

April 2013

NCEE Guidance for REL Study Proposals, Reports, and Other Products

NCEE Guidance for REL Study Proposals, Reports, and Other Products

Version 1.0

Background

This document provides guidance for the development and review of project proposals, reports, and other products that will be published by IES. These criteria will be used to fulfill the requirement in the Regional Educational Laboratory (REL) Statement of Work that states that “Any product produced with REL support shall be reviewed by IES according to its scientific standards and made publicly available.” All guidance is drawn from IES research standards [PL 107-279 Sec 102 (18)].

Contents

The guidance in this document is arranged in a table, with questions in the left column and ways to address each question in the right column. These are the same questions that REL CORs and external reviewers will use to review proposals, reports, and other products. Thus, prior to submitting a proposal or report to IES, REL product authors and REL Directors should confirm that every document submitted for review adequately addresses all of the questions raised in the relevant sections of this document.

The table is organized into six sections that correspond to different types of documents. Section I provides guidance about the review criteria that will be used for all submissions. Sections II–VI contain specific additional guidance according to the type of document being submitted for review.

Section	Title	Page
I	Guidance for all REL proposals, reports, and other products to be submitted to IES	2
II	Guidance for proposals for any publicly available materials	10
III	Guidance for reports and other products for publication	13
IV	Guidance for randomized controlled trial (RCT) proposals and reports	15
V	Guidance for literature reviews	25
VI	Additional guidance for systematic evidence review proposals and reports (use along with Section V)	29

I. GUIDANCE FOR ALL REL PROPOSALS, REPORTS, AND OTHER PRODUCTS TO BE SUBMITTED TO IES

A. Relevance and Utility of The Study

1. Does the information have the potential to inform an improvement effort?

The proposal and product should indicate how the findings can inform an action or decision. Although the information and actions that will be affected should have the potential to ultimately benefit students, a study could also focus on intermediate outcomes—for example, a study that examines changes in teacher outcomes that have the potential to benefit students would be acceptable.

2. Do the proposal and report include a summary of the literature and identify which topics have not been previously addressed?

In most cases, a summary of the existing literature should be presented as justification for undertaking the study. The literature review should:

- Be balanced, but not extensive.
- Accurately identify which relevant issues have and have not been addressed previously.
- Focus explicitly on the key study research questions rather than a broad review of the entire field.
- Place more weight on peer-reviewed research. (This includes papers in peer-reviewed journals as well as government published reports that have been reviewed by other researchers, including IES reports). If peer reviewed papers and reports are not available, this should be noted when other research is cited.

For additional details on criteria relevant to literature reviews and systematic evidence reviews please see Sections V and VI below.

3. Are the research questions, hypotheses, and/or study objectives clear and relevant to the policy issue and regional need?



The research questions should:

- Be stated in a way that is clear and relevant to the intended audience (typically State Education Agency [SEA] and Local Education Agency [LEA] staff).
- Be phrased so that the answers are able to clearly address a policy issue or educational need.
- Be fairly compact. Details related to the outcomes, the target population, and the relevant timeframe should be spelled out clearly in the narrative but need not be specified in the research questions.

4. Do resulting study products address the specified research questions in ways that are useful to the intended audience(s)?



Most projects are expected to result in one or more written products suitable for publication on the IES website. (See sections I.C and III below for more details on how products should be organized and written to ensure they are accessible to the intended audience). Each product should:

- Be focused on a single question or closely related set of questions.
- Address the specified research questions in a way that provides actionable information to the intended audience.

B. Study Design, Methods, and Data

5. Is the overall study design appropriate given the research questions?

Authors should select the most rigorous research design possible that addresses each of the main study questions, while also using the simplest approach that will answer the questions (see item 11 below for further discussion of this issue).

If a study includes an impact evaluation, the study must use a randomized controlled trial design and report all information necessary to meet the highest standards of the What Works Clearinghouse (WWC). The proposal will be approved only if it is likely to result in a final product that will meet these standards (see section IV below).

For more detail on analytic methods see #11 below.

6. Are the data sources and variables clearly identified and appropriate for the research questions?

The proposal and report should:

- Provide details on all of the data sources.
- Describe which data sources and variables will be used to address specific research questions.
- Identify intended respondents and other sources (existing databases, websites, etc.) explicitly. In proposals, it must be clear that the researchers know the data are available and appropriate for capturing each of the measures identified.

7. Do the proposal and product clearly explain how and why samples or subsamples (of data or respondents) are selected?

a. Are the samples/subsamples appropriate for the research questions?

b. If a sample of respondents cannot be drawn to represent a relevant universe, is this acknowledged and explained?

The sampling universe and the populations of interest must be clearly defined. Any limitations in the representativeness of the sample should be acknowledged.

