



**USAID**  
FROM THE AMERICAN PEOPLE



TOOL

# ASSESSMENT OF STUDY QUALITY (ASQ)

2023

This publication was produced for review by the United States Agency for International Development (USAID). It was prepared by EnCompass LLC and its partner MSI, a Tetrattech company, for the Data and Evidence for Education Programs (DEEP), Contract No. GS-10F-0245M. The views expressed herein do not necessarily reflect the views of USAID.

# BACKGROUND AND PURPOSE

In 2016, the Office of Education in the United States Agency for International Development’s Bureau for Economic Growth, Education, and Environment (USAID/E3) commissioned a team led by Management Systems International to conduct an evaluation synthesis, the first step of which was to assess the quality of USAID-funded evaluations in the education sector through a participatory approach that involved evaluation practitioners in the review of education evaluations. One of the key results of this study was the development of a tool to appraise the quality of evaluation reports in a way that was reflective of international best practices, responsive to USAID’s cross-sector guidance on evaluations, and applicable to sector-specific education evaluations. In 2022, USAID commissioned a team under the Data and Evidence for Education Programs (DEEP) project to revise the tool to be more broadly applicable to research and evaluation in the social sciences and align with updated USAID guidance.

The Assessment of Study Quality (ASQ) Tool (“the tool”) was designed to provide a **common framework on the quality of research and evaluation studies**, codifying best practices in designing, implementing, and reporting on studies. While this tool was produced with funding from the USAID Center for Education, it was developed to be broadly applicable to any social science research and evaluation study, regardless of funding source or social science sector. This tool is mainly **intended for social science researchers and evaluators** to provide guidance and give insight into the elements of research that USAID considers important to quality. Commissioners of social science research and evaluations may also use the tool to guide their expectations of research and evaluation, and users of social science research and evaluation may use it to support their assessment of the quality of studies they reference.

In the remainder of this section, we will provide an overview of the structure of the revised tool and guidance on how to use the tool. We then present the revised tool, followed by a use case to demonstrate how the tool may be adapted for a study using a systems thinking approach. We then include a one-page checklist of the tool for quick reference. Next, we provide the complete list of sources we referenced to develop the tool. In the annex we include a question description table with detailed descriptions to provide in-depth guidance on what to address or look for under each ASQ question.

## A NOTE ON TERMINOLOGY

Following the tradition set forth in the social sciences<sup>i,ii</sup>, we consider evaluation research as a type of social science research. Research, including evaluation research, is defined by standard practices from which rigor and expectations of quality may be derived. Since there are methods shared across all types of research, in this tool **we use the term “study” to be inclusive of all types of primary empirical research, including evaluations**, as well as their corresponding forms of documentation, such as reports.

## STRUCTURE OF THE TOOL

We **designed the tool around the “Principles of Quality” of evidence in education framework** developed by the Building Evidence in Education ([BE2](#))<sup>iii</sup> working group. For the updated version, we added a new principle, ethics, to address the importance of protecting human subjects. From these principles, USAID

identified key elements of the research and evaluation process to include in the tool, each of which are based upon international best practices in social science research and evaluation and are aligned with current USAID policies. The principles are ordered based on the sequence of the study process<sup>iv</sup> and the USAID Evaluation Report Template.<sup>v</sup> **The eight principles are defined in Exhibit I.**

