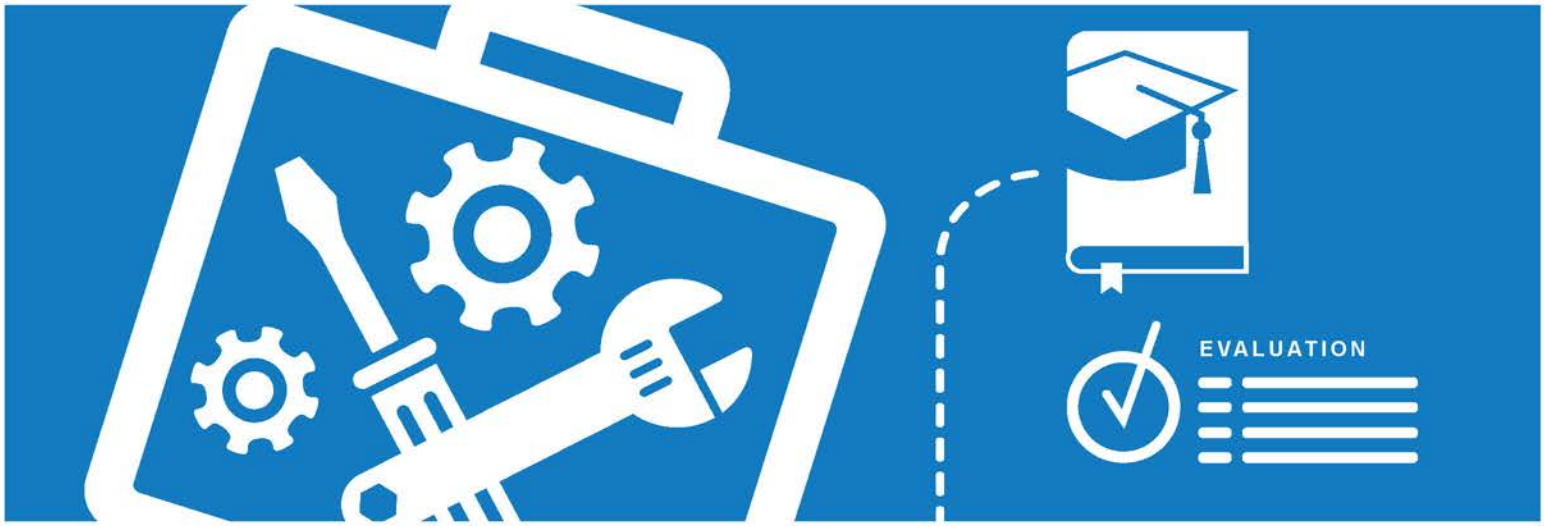
Evaluation Plan Template

The Evaluation Plan Template identifies the key components of an evaluation plan and provides guidance about the information typically included in each section of a plan for evaluating both the effectiveness and implementation of an intervention. Evaluators can use this tool to help develop their plan for a rigorous evaluation, with a focus on meeting *What Works Clearinghouse*™ evidence standards. The template can be used in combination with the Contrast Tool, a tool for documenting each impact that the evaluation will estimate to test program effectiveness.

Prepared by:

Abt Associates
Cristofer Price
Barbara Goodson
Anne Wolf
Beth Boulay

August 2016



The Institute of Education Sciences (IES) has made this tool publicly available as a courtesy to evaluators. However, the content of this tool does not necessarily represent IES's views about best practices in scientific investigation.

This tool was developed under U.S. Department of Education Institute of Education Sciences (IES) contracts ED-IES-10-C-0064 with Abt Associates and ED-ODS-12-A-0019/0031with AEM Corporation and its subcontractor Abt Associates. These contracts provided evaluation technical assistance for evaluations of interventions funded by the Investing in Innovation and First in the World programs. Tools, webinars, and other materials were developed to help grantees engage in good scientific practice and produce evaluations that meet *What Works Clearinghouse*™ evidence standards.