Ideally, the sample should correspond to the relevant universe that is the focus of the research questions. If the research questions refer to key subgroups, the proposal and product should identify the size of the corresponding subsamples and confirm whether they are adequate to obtain precise estimates. If the unit of sampling is different from the units whose outcomes or status will be examined—for example, if schools will be sampled and outcomes of students in the selected schools will be examined—the proposal must describe how respondents will be selected within each unit and how clustering affects precision.

For all studies, the proposal and product must be explicit about what issues can and cannot be addressed with nonrepresentative samples. For example, if the study makes use of a convenience sample to answer some research questions using either qualitative or quantitative methods, the proposal and product must be clear about the limitations of the data. (For example, the study team should note if a dataset includes only those students who participated in a standardized test, while the research questions are focused on a larger population of interest).

8. Are the data collection methods, sources, and instruments clearly described and appropriate for the research questions? Are any data collection instruments that are required included with the proposal?

Data collection methods should be selected to elicit the information needed to address each research question.

Generally, instruments should be included in the appendices of the proposal and study product. In the rare cases where draft instruments are not included in the proposal, authors should explain when and how the instruments will be developed and when they will be submitted to IES for review.

In cases where studies are using an existing instrument, authors should provide information on previous studies where the instrument was used as well as information on the validity of the items contained in the instrument.

In all cases, variables should be measured appropriately and methods for ensuring consistency across multiple data collectors should be discussed.

9. Does the proposal adequately address protection of confidential data, if applicable, and does the team have a plan to secure consent if necessary?

Reports and searchable databases should not reveal any personally identifiable information about individual students, families, teachers, school staff, or local and state administrators. Information about individual schools or districts that is not publicly available must also not be revealed in IES products. IES has primary responsibility for ensuring these guidelines are observed, but the proposal should identify any instances in which you believe the planned study may deal with confidential data and how you will keep it secure.

The study proposal should adequately address whether informed consent is necessary (from parents, students, and/or school staff). If consent is necessary, the study proposal should describe a strategy for securing it.

For more information, see the NCEE guidelines for development of Restricted Use Files and Disclosure Analysis Plans.

10. Are the analysis methods clearly described and appropriate for the research questions? Have the authors avoided the use of any unnecessarily complex methods?

Authors should use the simplest methods that will fully answer their research questions. In both the proposal and report they should clearly describe their methods; in the report, some of the details of the methods can be confined to an appendix.

The proposal and report should describe any models and procedures for testing statistical significance of differences across groups or observed changes over time.

For studies using qualitative data (e.g., interview write-ups or other documents), the proposal and report should:

- Describe coding protocols (including any plans for ensuring inter-rater reliability).
- Identify how qualitative data are analyzed including specific tools/programs used.
- Explain the method for selecting quotes or vignettes.
- Identify respondents (using pseudonyms as needed) when citing individuals.
- Avoid generalizations and provide information about the extent of the activity, sentiment, or opinion that was described.

If the study will assess alignment between curricula, standards, and/or assessments, authors should:

- Describe the set of dimensions by which alignment will be determined.
- Identify plans for developing the dimensions.

11. Are the key strengths and especially the key limitations of the data sources and analysis methods described?

The authors should describe the implications of any limitations in the approach, including how they might affect the interpretation of the findings.

C. Accessibility and Readability

12. Is the document well organized, well written, and free of errors?

The document should be well organized, clear, and carefully written. Authors are also expected to ensure that the document is free of editorial errors.

13. Does the document clearly indicate why the reader should care about the issues and findings covered?

The document should explain the research questions and context in a manner that motivates the intended audience to read the document and ensures that readers understand why the issues addressed are important.

14. Does the document provide engaging examples and details?

The document should clearly describe the nature of the issues being studied and contain details that make the findings meaningful and understandable to readers. For example, this can be done by:

- Including compelling vignettes or possible scenarios.
- Presenting data that describe how the current situation in the region is similar or different from the one analyzed in the report.

15. Is the document easy to understand, and are its main messages easy for the reader to ascertain?



The document should:

- Include a useful summary and introduction.
- Contain paragraphs that elaborate on a single point and generally begin with that point. (Short paragraphs are easier to grasp. Start a new paragraphs each time the topic changes.)
- Avoid using unscientific adjectives to describe quantities, frequencies, extent, or significance (such as *many*, *some*, *a few*, *high*, *low*, *very*).
- Include definitions of key terms early in the document.
- Use bulleted lists or other nontraditional formats to make information accessible.
- Avoid jargon and abbreviations where possible.
- Avoid making sources, figures, or tables the subject of a sentence (e.g., “Smith found . . .,” “Table X shows”), instead putting citations and callouts at the end of the sentence in parentheses. (e.g., “Enrollment declined dramatically between 2002 and 2012 (table X).”
- Minimize use of the first person, focusing on the study or findings instead of on the role or actions of the author.