Exhibit I: Principles of Quality

RESEARCH AND EVALUATION PHASE	PRINCIPLE OF QUALITY
DESIGN	<p><b>Conceptual Framing:</b> High quality studies are situated within a theory, acknowledge existing research, and pose specific questions. Questions in this principle address study questions, hypotheses, theoretical frameworks, existing research, and context.</p>
	<p><b>Robustness of Methodology:</b> High quality studies use designs and methods that are appropriate to the stated purpose and questions. Questions in this principle address the extent to which the method is able to answer study questions, the rigor of counterfactuals if applicable to the methods, triangulation, and the rigor of the sampling approach.</p>
	<p><b>Cultural Appropriateness:</b> High quality studies are designed with local stakeholder participation to be contextually appropriate and culturally relevant. Questions in this principle address the cultural relevance of methods, cultural relevance of data collection tools, local validation of findings, and locally relevant stratifiers in data collection and analysis.</p>
	<p><b>Ethics:</b> High quality studies adhere to the highest ethical standards, protect the human subjects involved, and do no harm to children, vulnerable populations, or study participants. Questions in this principle address the protection of human subjects and the research clearance process.</p>
IMPLEMENTATION	<p><b>Validity:</b> High quality studies produce credible and accurate results. Questions in this principle address construct validity, ecological validity, credibility of the findings, external validity or transferability of findings, and confidence intervals.</p>
	<p><b>Reliability:</b> High quality studies use consistent approaches and produce consistent results. Questions in this principle address data collection reliability, internal consistency of data collection instruments, inter-rater reliability, inter-coder reliability, and missing data.</p>
REPORTING	<p><b>Openness and Transparency:</b> High quality studies are transparent about design, methods, data, and limitations. Questions in this principle address study limitations, intervention limitations, biases, transparency about methodological decisions, and cost transparency.</p>
	<p><b>Cogency:</b> High quality studies provide a clear, logical thread linking the purpose to the methods and data to the conclusions. Questions in this principle address the presentation and clarity of the study report, including the writing style, the selection of visuals, the completeness of the study, and the logic connecting the entire study.</p>

We updated the structure of the tool to capture key components in each of the principles as **applicable to different methodological approaches** that are typically used in social science research and evaluation. While most of the questions in this tool are applicable across all methods, the tool is loosely structured around methodological groups to address elements of quality which are unique to a specific set of methods. We used the methodological groups “Experimental/Quasi-Experimental,” “Observational—Quantitative,” and

“Observational—Qualitative” to be consistent with the BE2 guidance as well as to adhere to USAID’s Evaluation Policy. A brief yet non-exhaustive explanation of the three methodological groups is provided in Exhibit 2.<sup>vi</sup>

Exhibit 2. Brief description of methodological groups

Experimental/Quasi Experimental	Observational—Quantitative	Observational—Qualitative
<ul style="list-style-type: none"> <li>• Designs involving the comparison of a treatment group to a non-treatment group. Includes impact evaluations.</li> <li>• Typically explore the effect of a treatment, intervention, or program on a group.</li> <li>• Selected Methods: randomised control trials (experimental), case-control studies (quasi experimental), regression discontinuity design (quasi experimental), other experimental/quasi experimental methods.</li> </ul>	<ul style="list-style-type: none"> <li>• Quantitative designs which do not assign subjects to treatment or control groups. Includes performance evaluations.</li> <li>• May explore quantitative patterns, causal relationships, or correlations, among others.</li> <li>• Selected Methods: case studies, analyses using cross-sectional or panel data, other non-experimental quantitative methods.</li> </ul>	<ul style="list-style-type: none"> <li>• Qualitative designs. Includes performance evaluations.</li> <li>• May explore patterns, behaviors, beliefs, and perceptions.</li> <li>• Selected Methods: case studies, methods using interview/focus group data, ethnographies, participatory designs, other qualitative methods.</li> </ul>

## HOW TO USE THE TOOL

The tool can be used (1) when commissioning a study, (2) when designing and conducting a study, or (3) when reading or reviewing a study report. **Not all questions in the tool will be applicable to all studies; they should be used as relevant and appropriate,** depending on the study parameters.

### WHEN COMMISSIONING A STUDY

USAID Operating Units, donor agency staff, or other actors involved in commissioning a study, may reference this tool to indicate that the funder **expects a researcher or evaluator to design and implement a study with attention to quality standards.** This tool can act as a reference, providing a list of methodology-specific requirements to include in a scope of work for a research or evaluation activity. For example, USAID Operating Units may reference this tool when procuring research or evaluation studies. It is not expected that those commissioning a study will understand all the technical details of this tool, but they can refer to this tool to clarify the expectations they have for those conducting studies and writing reports.

### WHEN DESIGNING AND CONDUCTING A STUDY

Researchers can use this tool when designing and conducting a study, to help **identify what steps should be taken to ensure the study is of high quality.** It is important for those who implement the study to document evidence that each relevant question has been addressed. It is also **important that the**

**documentation is available to the stakeholders who are authorized to access the study products.**

Documentation may exist in various ways, such as in a study design report or inception paper, through explicit reference in a study report, or in annexes to a report.