CONTENTS

Grantee/Project Name: _____

WHAT TO INCLUDE IN EACH SECTION

The Evaluation Plan Template provides guidance about the details typically included in each section of a plan for evaluating the effects of an intervention. The guidance appears in italics in a box under each section heading. Throughout, there are references to additional resources or tools that are available to assist you as you develop your evaluation plan, including the U.S. Department of Education's *What Works Clearinghouse*™ Procedures and Standards Handbook, Version 3.0.

You can use this document as a template for your evaluation plan by adding your text below the guidance box. After adding your text, you can delete the guidance box. After editing, update the table of contents, so headings and page numbers remain accurate.

There are some priority sections that should be completed as part of initial evaluation planning. The priority sections are:

- Evaluator information;
- Summary of intervention(s);
- Impact/Effectiveness evaluation, specifically the subsections on
  - research questions,
  - comparison condition,
  - study sample and how intervention and comparison groups are selected/assigned,
  - key measures and plan for obtaining data; and
- Implementation evaluation, specifically the subsections on
  - logic model, and
  - research questions.

The remaining sections, which address your analytic approach, are not as urgent at the beginning of an evaluation and can be completed later in the process. These remaining sections are:

- Subsections of the Impact/Effectiveness evaluation section
  - statistical analysis of models,
  - attrition, and
  - baseline equivalence testing;
- Subsections of the Implementation evaluation section
  - data collection plan and key measures, and
  - analysis approach; and
- Other investigations.

1. Evaluator Information

1.1 Contact Information

List the name and address of the organization or person conducting the independent evaluation. Also include the name, phone number and email information for the primary contact person(s) responsible for carrying out the evaluation for reference.

[Text can be added here.]

1.2 Independence

For some, the organization or person conducting the evaluation must be independent of and external to the entity implementing the intervention and any partners. To be considered an independent and external evaluation: (a) findings reported should not be subject to the approval of the project director or staff conceptualizing/implementing the intervention; (b) the evaluator should independently conduct all key aspects of the evaluation, including random assignment, collection of key outcomes data (other than from administrative records), analyses, and reporting of study findings.

[Text can be added here.]

1.3 Confidentiality Protections

Indicate whether the study has secured relevant Institutional Review Board (IRB) approvals. Describe the plans for protecting confidential data.

[Text can be added here.]

2. Summary of Intervention(s)

This section would typically be a few paragraphs that include the following:

- *Objectives of the intervention (what key outcomes are expected to be affected),*
- *Activities and services that will be newly introduced or expanded and are intended to achieve those objectives,*
- *Target population eligible to receive intervention services (e.g., schools serving disadvantaged students; transfer students; low-achieving students), and*
- *Intended duration of the intervention.*

Make sure to specify how the intervention services work together to achieve the intended change in outcomes for the targeted set of students or faculty. Describe specifically what is given to the intervention group that is not available to the comparison group.

[Text can be added here.]

3. Impact/Effectiveness Evaluation

Note that if the evaluation will include more than one impact study design (e.g., a student-level RCT testing the impact of one component of the intervention and a QED comparing intervention and comparison schools), it's helpful to repeat sections 3.1 through 3.7 below for each design.

3.1 Research Questions

The research questions would typically include the following:

- *The name of the intervention (or, if there is no formal name, the intervention or combination of components),*
- *The comparison condition, (see definitions below),*
- *The outcome expected to be affected, and*
- *The educational level of the student sample (if relevant).*

For example, a research question that includes all of this information might read as follows: "What is the effect of the STEM Academy on college freshmen's persistence and ultimate completion compared to the business-as-usual condition?" Or, another example: "Does providing potential transfer students with access to formal supports (including X, Y, and Z) compared to providing no formal supports increase their rates of college completion?"

[Text can be added here.]