See the NCEE Writers Guide for further discussion of these issues.

II. GUIDANCE FOR PROPOSALS FOR ANY PUBLICLY AVAILABLE MATERIALS

16. Does the study proposal clearly describe the topic and how it relates to a regional need? Will/do the study products provide new information not currently available to key audiences?



The proposal should:

- Clearly articulate why the issue is being investigated.
- Provide sufficient contextual information to make the case that this topic merits investigation in this region. (For the majority of REL work, the investigation should grow out of collaboration with a research alliance; alliance members should agree that the project being undertaken will be of value and will inform their own work).
- Articulate the specific use that a project may have for alliance members. (Where work is occurring outside of an alliance, authors should articulate how the information will be used by other regional stakeholders.)

17. Does the proposal include a timeline for the project or otherwise explain when all key activities will take place? Does the schedule appear feasible, in line with the scope and resources of the planned study?



Study proposals should include sufficient information that gives IES the assurance that the study is within scope and likely to be completed appropriately and on time. The schedule should include time for NCEE review, OMB clearance where necessary, and account for real-world constraints such as school schedules and alliance partners' other time commitments.

18. Does the proposal adequately identify and describe major challenges that could arise in conducting the study, and how the researchers would address them?

Plans should provide key information that gives IES the assurance that the study leaders have considered possible challenges that may arise during data collection, analysis, and reporting. Such challenges often include accessing necessary data, recruitment of appropriate respondents or sites, and maintaining alliance interest over the course of the project. Proposals must describe strategies to address these challenges.

19. Does the proposal specify the type and content of product(s) that will be produced at the study's conclusion?

Project proposals should clearly describe the product(s) that are expected to result from the study, the intended audiences for the product(s), and which products will be made publicly available on the IES website. For projects that will result in multiple products, the authors should clearly articulate the content of each product, including the format and focus of both written and nonwritten products.

20. Are the citations fully referenced?

Citations should follow APA format and must provide adequate information to ensure accessibility to readers and reviewers. (Please see the NCEE Style Guide for further information on citations.)

21. If IRB approval is necessary, is this included in the proposal?

The proposal should indicate whether IRB approval is necessary and when it will be secured.

22. Is there any indication that the data collection will require OMB clearance? If so, does the proposal explain how the study team will limit data collection burden on respondents and schools and justify any planned use of respondent incentives?



IES has primary responsibility for ensuring that OMB guidelines are observed, but the proposal should identify any instances that may require OMB clearance. Specifically, the authors should:

- Identify any plans for similar data to be collected from more than nine respondents.
- Describe efforts to limit data collection burden.
- Provide information about incentives and how they the plan for incentives aligns with NCEE guidelines that require incentives to be modest and justified (for example, small incentives may be needed in order to secure adequate response rates).

For additional information, please consult the NCEE guidance document on OMB submissions.

III. GUIDANCE FOR REPORTS AND OTHER PRODUCTS FOR PUBLICATION

23. Is the need for this product clearly articulated?

The report should make clear why this information is of interest to a broad audience. The proposal made clear how the information could be used by a specific alliance, but the report should make clear to the reader why the information is important to a more national audience. Your report should engage the reader by conveying the possible uses of the information in circumstances likely to be found in many regions.

24. Can the document be readily understood by the intended audience?

The document should include:

- An executive summary of not more than two pages that states the main messages and findings and serves as a roadmap for the rest of the document.
- Information accessible to a broad audience in the body of the report, and technical details and more finely grained findings in appendices.
- A description of the research methods in the body of the report in language that is comprehensible to a reader with interest and familiarity with the content, though not necessarily deep familiarity with research methods.
- Report elements (title, headings, and titles of tables, figures, and boxes) that are content rich and advance the narrative. (A sample heading and subheading: “Academic and graduation outcomes for re-enrollees are mixed,” with subheading: “Most re-enrollees did not earn enough credits to graduate.”).
- Titles of figures and tables that focus on the findings they support (“District A students scored highest on reading comprehension,” not “Comparison of students across districts on exam X”).
- Data displays presented as simply as possible. Aim for focused rather than exhaustive displays to convey information.
- Refer to the NCEE “Writers and Style Guide” for more information on report organization.

25. Are conclusions fully supported by the findings?

Authors should:

- Be sure that the “take-away” messages of the report are clear.
- Discuss potential implications of the study if they are evidence-driven and clearly implications of the findings, rather than demands or policy recommendations. For example, a report should say “The findings indicate that students who participate in X are more likely to do Y” instead of “Schools should do X...”
- Avoid conclusions that are unrelated to the findings or that relate to issues not explicitly related to the data that were analyzed.

