## GUIDELINES FOR REVIEWING IMPLEMENTATION STUDIES

These guidelines are intended to promote quality and consistency in Clearinghouse for Labor Evaluation and Research (CLEAR) reviews of selected reports that describe the implementation process of a program or policy, or that measure implementation inputs and outputs to assess the quality and fidelity of implementation of a planned program. The guidelines describe the characteristics that reviewers assess for each selected study and are a framework to support consistency across reviews. Given the range in the intent, approach, and findings of implementation studies, the guidelines are not expected to pertain in full to each study. For studies that combine implementation and other types of analysis (such as impact analysis), reviewers use these implementation guidelines in addition to other relevant review guidelines.[1]

The CLEAR website conveys the results of these reviews through profiles of selected studies.[2] Profiles provide clear and concise information on implementation practices, challenges, and successes, with enough information on the quality of the study and its limitations to place the findings in the appropriate context.

The guidelines are organized into two stages: criteria in the first stage are used to assess the technical qualities of the analysis in relation to the findings it produces, and categories in the second stage are used to describe findings to be reported in the profile produced by CLEAR.

### Stage 1: Technical Qualities

CLEAR does not use a rating system for implementation studies, nor is there a minimum bar of quality or rigor that a study must meet. The purpose of the review of technical qualities is to assess whether the study findings are accurate and appropriate for the design. The criteria for technical adequacy help identify the strengths of the study and any important limitations. The criteria were developed using existing guidelines and similar resources for implementation studies and qualitative research (listed in the final section of this document).

The criteria for technical adequacy assess the following study aspects:

- Scientifically sound design appropriate for addressing the research questions

- Explicit and rational explanation of the selection of sites and respondents, and other sampling strategies, that are appropriate to the design

- Appropriate data sources and systematic, ethical, and unbiased data collection

---

[1] Reviewers examine take-up rates for the treatment group and the conditions for the comparison group as part of the review of the impact study, if applicable.

[2] CLEAR produces *Highlights* on the website for all studies. The highlights include basic information on a given report's objective, setting, methods, and findings. Selected studies receive the more comprehensive review against implementation guidelines discussed in this document.

- Transparent, systematic, and sufficiently rigorous analysis methods

- Explicit and credible findings

The technical qualities guidelines are presented in the form of a checklist that reviewers examine and complete in the course of their reviews (Table 1). For each technical criterion, reviewers indicate whether the issues were appropriately addressed in the study (yes, no, or mixed) and briefly note the rationale. At the end of the technical adequacy section, reviewers summarize the study's key technical strengths and limitations along with their implications for the findings. This information is used to develop the "Considerations for Interpreting the Findings" section of the study profile. For some studies, these considerations can be well aligned with the limitations reported by the study authors; for others, the considerations noted by CLEAR might be different or more comprehensive than those of the authors.

Please note that some considerations are especially relevant to fidelity assessments. For the purposes of this review, fidelity assessments are those studies that measure the degree to which delivery of an intervention adheres to the protocol or program model originally developed and assign ratings (for example, 7 of 10) or rankings (for example, high, medium, or low) to the study sample based on the degree of fidelity.

**Table 1. Checklist for Assessing Technical Quality of Implementation Studies**

| 1. Study Design |
|---|
| **Criterion 1.1:** Is the overall study design appropriate for addressing the research questions? |

| | Is the proposed research strategy appropriate for meeting the aims of study? For example, what will the study do to assess the implementation process or implementation fidelity?<br><br>Does the study present a convincing argument for different features of its design? For example, what are the reasons for different components or stages of research, or the purpose of particular methods or data sources, multiple methods, and time frames?<br><br>Does it discuss limitations of the design and appropriate implications for interpreting findings? For which topics does it note limitations?<br><br>**In addition, for fidelity assessments, does the following occur:**<br><br>Does the study include a logic model? Does it use the logic model to illuminate research questions and guide the study? The model should specify the implementation inputs and outputs, the expected outcomes, and contextual factors, and it should specify the focal elements of the implementation study. |
|---|---|

| **Criterion 1.2:** Are the data sources clearly identified and appropriate for addressing the research questions? |
|---|

| | Does the study do the following:<br><br>• Clearly identify and describe all data sources and methods? For example, does it describe respondent interviews, surveys, or focus groups; observations; documents; and existing data bases?<br>• Document the sample sizes/coverage of each data source?<br>• Indicate the strengths and weaknesses of each data source?<br>• Provide evidence of triangulation of data sources across respondents and/or with administrative or observational data?<br><br>**In addition, for fidelity studies, does the study do the following:**<br><br>• Describe the domains and data sources used to construct measures of implementation inputs (for example, organizational characteristics, staffing structures, and service delivery supports); contextual factors; and implementation outputs (such as content, quantity, quality, and delivery mode of the intervention)?<br>• Discuss the range of values of the implementation fidelity measures that can be computed with these data and methods, and the thresholds that are used to make an assessment of implementation fidelity (by program component)?<br>• Discuss measure validity, particularly those assessing the quality of service delivery? |
|---|---|

| 2. Study Sample |
|---|

**Criterion 2.1:** Is the design and method for selecting sites clearly explained and appropriate for the purposes of the study?