## WHEN USING A STUDY

Those who are interested in using the evidence generated through studies, such as policy decision-makers, donor agency staff, practitioners, students, or other researchers and evaluators, can use the tool to assess the quality of an individual study design and implementation. For example, an implementing partner may **determine the strength of the evidence generated by a study** by using the tool as a rubric to assess how well a study meets the requirements of applicable questions on the tool under each principle of quality. This will enable the implementing partner to **determine what evidence can be used with confidence** to make decisions about interventions. In a similar way, those who wish to conduct a systematic review of the evidence about a certain topic in the social sciences can use the tool to assess the quality of multiple studies. For example, a researcher may use this tool as a rubric to determine which studies meet minimum quality standards to be included in a systematic review. As a result, the researcher can produce a review with **confidence in the strength of bodies of evidence** and make appropriate evidence-based recommendations.

# 2023 ASSESSMENT OF STUDY QUALITY (ASQ) TOOL<sup>vii</sup>

PRINCIPLE OF QUALITY	EXPERIMENTAL/QUASI-EXPERIMENTAL, INCLUDING IMPACT EVALUATION	OBSERVATIONAL, INCLUDING PERFORMANCE EVALUATION	
		QUANTITATIVE	QUALITATIVE
CONCEPTUAL FRAMING	[1] Are clear study questions that are appropriate to the stated purpose of the study included in the report?		
	[2] If applicable, are study hypotheses included in the report?		
	[3] Are the study questions appropriate to the conceptual/theoretical framework or theory of change?		
	[4] Does the report acknowledge and draw upon existing relevant research?		
	[5] Does the report explain the local context in sufficient detail as it relates to the study purpose and questions?		
ROBUSTNESS OF METHODOLOGY	[6] Is the methodology appropriate for answering posed study questions?		
	[7] Does the counterfactual meet standards of rigor?		
	[8] Does the analysis include triangulation of data from different sources?		
	[9] Does the report mention steps to mitigate common biases or threats to the integrity of the study?		
	[10] Are the sampling approach and size appropriate to the study objectives, calculated to sufficiently accommodate necessary disaggregations, designed to be generalizable/transferable or sufficiently representative of the target population(s), and presented in sufficient detail?		
CULTURAL APPROPRIATENESS	[11] Does the report list the steps taken to ensure that study questions and methodology are informed by local stakeholders, culturally relevant, contextually appropriate, gender-sensitive, and inclusive as appropriate?		
	[12] Does the report demonstrate that data collection tools were developed/adapted with participation of relevant local stakeholders, were piloted with representatives of the target populations and revised as needed, are culturally appropriate, gender-sensitive, and inclusive as appropriate?		
	[13] Does the report list steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders and incorporate stakeholder feedback in the report?		
	[14] Was the study designed to take into account locally relevant stratifiers, such as political, social, ethnic, religious, geographical, sex/gender, disability status, displacement status, socio-economic status, and/or other relevant phenomena, during data collection and analysis?		
ETHICS	[15] Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report?		
	[16] Was/were research clearance(s) appropriate to the study obtained and documented prior to starting data collection?		
VALIDITY	[17] Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated?		

PRINCIPLE OF QUALITY	EXPERIMENTAL/QUASI-EXPERIMENTAL, INCLUDING IMPACT EVALUATION	OBSERVATIONAL, INCLUDING PERFORMANCE EVALUATION	
		QUANTITATIVE	QUALITATIVE
	[18] Is the report open and clear about how the act of doing the study may have biased the findings?		
	[19] Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections?		
	[20] Does the report address the external validity (for quantitative studies) or the transferability (for qualitative studies) of findings?		
	[21] If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates?		
RELIABILITY	[22] Does the report document the steps taken to ensure that data were collected with a high degree of reliability?		
	[23] If applicable, was internal consistency of the instrument(s) established and documented?		
	[24a] For studies where data is collected by a team, was inter-rater reliability established and documented?		[24b] If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team?
	[25] Does the report adequately address missing data/non-response?		
OPENNESS AND TRANSPARENCY	[26] Is the report open and clear about limitations inherent to the study design and with its implementation?		
	[27] For evaluations, is the report open and clear about study limitations due to issues with the implementation of the intervention being evaluated?		
	[28] Is the report open and clear about potential biases due to the study team composition?		
	[29] Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made?		
	[30] For impact evaluations, is a cost analysis of the intervention being evaluated included?		
COGENCY	[31] Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings?		
	[32] Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)?		
	[33] If recommendations are made, are they specific, relevant, actionable, and based on the findings?		
	[34] Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations?		
	[35] Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings?		