3.2 Comparison Condition

Although you've already named the comparison condition in the previous section, this section can include a more complete description of the condition that the intervention is being compared to. It is helpful to describe the likely instruction, services, or experience of students in the comparison condition, and how they differ from those in the intervention.

Note that study designs that have the potential to meet What Works Clearinghouse Standards (with or without reservations) can have a comparison group that may receive any of the following:

1. *An alternative intervention (e.g., an existing tutoring program, when the intervention being evaluated is a new one)*
2. *"Business-as-usual"(whatever would ordinarily be available to the group targeted for the intervention)*
3. *No treatment at all (when the intervention represents a totally new type of service or activity)*

Evaluations that compare different amounts of exposure to an intervention (e.g., studies of dosage or moderator effects) are not eligible under the WWC standards. Such investigations, which do not examine intervention impacts, can be described in section 5, "Other Investigations".

This difference between the comparison condition and the intervention, also called the "contrast," will be important for determining whether your evaluation has a sufficient sample size and how to interpret the impacts estimated. Typically, smaller differences require larger sample sizes.

[Text can be added here.]

3.3 Study Sample and How Intervention and Comparison Groups are Selected/Assigned

The description of how the intervention and comparison groups are formed typically includes the following information:

Sample:

- *Identification of units eligible for participation (e.g., schools with low graduation rates; students who need developmental education)*
    - *Inclusion/exclusion criteria (e.g., grade level, test scores, demographics, major)*
- *Unit at which groups are to be selected/assigned (e.g., school, faculty/class, student)*
- *Target sample size at each level*
- *Extent to which the sample represents the full population and settings being served by the intervention (e.g., sites, grade levels, demographic characteristics).*

Selection/Assignment:

- *Method of assignment (e.g., random assignment, matching or other non-random approach)*
- *Timing of when intervention and comparison groups are to be selected/assigned*
- *Procedure for selecting/assigning groups (e.g., grouping eligible units on some common dimensions and then randomly assigning within those groups ("blocking" or "stratification"); characteristics used for matching)*
- *Procedures for tracking units after selection/assignment and ensuring intervention delivery to correct group (e.g., monitoring "cross-overs" (comparison group members who inadvertently participate in the intervention) and "no-shows" (intervention group members who do not wind up participating in the intervention offered/available to them)).*

Multi-semester/year and multiple cohort studies:

- *For multi-semester interventions (where the intervention is intended to be provided to the sample for more than one semester) and multiple-cohort studies (where multiple samples receive the intervention in different years), how the sample will be followed over time, including:*
    - *Length of intervention when outcomes are measured (e.g., after the intervention has been in place for one year, two years, and three years)*
    - *Grades/school year when outcomes measured (e.g., grade 12, freshman year)*
    - *Length of exposure for units measured at outcome (e.g., students who have participated in the intervention for two years)*
    - *Number of cohorts that will be included in the sample (e.g., college freshmen from fall 2015, fall 2016, and fall 2017)*
    - *Whether students will join the sample after the intervention and comparison groups have been assigned*.*

If cluster-level selection/assignment:

- *If groups are to be selected/assigned at the cluster level (e.g., school districts, schools, faculty, classes), criteria and procedures for identifying/selecting units within the cluster (e.g., all students in all developmental education classes will be included in the sample; students from one class section for each instructor will be included in the sample; a random sample of 10 students from each class will be included in the sample).*
    - *Inclusion/exclusion criteria*