26. Is the document properly formatted, following the guidance established by the NCEE?

The document should:

- Adhere to NCEE style conventions for capitalization, headings, tables and figures, callouts, lists, notes, references, typography, and other style elements.
- Contain editable versions of tables and figures (not pictures pasted into the file) and alternative text describing figures and equations. These are required for compliance with Section 508 of the Rehabilitation Act, as amended by the Workforce Investment Act of 1998.
- Include all the required front matter in the correct order and format (title page with authors and affiliations, imprint page, table of contents, summary).
- Include a newsflash summary, subject and keywords, editable vector files (for example, .eps, .ai, or .svg files) for maps, list of tables, figures, and boxes after the table of contents.

(For additional details, see the NCEE Writers and Style Guide.)

IV. GUIDANCE FOR RANDOMIZED CONTROLLED TRIAL (RCT) PROPOSALS AND REPORTS

A. Relevance and Specificity of the Intervention and Research Questions

27. Are the intervention and the existing policy, strategy, or program (hereafter, the counterfactual) clearly described and policy relevant? Could the intervention be replicated elsewhere?

The document should clearly describe the intervention and the counterfactual in sufficient detail to understand how they differ, how each condition could be replicated elsewhere, and how the comparison between the conditions is policy relevant. Simply stating that the counterfactual is “business as usual” is not sufficient. The proposal must demonstrate knowledge of the actual existing practices in the specific contexts where the experiment is to be conducted. Additionally, the intervention should be relevant to many schools in the region: it should address a challenge faced by many schools and should be affordable and feasible to replicate.

28. Are the research questions, hypotheses, and/or study objectives stated early and relevant to the policy issue and needs identified by stakeholders?

Authors should clearly identify how the research questions, target population, and measured outcomes are directly related to the underlying regional need. The need should go beyond an interest in knowing whether a specific intervention is effective. This need should be substantiated with data from the region, and, in the proposal, through evidence of commitment from alliance members. In addition, the proposal should relate the specific intervention to the underlying research question and policy issue.

A study proposal can have both primary (confirmatory) and secondary (exploratory) research questions, but the questions should be limited in number to provide focus to the study. Exploratory questions should only be included if:

- No additional data collection is necessary.
- Additional data analysis is limited.
- The analyses corresponding to the exploratory questions are useful for interpreting the answers to the primary research questions.

29. Is there existing empirical evidence that suggests the intervention can have a positive impact?

Existing empirical evidence should be included here, if available. Examples include:

- Evidence that one or more of the components of the intervention have had a positive impact in some schools.
- Pre-post evidence suggesting that the intervention has improved outcomes relative to a given benchmark (for example achievement gains exceed the average for similar schools or populations of students).

30. Is this an efficacy or an effectiveness study and how does that relate to the conditions under which it is being implemented?

An efficacy study is designed to show an intervention can have a positive impact under highly controlled conditions, often in a limited number of schools. If the RCT is an efficacy study, the proposal and report should describe how the study team will manage (or has managed) implementation to ensure a high level of fidelity. Ideally, the proposal also should provide evidence that the schools to be included in the study are willing to participate and agree to the level of implementation needed to assess efficacy.

Effectiveness studies are typically implemented in a larger number of settings and attempt to capture the real-world impact of a new policy or the impact of a replication or scale-up effort. If the RCT is an effectiveness study, the team should discuss the extent to which participating sites are representative of a broader relevant universe of districts and schools (perhaps limited to those in the REL's region). If participating schools are purposefully selected, the authors should discuss the extent to which the findings will be applicable to a broader set of schools; for example they should describe the extent to which any professional development, technical assistance, or other support for implementation is similar to what other schools outside of the study are likely to receive if the relevant policies are adopted on a broader scale.

31. Does the study include an analysis of the total or incremental costs of the intervention?

NCEE would like RELs to consider including cost effectiveness analyses, though they are not a requirement at this time. If cost analyses are proposed, authors should provide details that indicate that enough information is being collected to measure all costs associated with the intervention (e.g. total cost of the intervention as well as the incremental cost of the intervention relative to the counterfactual). NCEE plans to provide guidance/exemplars on conducting cost analyses in the future.

B. Methods and Data Sources

32. Is the study's target effect size relevant to policy? Can the target effect be detected, given the proposed design and anticipated sample size?

The proposal should present the minimum detectable effect size for each confirmatory analysis of a primary research question, given the study's design, anticipated sample size, expected crossover rates, and anticipated attrition. Minimum detectable effects should be materially significant; studies should be designed with the practical relevance of outcomes in mind. When reviewing calculations for effect sizes, confirm that the anticipated sample size is net of any expected attrition and other data nonresponse. Also, confirm that intra-class correlations were included for group randomized designs, and that reasonable assumptions for the intra-class correlations were used. If the power calculations assume that baseline characteristics of participants will be included in the impact estimation to increase statistical precision, confirm that the explanatory power of the characteristics (R-square) is reasonable. NCEE would like all effect sizes to be stated in real-world terms (such as the amount by which an effect improves student achievement in percentile points).