# USE CASE: SYSTEMS THINKING APPROACHES

In this use case, we first provide an overview of systems thinking approaches, then present the eight principles of quality through the lens of systems thinking to provide a frame for how the questions in the tool will shift when applying a systems thinking approach.

Systems thinking is a set of approaches that are used to understand complex questions or problems by examining the different components and interactions in a system which could contribute to a possible

Exhibit 2: Systems Thinking Resources

### Find out more: Systems thinking resources

Many resources exist about the systems thinking approaches, including introductory resources, such as [The Systems Thinker](#) and [Learning for Sustainability](#), as well as tool-specific resources, such as [using causal loop modeling for a labor market assessment](#).

outcome.<sup>viii,ix,x</sup> Systems thinking approaches use a wide range of quantitative and qualitative methods, many of which overlap with traditional research and evaluation methods. They take a **holistic approach** to answering a question or solving a problem **while addressing complexity**. Differently from traditional research, systems thinking approaches examine the interactions and links between different elements of a system as

they relate to the question at hand.<sup>xi, xii,xiii, xiv</sup> Foundational to systems thinking approaches is the exploration of the **boundaries** of the system and subject being studied, the **perspectives** of various stakeholders about the subject being studied, and the **inter-relationships** between sub-systems and stakeholders that affect the subject being studied.<sup>xv</sup> Systems thinking is called out as a use case for two reasons:

- To respond to a growing interest in using systems thinking approaches in international development and the social sciences
- To acknowledge the unique terminology of the systems thinking body of inquiry

Exhibit 3 presents the principles of quality through a systems thinking lens to demonstrate how users should shift their interpretation of the ASQ tool when applying a systems thinking approach. The questions in the tool are still applicable, based upon the method selected.

Exhibit 3: Principles of Quality Framed through a Systems Thinking Lens

PRINCIPLE OF QUALITY	SYSTEMS THINKING APPROACH
CONCEPTUAL FRAMING	A study using a systems thinking approach should be framed around the boundaries of the system studied, the perspectives of various stakeholders in the system, and the inter-relationships between sub-systems and agents.
ROBUSTNESS OF METHODOLOGY	A study using a systems thinking approach should use a methodology within the systems approach toolbox that is suited to answer questions about boundaries, perspectives, and inter-relationships. The sample should be designed to represent points of view that bring in various perspectives relevant to the study.
CULTURAL APPROPRIATENESS	A study using a systems thinking approach should ensure that study questions are informed by local stakeholders included and/or excluded from within the system boundaries and that the perspectives from relevant stakeholders within each bounded system are included. The study



PRINCIPLE OF QUALITY	SYSTEMS THINKING APPROACH
	should be designed to examine the inter-relationships between different stakeholder groups.
ETHICS	All studies, including those using a systems thinking approach, must integrate ethical principles for the protection of human subjects into the study approach. Risks to human subjects should be mitigated, and the study should apply principles of “Do No Harm”. <sup>xvi</sup> Perspectives from traditionally marginalized populations should be included in a study using a systems thinking approach, which requires careful attention to risk mitigation and “Do No Harm” so that these populations are not further marginalized through their engagement in the study.
VALIDITY	A study using a systems thinking approach should address the generalizability or transferability of the results to the population defined by the boundaries of the systems/sub-systems. Alternative interpretations about the inter-relationships between sub-systems should be discussed in the findings section.
RELIABILITY	A study using a systems thinking approach should take steps to ensure that different stakeholder groups agree that the system diagrams developed represent their perspectives.
OPENNESS AND TRANSPARENCY	A study using a systems thinking approach should be open and clear about decisions made regarding the boundaries of the study, who made those decisions, and the implications of excluding or restricting any of the system’s agents.
COGENCY	A study using a systems thinking approach clearly connects the study questions, framework, data, analysis, findings, and conclusions through the frame of boundaries, perspectives, and inter-relationships. The systems maps, rich pictures, causal loop diagrams, and other visuals produced in the study should be included in the report and described so that non-technical audiences can understand the inter-relationships between actors and sub-systems and the different perspectives within the system’s boundary.