*****A note about sample selection and assignment, specific to cluster randomized designs:*** *In cluster random assignment designs (e.g., schools randomly assigned to intervention and control conditions), it is important to determine whether the lower level units (e.g., students or teachers in the randomized schools) joined the randomized clusters after randomization. The WWC has a very strict approach to defining* "joiners" *for cluster randomized designs. In short, unless the membership of clusters is clearly defined and determined prior to randomization, then the WWC assumes that the cluster includes joiners. For example, if schools are randomized in the summer, and school enrollment is determined in the fall, the WWC assumes that students may have joined the clusters (schools) after random assignment. When schools are randomized, the WWC defines the sample as not including joiners only if the analysis sample is limited to students who were demonstrably enrolled in the schools prior to random assignment. Therefore, it is critical for cluster randomized designs to be as clear as possible regarding the timing of when cluster membership is determined relative to the timing of the randomization of clusters. For example, can it be demonstrated that all students in the analysis sample were enrolled in schools prior to school randomization (or in classes prior to class or teacher randomization)?*

[Text can be added here.]

3.4 Key Measures and Plan for Obtaining Data

The plan for obtaining data would typically include the following:

- *Time period(s) the data will represent (e.g., spring 2016 and spring 2017 for cumulative credits earned)*
- *Plans for transfer of administrative data or primary data collection, if applicable (e.g., spring 2016 credits earned will be obtained from school records during July-Aug 2016; a freshmen survey will be administered in fall 2016)*
- *Plans to ensure the quality and consistency of data (e.g., strategies for common coding of records from multiple schools; procedures for training data collectors, such as those administering assessments or surveys; procedures for consistency across data collectors, such as those coding responses or observations)*
- *For each outcome measure and baseline measure**
    - *Name and/or description of measure (e.g., instrument and subtest; number of credits accumulated; graduation rate; SAT/ACT score)*
    - *Domain being measured (e.g., math achievement, persistence)*
    - *Unit of measurement (e.g., school, teacher, class, student)*
    - *How variables will be constructed for analysis (e.g., summing/averaging survey items; creating a composite measure; calculating z-scores)*
    - *Information about reliability and face validity*
    - *Indication of the baseline measure(s) that corresponds to each outcome (or if there is no corresponding baseline measure for a given outcome)*
- *Other student, teacher, and/or school characteristics that will be included in impact analysis models as covariates or control variables.*

Note that the **WWC Postsecondary Education Review Protocol (http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_pe_protocol_v3.1.pdf) and the **WWC Developmental Education Review Protocol** (http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_de_protocol_v3.1.pdf) acknowledge that pretest measures of the outcome are often unavailable or not relevant, in which case baseline equivalence for QEDs and RCTs with high attrition (see p.5-6 in each protocol) must be demonstrated in two domains: a continuously-scaled measure of **academic achievement** (e.g., high school grade point average, SAT/ACT scores) and a measure of student **socio-economic status** (e.g., FAFSA expected family contribution, family income, free- or reduced-price lunch status, parent education levels, Pell Grant eligibility).*

[Text can be added here.]

3.5 Statistical Analysis of Impacts

The description of the analytic approach, typically includes the following:

- *For each impact that will be estimated*
  - *A very brief description of the intervention sample and comparison sample (e.g., transfer students that participate in the CollegeConnect support services intervention compared to transfer students that participate in the business-as-usual support services)*
  - *Outcome domain, outcome measure, and unit of observation (e.g., school, teacher, class, student)*
  - *Timing of outcome measurement (i.e., duration of intervention or exposure)*
  - *Measures to be used to establish baseline equivalence (remember to check the most relevant WWC topic protocol, which lists the measures expected to be used)*
  - *Unit of observation for baseline measure(s) (e.g., student, teacher, class, school)*
  - *Timing of baseline measurement (e.g., how long before the intervention was offered or began)*
- *Specification of the statistical model used to estimate the impact of the intervention*
- *Approach to handling missing data (for outcomes, baseline variables, and other covariates)*
- *Strategy for dealing with multiple comparisons (i.e., adjusting the threshold for statistical significance for multiple tests with the same outcome domain)*
- *Subgroups for which the study plans to estimate intervention impacts, if applicable*
- *For RCTs, plans for handling crossovers and no-shows*
- *The estimated minimum detectable effect size, based on the design, planned sample size, and stated assumptions about other relevant factors affecting power (e.g., ICC, R-square)*

[Text can be added here.]