33. Does the unit of random assignment (for example, student, teacher, or school randomization) balance the goal of attaining an efficient design—that is, requiring the smallest sample size to detect the target effect—with the need to limit contamination and implement the program in a realistic, policy-relevant manner?



The choice of the units to be randomized should balance the need to implement the treatment in a realistic manner, achieve adequate power to detect impacts, and limit contamination. For example, randomizing *districts* to study a *school-based* intervention would minimize the possibility of interfering with implementation and contamination, but randomizing *schools* may suffice for meeting these goals and would be more efficient than *district* randomization.

34. Is random assignment stratified (or blocked) to improve the precision of the impact estimates (or, if needed, to answer research questions relating to subgroups)?



If the random assignment will be stratified variables that represent the strata should be included in the impact calculation, consistent with the sample design.

35. Do/did the design choices appropriately limit the risk of violations to random assignment, contamination, crossovers, or other confounds?



The random assignment process should be truly random, and assignment probabilities should not be susceptible to manipulation by study participants; ideally, these probabilities should be known in advance. If there are exceptions to random assignment, these should be identified in advance, and documented as they occur.

If there is a possibility of crossovers, contamination, or no-shows, the study team should confirm that there will be sufficient data to identify them in the analyses and that the study is still powered to detect impacts. The study should discuss whether there are plans to provide the intervention to the control group in the future and, if so, that the period before this occurs is adequate to measure impacts.

The authors should limit any potential confounds and describe any that remain. Comparable data should be collected from treatment and control groups.

36. Is the timing of random assignment sensible? Is there sufficient time to plan and implement prior to random assignment?



The timing of random assignment (including the process for identifying sample members) should be defined to balance the needs to limit crossovers and no-shows, secure consent (if necessary), and collect baseline data.

Consent to participate in the study should be obtained prior to random assignment.

Ideally, schools should be given sufficient time to plan and train staff prior to random assignment to ensure the treatment is implemented appropriately.

37. Is the follow-up period long enough to detect impacts?



The proposal and report should provide a logic model or theory of action that specifies the intended intermediate and final outcomes including those that will not be measured by the study. Authors should describe the anticipated timing of these outcomes and why one might see changes in the measured outcomes within the specified follow up period. Ideally, authors will include references to previous relevant studies (either implementation or impact studies).

38. Are all members of the treatment and control groups included when calculating the main impacts? Is attrition/nonresponse addressed appropriately?



When calculating the main impacts, the goal should be to include all study units that were randomized. Every effort should be made to minimize attrition and also limit differential attrition (differences in the rates of attrition for the treatment and control groups). The acceptable amount of overall and differential attrition differs across the WWC topic areas and should be applied here too. For example, see page 4 in http://ies.ed.gov/ncee/wwc/pdf/reference_resources/esm_protocol_v2.0.pdf for acceptable levels of attrition for studies about elementary math interventions.

If attrition/nonresponse could occur or has occurred, the proposal and report should describe an attrition/nonresponse analysis including an investigation of treatment-control differences in attrition/nonresponse. The authors also should discuss whether and how weights are used in the impact estimation to account for attrition/nonresponse and how the weights will be or have been constructed.

Study reports should include a CONSORT diagram indicating the initial study sample and the amount of sample lost to various forms of attrition (schools that drop out of the study, sample members not consenting, sample members not providing key data items

Authors may propose to impute missing data, but imputation must be well justified and WWC standards for attrition must still be met. Researchers may not impute outcomes as part of their main analyses. For WWC standards on attrition see page 13 in

http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf. For a useful resource on imputation see: <http://ies.ed.gov/ncee/pdf/20090049.pdf>.

39. Has the study team confirmed baseline equivalence of the treatment and control groups?



After baseline data are collected, the authors should confirm the baseline equivalence of the treatment and control group in terms of the key background covariates and baseline assessments. This analysis should be summarized in the report.

40. Have the authors specified the approach for calculating impacts—in particular, have they specified the statistical model that will be used? Does the document describe how the study will assess or has assessed the sensitivity of impact estimates to alternative model specifications?



The proposal should specify the statistical model that will be used to estimate impacts. Every variable in the model should be explicitly defined and the data source for the variable should be clear.

Study plans and reports can include sensitivity analyses, which examine the robustness of the impacts to different ways of estimating them, but they should identify the main model used to estimate the primary findings.

41. Does the proposal and report include a participation analysis? Are data collected from both the treatment and control groups needed to determine treatment control differences in the actual receipt of intervention and “intervention-like” services?



The proposal and report should include a participation analysis. The study should include data about the receipt of intervention-type services from both the treatment and control groups. Any plan to measure impacts must include a sufficient plan to document the relevant conditions in the treatment and control groups. For example, if the impact of a reading program is being measured, the study must document reading instruction in both groups. Without such information, it is not possible to determine what caused any observed differences in the average outcomes of the two groups.