## ASSESSMENT OF STUDY QUALITY ONE-PAGE CHECKLIST

Study reports should address and document the items under each principle of quality listed below. Not all items will be applicable to all study purpose and methods. The applicability of each item to different study types is indicated in [brackets]. Detailed guidance is included in the Assessment of Study Quality (ASQ) Tool.

PRINCIPLE OF QUALITY	EXPERIMENTAL/QUASI-EXPERIMENTAL	OBSERVATIONAL	
		QUANTITATIVE	QUALITATIVE
CONCEPTUAL FRAMING	<input type="checkbox"/> clear and useful study questions [all studies] <input type="checkbox"/> hypotheses (if applicable) [all studies] <input type="checkbox"/> conceptual/theoretical framework or theory of change [all studies] <input type="checkbox"/> existing relevant research [all studies] <input type="checkbox"/> local context [all studies]		
ROBUSTNESS OF METHODOLOGY	<input type="checkbox"/> selection of appropriate method(s) [all studies] <input type="checkbox"/> rigorous counterfactual [experimental/quasi-experimental studies and impact evaluations] <input type="checkbox"/> triangulation of data [all studies] <input type="checkbox"/> mitigation of common biases or threats [all studies] <input type="checkbox"/> appropriate sampling approach and size		
CULTURAL APPROPRIATENESS	<input type="checkbox"/> cultural relevance of questions and methods [all studies] <input type="checkbox"/> culturally appropriate tools [all studies] <input type="checkbox"/> validated findings, conclusions, and recommendations [all studies] <input type="checkbox"/> account for locally relevant stratifiers [all studies]		
ETHICS	<input type="checkbox"/> human subjects' protection and risk mitigation [all studies] <input type="checkbox"/> research clearance(s) or IRBs/ERCs [all studies]		
VALIDITY	<input type="checkbox"/> indicators/constructs capture the relevant phenomenon [all studies] <input type="checkbox"/> influence of doing the study on the findings [all studies] <input type="checkbox"/> alternative explanations or credibility of findings [all studies] <input type="checkbox"/> external validity or the transferability of findings [all studies] <input type="checkbox"/> standard errors and confidence intervals [experimental/quasi-experimental and observational quantitative studies]		
RELIABILITY	<input type="checkbox"/> reliable data collection procedures [all studies] <input type="checkbox"/> internal consistency of data collection instruments [experimental/quasi-experimental and observational quantitative studies] <input type="checkbox"/> inter-rater reliability [experimental/quasi-experimental and observational quantitative studies] <input type="checkbox"/> inter-coder reliability [observational qualitative studies] <input type="checkbox"/> missing data/non-response [all studies]		
OPENNESS AND TRANSPARENCY	<input type="checkbox"/> limitations to study design and implementation [all studies] <input type="checkbox"/> limitations due to the intervention [all evaluation studies] <input type="checkbox"/> bias due to study team composition [all studies] <input type="checkbox"/> clearly explained methodology [all studies] <input type="checkbox"/> cost analysis of the intervention [impact evaluations]		
COGENCY	<input type="checkbox"/> answers to all study questions and sub-questions in the Executive Summary and the report [all studies] <input type="checkbox"/> written for the intended audience [all studies] <input type="checkbox"/> relevant and actionable recommendations [all studies] <input type="checkbox"/> clear progression from questions to methodology to findings to conclusions [all studies] <input type="checkbox"/> relevant visualizations [all studies]		