3.6 Attrition (RCTs Only)

Attrition occurs when eligible units (schools, teachers, students) are randomly assigned but, for whatever reasons, data cannot be collected from them. Significant amounts of attrition from either the intervention or comparison/control group or both groups can compromise the initial comparability of the groups resulting from random assignment and potentially lead to biased estimates of the intervention's impact.

For RCTs, this section should describe the strategies that will be used to minimize attrition (school, teacher, student) from the sample. Specifically, the plan should address efforts to maximize the number of units in the sample used to analyze impacts (i.e., those with outcome data).

In addition to describing strategies for minimizing attrition, it is also typical to describe the plan for calculating attrition (i.e., the difference between the number of units assigned and the number of units in the analytic sample – for the sample overall and for each condition separately). For RCTs in which students are randomly assigned to intervention and comparison/control conditions, attrition of ***students*** *should be calculated, both the overall attrition (average, combining both groups) and the differential attrition (the difference in the attrition for each group separately). For RCTs in which clusters (schools, teachers/classrooms) are randomly assigned and joiners are not*

*included in the sample, the overall and differential attrition of both **clusters and students** should be calculated. For RCTs in which clusters are randomly assigned and joiners are included in the sample, the overall and differential attrition of **clusters** should be calculated.*

Note that an RCT cannot meet WWC evidence standards without reservations, if it does not provide clearly stated results on attrition or if attrition rates are above the established attrition thresholds.

[Text can be added here.]

3.7 Baseline Equivalence Testing (QEDs and RCTs with High Attrition)

The section on baseline equivalence testing should include a detailed description of plans for assessing whether the intervention and comparison groups in the analytic sample (i.e., those with outcomes data and included in the impact analysis) are equivalent at baseline on the baseline measure(s) identified for each outcome. The plan for assessing baseline equivalence typically describes the analytic approach that will be used to estimate the intervention-comparison group difference at baseline. Examples of these analytic approaches include:

- *Difference between mean values for each group*
- *Intervention parameter estimate from a regression model with a baseline measure on the left-hand side, as a dependent variable, and the same structural components as the impact model, such as stratum dummies, random terms, or multilevel structure).*

This section should also indicate that the standard deviation of the baseline measures will be reported for both the intervention group and the comparison group, because the WWC will standardize the reported baseline group difference by dividing that difference by the pooled standard deviation of the baseline measure. They will then compare that standardized difference to WWC baseline equivalence thresholds.

[Text can be added here.]

3.8 Implementation Evaluation

An implementation study provides essential contextual information for understanding the intervention and interpreting its impacts. Implementation studies yield important information about when, why, and how interventions work. An implementation study plan should include (a) a well-specified logic model that outlines the key components of the intervention/program; (b) research questions linked to the logic model, including questions about the extent to which the implementation of the program matches the program as planned (fidelity of implementation); (c) plans for systematic, valid data collection; and (d) plans for how the data will be analyzed.

[Text can be added here.]

3.9 Logic Model for the Intervention(s)

The section on the logic model for the intervention would typically include both the graphic illustration and a narrative description of the logic model (i.e., the intervention as planned). Although a narrative description is sufficient, a graphic illustration often helps summarize key intervention components, mechanisms of change, and targeted outcomes. In some cases, aspects of the intervention, target sample, or outcomes may change throughout the planning stages, so carefully review and update your logic model so that it's consistent with your current evaluation plans.