42. Will or has the study team examined impacts on those treated? If so, what is the rationale for conducting these analyses, and are the methods appropriate?



If the proposal and report include an analysis of impacts of the “treatment on the treated,” (TOT) they should include a justification of this approach and how it addresses a key policy question. For example, estimating TOT impacts may be useful if it is likely that only a modest fraction of individuals in the treatment group will participate in a pilot of the intervention but full implementation of the new program or policy is likely to lead to a considerably higher rate of participation.

43. Is the hypothesis testing plan appropriate?



Consistent with IES criteria, results should only be considered statistically significant if there is no more than a 5 percent chance that treatment-control differences occurred by chance. If random assignment was stratified, degrees of freedom for the statistical tests should reflect the number of strata created.

44. Does the plan include appropriate statistical tests that are adjusted for multiple comparisons?



If multiple comparisons are made (such as examining multiple outcomes within a single domain), the report should present statistical tests with and without multiple comparisons adjustments; the Benjamini-Hochberg approach should be used for these adjustments.

45. Does the proposal and report include an assessment of implementation, such as fidelity to the intervention, program reach, retention, organizational characteristics, etc.)?



To measure implementation fidelity, the proposal should begin by establishing a clear definition of fidelity—that is, how the fidelity measures will be used to assess whether a critical threshold of implementation was achieved. This should be determined in advance, so that decisions about fidelity thresholds are not made based on observed data. In addition, the proposal and report should describe:

- The intended implementation of the intervention (typically based on developer’s expectations),
- How the research team assess how actual implementation compares with planned implementation,
- The reliability of the instruments/measures used to assess implementation fidelity.

46. Are all the key outcome measures reliable and valid? If the outcome differs across sites, such as achievement tests, does the proposal and report describe how different measures are used in the analysis—for example, how different assessments are placed on the same scale?



Outcome measures should not be excessively aligned with the treatment and should have an established relevance to future success (academically, professionally, and/or personally).

The outcome measures related to primary research questions should meet WWC standards for validity and reliability. For an example, see page 3 in http://ies.ed.gov/ncee/wwc/pdf/reference_resources/esm_protocol_v2.0.pdf for the types of outcomes the WWC considers acceptable for studies of elementary math interventions.

V. GUIDANCE FOR LITERATURE REVIEWS

A. Search of Literature

47. Is the search for relevant studies detailed and comprehensive? Are databases and search terms appropriate? Are the time periods for each search indicated? Was the 'gray literature' covered?



The search should ideally cover published and unpublished sources of literature. Sources may include significant scholarly articles, books, and book chapters, trials registers, conference proceedings, general-purpose databases, search engines, meta-search engines, journals, personal or published bibliographies, and "gray literature." Gray literature refers to papers, reports, technical notes or other documents produced and published by governmental agencies, academic institutions and other groups that are not distributed or indexed by commercial publishers.

B. Introduction

48. Does the introduction of the review define and clarify the topic and question to be answered? Does the review inform the reader of the current state of research?



The introduction of the review should provide a statement of the problem, set out the purpose of the proposed review, and state the review question (or hypotheses). The introduction should summarize previous literature reviews that have been done on the same or similar topics.

49. Does the review identify relations, contradictions, gaps, and inconsistencies in the literature? Are terminology and concepts clearly defined?



The introduction should also establish the need for investigation, typically by identifying how it fills a gap in the knowledge accumulated about the subject area. Ideally, it should clarify how it builds on and adds to any previous literature reviews on the topic.

C. Main Body: Analysis and Synthesis of the Literature

50. Does the review assess the quality of the study or the methods used?



The review should assess the quality of studies and describe the strength of evidence associated with any reported findings. It should identify and highlight studies that are the most methodologically rigorous.

51. How does the review categorize the literature?



The literature can be categorized in the following ways:

- In the main body of literature review, studies can be grouped according to common characteristics such as specific aspects of the topic, variants of an intervention or policy, or analytic methods.
- Within each chosen category, the review should provide essential information about each study: specifically, the study's supporting arguments and major evidence.
- If the review is focused on studies measuring the efficacy of an intervention it should employ the standards developed by the WWC to assess the quality of these studies.

52. Is the literature review balanced and impartial in presenting different views on the issue or theory under investigation?



The review should be objective and impartial. It should cover alternative theories about the main question, particularly theories that have been subjected to empirical tests. It should identify the missing elements in the literature.

D. Conclusion

53. Are the results of primary studies clearly displayed? Were results of the review combined across studies in table(s) or concept map format?



Building tables is often a helpful way to organize and summarize findings of literature reviews. For example, tables may include definitions of key terms and concepts, research methods, and summary of research results.