*This criterion is relevant only for studies in which a subset of program sites is included in the study. For studies that include the universe of program sites, this is not relevant. If different data sources include different study sites, address these questions for all major data sources.*

Does the study do the following:

- Describe the breadth of the study in terms of the population of interest and how sample selection relates to it (for example, purposive or representative sample)?
- Specify clearly the criteria and rationale used in site selection?
- Describe a systematic process to obtain eventual sites (including methods of site replacement, if appropriate)?
- Explain how site selection is aligned with the purpose of the study?
- Factor site selection criteria into the interpretation of the findings?
- Discuss what findings on sites can be generalized to the wider population from which it is drawn, as well as limitations to drawing wider inferences?

**Criterion 2.2:** Are the methods for drawing samples from data sources and identifying and sampling respondents clearly defined and defensible?

For each data source in the study, does the study do the following:

- Explain the method/criteria for selecting a sample of respondents?
- Explain how the most appropriate respondents who held the type of knowledge sought by the study were identified and selected for qualitative interviews?
- Discuss missing observations or information and implications for study findings?

For surveys, does the study discuss representativeness, response/participation rates, and potential reasons for nonresponse/nonparticipation?

For extant data, does the study discuss reasons for and methods of drawing samples?

**In addition, for fidelity assessments, does the study do the following:**

- Describe reasons for and methods of selecting examples of program activities for observation? (The importance of randomness and representativeness in this process will depend on the prominence of the activity for the study purpose; representativeness is important in assessing implementation quality and fidelity.)

| 3. Data Collection |
| --- |

**Criterion 3.1:** Does data collection reflect sound and systematic methods to produce reliable data?

Does the study do the following:

- Discuss who conducted data collection and training provided for data collectors?
- Describe interview methods (such as semistructured interviews, use of topic guides) and conventions for taking notes?
- Describe quality assurance procedures in data collection and verification of qualitative data? (Such procedures can include use of audio recording of interviews/group discussions/focus groups to verify completeness and check discrepancies with notes; notes reviewed by a second interviewer; and respondent reviews, also known as member checks, of summary information to ensure accuracy.)
- Describe quality assurance procedures in data collection and verification of quantitative data, such as training and quality checks for conducting surveys, methods of training on observational measures, and achieving and ensuring reliability based on developer standards?
- Discuss how on-site data collection settings or methods might have influenced data collected?
- Identify limitations in data collection methods and steps taken to address them?

**Criterion 3.2:** Does the study include any information that raises the risk that results are biased or that there might be other ethical issues?

Potential issues could include the following:

- Evidence that the research team is not fully independent or objective
- Consent was not obtained from respondents or the study was misrepresented to them
- Data confidentiality procedures and procedures for maintaining respondents' anonymity were compromised
- Necessary approval from an institutional review board was not obtained

| 4. Data Analysis and Findings |
|---|

**Criterion 4.1:** Are the analysis methods clearly described and sufficiently rigorous?

Does the study do the following:

- Describe the analysis methods, such as the use of specific qualitative analysis tools or software, or statistical tests to test hypotheses using quantitative data?
- Present evidence of evenness in use of all data in the analysis to include multiple perspectives and alternative positions (or contradictory data)?
- Describe a consistent and systematic coding process implemented across data sources?
- Describe the original form of data and how the initial descriptive analytic categories/variables were generated and used (for example, use of a protocol to construct descriptive categories a priori, variables derived directly from survey questions)?
- Present evidence of how categories or themes were derived from the data (for thematic analysis of qualitative data or to construct analytic variables from quantitative data)?
- Explain how the data presented were selected from the original sample to demonstrate the analysis process? For example, does the study avoid generalizations and provide information about the extent of an activity or theme? Does it explain the selection of textual abstracts such as quotes or vignettes?

**Criterion 4.2:** Are the links among data, interpretation, and findings clear? Can the route to the findings be explicitly followed?

Are findings and conclusions aligned with the study's research questions?

Are findings clearly supported by the data presented? Does the study present findings accurately and objectively and not introduce speculative hypotheses that are not supported by the data?

Does the study describe how or why particular interpretation or significance is assigned to aspects of the analysis and findings?

Do the findings make sense as a whole? Are seemingly odd or inconsistent findings acknowledged and addressed appropriately?

Does the study demonstrate credibility of the findings? For example, triangulation of data, respondent validation, and more than one researcher involved in analysis?

Are findings placed in the appropriate context with regard to generalizability given limitations in design, data sources, and methods of the study?

Does the study discuss considerations relevant to replicability of findings in different contexts or circumstances?

## Stage 2: Findings

The categories for describing findings (Table 2) align with elements of a basic logic model: contextual factors, planning and design, inputs and resources, activities, and outputs. Additional categories describe other specific characteristics of implementation: use of fidelity measures, costs, perceived benefits, collaboration, challenges, and solutions. These categories guide reviewers in summarizing study findings consistently and comprehensively. They apply to a range of implementation studies, including studies that (1) are descriptive in enumerating the planning and/or implementation process, and detailing the content, structure, and means of delivering services (and to whom); (2) assess the quality of implementation (and possibly identify factors that led to high quality implementation) by measuring implementation inputs and outputs, as well as contextual factors; and (3) assess a program's fidelity to a specific evidence-based model.