The narrative generally describes the intervention with a sufficient level of detail to allow for replication, if found effective, and includes the following:

- *A clear statement of the population for which it is intended;*
- *The theoretical basis for the intervention;*
- *The expected causal mechanisms by which the intervention should work;*
- *A detailed description of the intervention's content and organization, its duration, the amount required for each activity, intervention procedures, etc.*
- *The hypothesized connections between activities and intended outcomes.*

The graphic illustration generally shows all program inputs, program activities, and theorized short-term, intermediate, and long-term outcomes including:

- *The key components of the intervention (e.g., professional development model, curricular materials, administrative supports)*
  - *The activities associated with each key component*
  - *Professional development activities with faculty, coaches, administrative staff. Specify each type of PD activity, the amount offered/required, and who is providing the PD. This includes on-line training as well as in-person training, mentoring, coaching, etc.*
  - *Instructional strategies with students, including content of instruction, instructional materials, instructional approaches, uses of technology, formative assessment. Indicate the dosage or level of exposure students are expected to have to the programmatic elements.*
  - *Classroom environment elements, such as student groupings, use of time.*
- *Direct outcomes for faculty, administrators, students, etc.*
- *Intermediate outcomes (mediators) for faculty (e.g., changes in instructional practice), administrators, classroom environments, school climate, student attitudes, etc.*
- *Long-term outcomes for faculty, administrators, students (e.g., student academic achievement, persistence), etc.*

For help in creating a logic model, see the following resources:

- *Regional Education Laboratory – Pacific, Education Logic Model Application (An interactive tool for developing a logic model). Available at: http://relpacific.mcrel.org/resources/elm-app*
- *W.K. Kellogg Foundation (2004). W.K. Kellogg Foundation Logic Model Development Guide. Battle Creek, MI: W.K. Kellogg Foundation. Available at: http://www.smartgivers.org/uploads/logicmodelguidepdf.pdf*
- *Knowlton, L.W. & Phillips, CC. The logic model guidebook: Better strategies for great results. Los Angeles: Sage; 2009.*
- *McLaughlin, J., & Jordan, G. (1999). Logic models: A tool for telling your program's performance story. Evaluating and Program Planning, 22, 65-72.*

[Text can be added here.]

3.10 Research Questions for Implementation Evaluation

This section should include the research questions that the implementation study will address, and may include the following types of questions:

- *Were the key components of the intervention implemented with fidelity (i.e., were the key components implemented as planned)?*
- *What was the amount of variation in implementation fidelity?*
- *What was the relationship of fidelity of implementation to intermediate outcomes associated with changes in faculty, coaches, counselors, or other individuals implementing the intervention?*
- *What were the barriers to and facilitators of implementation?*

[Text can be added here.]

3.11 Data Collection Plan and Key Measures

The data collection plan for the implementation study would typically include:

- *Schedule of data collection for each source of data*
- *Who is responsible for collecting each data source*
- *Plans for transfer of data to evaluators*
- *The data sources that will be used to construct measure(s) of implementation fidelity*
- *Plan for coding and scoring data to construct measures of implementation fidelity, including*
  - *Measureable indicators within each key component*
  - *Plan for combining indicator scores to create a fidelity score for each key component*
  - *Unit of measurement*
  - *Plan for defining the threshold for determining whether each key component was implemented with (or without) fidelity at the sample level.*

[Text can be added here.]

3.12 Analysis Approach

The description of the analysis approach would typically describe how the fidelity data and other implementation data will be analyzed to address the research questions.

[Text can be added here.]

3.13 Other Investigations

The section on "other investigations" is the place to describe analyses that are intended to explore the intervention using analytic approaches that cannot (or are not intended to) meet What Works Clearinghouse Standards or are not formal implementation studies. Some examples include studies that compare different amounts of exposure to an intervention (e.g., studies of dosage); explore the relationship between implementation fidelity and outcomes; or examine moderator effects (e.g., interaction effects).There is no expectation that the evaluation will include these types of analyses, so this section may be blank.

To the extent possible, the descriptions here should follow the components of the impact evaluation sections (if the analyses are intended to examine effectiveness) or the implementation evaluation sections (if the analyses are intended to examine qualitative features of the intervention).

[Text can be added here.]

4. References

[Text can be added here.]