54. Do the authors identify trends and patterns in the results reported by literature review? How well do sources connect to each other?



Typically, authors should critically analyze existing literature looking for trends and patterns. If the results of previous studies are inconsistent or widely varying, this should be discussed.

55. Does the review summarize the main conclusions to be drawn based on the full set of literature?



The conclusion section should summarize major contributions of reviewed studies, maintaining the focus established in the introduction.

56. Does it identify the gaps and next steps in addressing the topic or solving the problem?



The review should summarize what the literature has accomplished, what has not been adequately studied, and what debates still need to be settled.

VI. ADDITIONAL GUIDANCE FOR SYSTEMATIC EVIDENCE REVIEW PROPOSALS AND REPORTS

The criteria pertaining to literature reviews (those in Section V above) also apply to systematic evidence reviews. The criteria listed in this section are additional ones that apply to systematic evidence reviews. Systematic evidence reviews are distinctive in that they follow a pre-specified protocol for searching, inclusion, and assessment of research and focus on the effectiveness of interventions and curricula.

A. Review Protocol

57. Is the review protocol developed? Is the systematic review consistent with the review protocol?



Authors should develop their own systematic review protocol using the template that has been created for the REL program. The systematic review protocol should:

- Document authors' proposed decision rules for the literature search, inclusion/exclusion criteria, data collection, and approach to synthesis (elaborated in criteria).
- Describe the essential procedures for conducting the review and state what evidence is eligible for inclusion, the sources and methods for searching, and how the evidence will be assessed.
- Contain all information that would be sufficient for another reviewer to conduct an identical systematic review.

For a complete example, refer to the WWC Evidence Review Protocol for Adolescent Literacy Interventions, version 2.1:

http://ies.ed.gov/ncee/wwc/pdf/reference_resources/adlit_protocol_v2.1.pdf.

B. Search of Literature

58. Is the search for relevant studies detailed and comprehensive? Are databases and search terms appropriate?

At minimum systematic reviews should make use of general-purpose databases, search engines, journals, bibliographies, and books, and if relevant gray literature, trials registers, and conference proceedings.

59. Are search methods described in enough detail to permit replication? Are the time periods for each search indicated?

Search methods should be described in enough detail to permit replication. The review protocol should thoroughly describe literature search methodology including databases, websites, search parameters and keywords used for electronic searches. Ideally, the full search strategy for at least one database should be listed including search terms and filters used. For a complete example, refer to the WWC Evidence Review Protocol for Adolescent Literacy Interventions, version 2.1: http://ies.ed.gov/ncee/wwc/pdf/reference_resources/adlit_protocol_v2.1.pdf. The indicated time periods covered by the search should balance the need for a thorough review covering that period with the need to cover the period when useful evidence might have been produced. It can also take into account expert knowledge on when the intervention was developed. For example, the search for evidence relating to an intervention developed 10 years ago need only cover the past decade.

C. Quality Assessment of Studies

60. Are the criteria used to assess quality of studies reported? Are study selection criteria reported and clear (types of interventions, outcomes, participants, and study designs)? Are rules for including/excluding findings consistently applied across studies?

Quality assessment of individual studies should proceed in two stages. First, the study team should decide whether to include or exclude studies based on relevance and study design characteristics as described in the developed review protocol (for example, one-group pre-post research designs usually would be excluded from review). Second, the study team needs to make a detailed critical appraisal to determine validity of included study designs, based on specified evaluation criteria.

61. Are studies assessed for methodological quality using the standards and study review guide developed by the WWC?

Reviewers should complete a WWC study review guide (SRG), which is a template used to collect information from primary studies (Please see <http://ies.ed.gov/ncee/wwc/studyreviewguide.aspx>). This guide indicates which key characteristics and findings of individual studies are collected and summarized. The completed SRG will specify study participants, assignment procedure, comparison groups, outcome measures, implementation factors, attrition and biases arising from confounding factors, and major statistical indices (including measures of baseline equivalence and effect sizes).

62. Are study authors contacted for additional information, where applicable?

The team should contact the authors of a paper if additional information is needed to assess the quality of their study (for example, to determine sample attrition rates).

63. Do at least two authors work independently to assess study quality?

Ideally, at least two reviewers should work independently to assess study quality and compile the study review guide information, with a third reviewer reconciling any differences.

D. Outcome Measures

64. Does the review collect quantitative measures of effectiveness from each study? Are the outcome measures of the intervention(s) clearly defined? Are outcome measures reliable, valid, and properly aligned with the intervention?

The validity and reliability of outcome measures (internal consistency, temporal stability/test-retest reliability, and inter-rater reliability) should be assessed using WWC standards. For group design studies: internal consistency score reliability is minimum of 0.60; temporal stability/test-retest score reliability is minimum of 0.40; inter-rater score reliability is minimum of 0.50 (percent agreement, correlation, Kappa; see page 8 in http://ies.ed.gov/ncee/wwc/pdf/reference_resources/adlit_protocol_v2.1.pdf.)