Profiles developed by CLEAR reviewers report the key findings, which include factors important to practitioners considering implementing similar programs or policies and findings highlighted by study authors. For studies conducted in conjunction with an impact analysis, the profile reports fidelity to the program model (that is, whether the intervention was implemented as intended).

**Table 2. Categories for Describing the Findings of Implementation Studies**

| 1. Elements of a Basic Logic Model |
|---|
| **Category:** Contextual factors |
| Summarize findings related to factors exogenous to the program that could affect implementation. These can include the demographics and socioeconomic status of sites and study participants, other contextual information on study participants, type and capacity of program providers, and so on. (Include any information that might be pertinent to practitioners interested in adopting the program). |
| **Category:** Planning and design |
| Summarize findings related to the rationale for decisions made about program design, key stakeholders in planning and design, and process. |
| **Category:** Inputs/resources |
| Summarize findings related to factors that are needed to support implementation (what goes into the program), such as the following:<br>• Funding entity and funding levels<br>• Types of organizations and facilities for service delivery<br>• Staff characteristics, training, and performance<br>• Availability of training, technical assistance, and written documents such as policy and program manuals to guide consistent implementation |
| **Category:** Activities |
| Summarize findings related to details of program activities/service components (what the program does), including the following:<br>• Progression of activities<br>• Structure of components<br>• Content of components<br>• Means of delivery of each component (mode, location, and by whom) |
| **Category:** Outputs or outcomes |
| Summarize findings related to measures of what actually happened in the program (did it do what it was intended to):<br>• Service delivery<br>  – Content delivered<br>  – Quality of services<br>  – Mode of service delivery<br>• Participation (prevalence)<br>  – Recruitment and engagement of participants and/or key stakeholders (for example, employers)<br>  – Who participates (relative to eligibility or target population)<br>• Dosage<br>  – Length of participation<br>  – Intensity of participation (hours per week or month and types of activities)<br>• Service receipt rates<br>  – Take-up/utilization of service components<br>• Participants' outcomes<br>  – Educational achievement outcomes<br>  – Labor market outcomes |

| 2. Additional Implementation Findings |
|---|

| **Category:** Fidelity measures |
|---|
| Summarize findings related to whether the program was implemented with fidelity to the program model, including information about the fidelity measures used. |

| **Category:** Costs |
|---|
| Summarize findings related to costs of the program and note whether data were collected by researchers or reported by respondents. |

| **Category:** Perceived benefits |
|---|
| Summarize findings related to perceived benefits as reported by respondents (administrators, staff, and program participants) but not measured by researchers. |

| **Category:** Collaboration |
|---|
| Summarize findings related to the role of collaboration and with whom. |

| **Category:** Implementation challenges |
|---|
| Summarize findings related to reported challenges to implementation. |

| **Category:** Implementation solutions |
|---|
| Summarize findings related to reported solutions to implementation issues (potential successes). |

## Resources

These guidelines were developed and synthesized (and in some cases parts excerpted) from the following sources:

Abt Associates. (2013). Studying implementation in the context of i3. Bethesda, MD: Abt Associates. Available at http://forumfyi.org/files/Abt%20i3%20Learning%20Community%20October%202013.pdf.

Critical Appraisal Skills Programme (CASP). (2013).CASP Qualitative Research Checklist 31.05.13. Oxford, UK: CASP. Available at http://media.wix.com/ugd/dded87_951541699e9edc71ce66c9bac4734c69.pdf.

Hannes K. (2011). Chapter 4: Critical appraisal of qualitative research. In Noyes, J., Booth, A., Hannes, K., Harden, A., Harris, J., Lewin, S., & Lockwood, C. (eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane Systematic Reviews of Interventions.* Version 1 (updated August 2011). Available at http://cqrmg.cochrane.org/supplemental-handbook-guidance.

Mathematica Policy Research. (2013). NCEE guidance for REL study proposals, reports, and other products. Project materials, not publicly available. Princeton, NJ: Mathematica Policy Research.

Spencer, L. Ritchie, J., Lewis, J., & Dillon, L. (2003). Quality in Qualitative Evaluation: A framework for assessing research evidence. London: Government Chief Social Researcher's Office. Available at http://www.civilservice.gov.uk/wp-content/uploads/2011/09/a_quality_framework_tcm6-38740.pdf.

Weiss, M., Bloom, H., & Brock, T. (2013). A conceptual framework for studying the sources of variation in program effects. New York: MDRC. Available at http://www.mdrc.org/publication/conceptual-framework-studying-sources-variation-program-effects.

Zief, S., Henke, J., Walker, K., & Zaveri, H. (2011). Evaluation technical assistance update. Frequently asked questions: Reporting implementation findings. Produced for Office of Adolescent Health and Administration for Children, Youth and Families Teen Prevention Program grantees. Washington, DC: OAH, ACYF. Available at http://www.hhs.gov/ash/oah/oah-initiatives/assets/ta_update_3.pdf.