In addition to the reliability with which an outcome is measured, the reviewers should assess whether any outcome measures are too closely aligned with the intervention. Over-alignment may occur when the outcome measure contains content that only the intervention group was exposed to during the intervention.

65. Do the reviewers report all relevant findings defined by the scope of the protocol?



To prevent any bias in the review the study author should report all relevant findings. For example, if the study author tested the intervention on 1st, 2nd, and 3rd graders, but only reported findings for 1st graders, this could give rise to biased findings.

66. Is the timing of the outcome data collection (e.g., posttest, follow-up) clearly specified?



Outcomes can also be measured at the end of an intervention (posttest), or any time thereafter (follow-ups). The review protocol must make clear which outcomes and follow up periods are covered and the rationale for these decisions.

E. Effect Size Calculations

67. Are effect sizes calculated from each primary study where possible?

Do reviewers use appropriate statistics for calculation of effect sizes?



Quantitative data on impacts should be converted into effect sizes. Studies included in a systematic review often use a range of different units for measuring change, which leads to differences in the effect size metric, requiring reviewers to recalculate effect sizes, transforming them into a common metric. The reviewers should describe their plan for this and should explain in the final report how this was done. For continuous outcomes, the WWC has adopted Hedges's g , the most commonly used effect size index. Hedge's g is defined as the difference between the mean outcome of the intervention group and the mean outcome of the comparison group divided by the pooled within-group standard deviation on that outcome measure. The WWC has adopted the Cox index as the default effect size measure for dichotomous outcomes.

68. Are corrections for clustering or multiple comparisons needed and implemented appropriately?



If there is a mismatch between the unit of assignment (e.g., classrooms) and analysis (e.g., students), the findings may overstate precision and a clustering correction should be applied. Similarly, when a study examines many outcomes simultaneously, the statistical significance of findings may be overstated. To correct for multiple comparisons within a domain, the WWC uses the Benjamini-Hochberg method.

For more information about WWC indices, effect size calculations, corrections for clustering, and multiple comparisons, please refer to the *WWC Standards and Procedures Handbook*, Version 2.1, Appendices B–D, pages 37–54:

http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf.

F. Synthesis of Evidence

69. Are the results of included studies clearly displayed?



Tables and figures help to summarize study findings in a systematic and clear format. They provide key information concerning the quality of evidence, the magnitude of effect of the interventions examined, and the available data on all important outcomes for a given comparison.

70. Were results of the review summarized within and across studies? Have the appropriate methods been used to pool effect sizes and standard deviations within primary studies where applicable?



WWC intervention reports provide good examples of research synthesis of evidence on the effectiveness of intervention (please see the *WWC Standards and Procedures Handbook*, Version 2.1, Part IV, pages 19–26:

http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf).

The Campbell Collaboration is a source of other relevant examples of research synthesis. See: <http://www.campbellcollaboration.org/library.php>.

71. Are the reasons for any variation in the results discussed?

If the included studies do not indicate similar effects, the review should identify potential explanations such as: differences in study target population, sample sizes, outcome measures, the treatment or counterfactual conditions, or the impact analysis methods.

G. Meta-analysis (only if relevant)

72. If meta-analysis is proposed, are the methods technically appropriate?

If there are enough studies that are similar with respect to population, outcome, and intervention, a meta-analysis may be conducted as part of systematic review. The meta-analysis aggregates findings across studies to produce a summary estimate of the overall average effect of the intervention and proceeds in two stages. The first stage is the extraction of data from each individual study and the calculation of both the overall average impact (the “point estimate” or “summary statistic”) and an estimate of the standard error or confidence interval of each impact. The second stage involves deciding whether it is appropriate to calculate a pooled average result across studies and, if so, how to take into account the standard errors of each estimate, giving greater weight to the results that are more precise.

73. Do the included studies seem to indicate similar effects? If the heterogeneity of effects is investigated, is this done appropriately?

A common method in meta-analysis is to conduct statistical tests of homogeneity to determine if the effect sizes obtained from the sample of studies are significantly different from what would be expected by chance or sampling error. If the effect sizes are significantly different, then the reviewer might consider whether there are meaningful subgroups or other moderating influences that drive these differences.

74. Is the use of a particular statistical model (fixed or random effects) reported and justified for the studies combined?



The random effects model provides a systematic methodology to manage between- and within-study variation, while the fixed effects model assumes no heterogeneity. Although this remains an area of active debate and work among statisticians, general advice within meta-analytic literature is to use random effects models.

75. Are any sensitivity analyses appropriate?



Ideally, meta-analyses should also include some sensitivity analyses, which examine the robustness of the impacts to different ways of combining them.