# REFERENCES

- Alcott, B., P. Rose, R. Sabates, and R. Torres. "[Handbook on measuring equity in education.](#)" UNESCO Institute for Statistics, 2018.
- Allen, Will. "[Systems Thinking](#)" Learning for Sustainability. Accessed 28 October 2022.
- Arnold, Ross D., and Jon P. Wade. "[A definition of systems thinking: A systems approach.](#)" *Procedia computer science* 44 (2015): 669–678.
- Barroga, Edward, and Glafera Janet Matanguihan. "[A Practical Guide to Writing Quantitative and Qualitative Research Questions and Hypotheses in Scholarly Articles.](#)" *Journal of Korean Medical Science* 37, no. 16 (2022).
- Blaikie, Norman, and Jan Priest. *Designing social research: The logic of anticipation.* John Wiley & Sons, 2019.
- Building Evidence in Education (BE2) Steering Committee. [Assessing the Strength of Evidence in the Education Sector.](#) 2015.
- Building Evidence in Education (BE2). "[Guidance Note on Qualitative Research.](#)"
- Chilisa, Bagele. *Indigenous research methodologies.* 2<sup>nd</sup> Edition. Sage publications, 2019.
- Coghlan, David, and Mary Brydon-Miller. "[Systems Thinking.](#)" In *The SAGE Encyclopedia of Action Research*, edited by Coghlan, David, and Mary Brydon-Miller, 753–754. London: SAGE Publications Ltd, 2014.
- Critical Appraisal Skills Programme (CASP). "[Qualitative Research Checklist.](#)" 2017.
- Glennerster, Rachel, and Kudzai Takavarasha. "[Running randomized evaluations.](#)" In *Running Randomized Evaluations.* Princeton University Press, 2013.
- Goodman, Michael. "[Systems Thinking: What, Why, When, Where, and How?](#)" The Systems Thinker. Accessed 31 October 2022.
- Greener, Sue. "[Research limitations: the need for honesty and common sense.](#)" *Interactive Learning Environments* 26, no. 5 (2018): 567–568.
- Institute of Education Sciences (IES). "[What Works Clearinghouse Procedures and Standards Handbook, Version 3.0.](#)" (2014).
- Institute of Education Sciences (IES). "[What Works Clearinghouse Standards Handbook, Version 4.0.](#)" (2017).
- Institute of Education Sciences (IES). "[What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.](#)" (2022).
- Kim, Daniel. "[Introduction to Systems Thinking](#)" The Systems Thinker. 1999.
- Lamont, Michèle, and Patricia White. "[Workshop on interdisciplinary standards for systematic qualitative research.](#)" In *National Science Foundation Workshop.* 2005.

Lincoln, Yvonna S., and Egon G. Guba. "Establishing trustworthiness." In *Naturalistic inquiry*. Beverly Hills, CA: Sage publications, 1985. 289–327.

RTI International. [Early Grade Reading Assessment \(EGRA\) Toolkit, Second Edition](#). 2015. Washington, D.C.: United States Agency for International Development.

Remler, Dahlia K., and Gregg G. Van Ryzin. *Research methods in practice: Strategies for description and causation*. Sage Publications, 2021.

Saldana, Johnny. *The Coding Manual for Qualitative Researchers*. SAGE, 2021.

Save the Children. "[Gender and Power Analysis Guidance](#)." 2021.

Shenton, Andrew K. "Strategies for ensuring trustworthiness in qualitative research projects." *Education for information* 22, no. 2 (2004): 63–75.

Springer Nature Editors. "[Research must do no harm: new guidance addresses all studies relating to people](#)." *Nature* 606 (2022): 434.

StataCorp, L. L. C. "[Stata Survey Data Reference Manual, Release 13](#)." 2013.

United Nations Evaluation Group (UNEG). "[Quality Checklist for Evaluation Reports](#)." UNEG, June 2010.

USAID ADS Chapter 201, "[Program Cycle Operational Policy](#)." USAID, September 2022.

USAID ADS Reference 201maa, "[Criteria to Ensure the Quality of the Evaluation Report](#)." USAID, September 2016.

USAID ADS Reference 201mah, "[USAID Evaluation Report Requirements](#)." USAID, September 2016.

USAID ADS Reference 201sae, "[USAID Recommended Data Quality Assessment \(DQA\) Checklist](#)." USAID, September 2016.

USAID. "[Ethics in Research and Evaluation in the Education Sector](#)." July 2017.

[USAID Evaluation Policy](#). USAID, October 2020.

[USAID Evaluation Report Template](#). USAID, June 2021.

USAID. "[Local Systems: A Framework for Supporting Sustained Development](#)." April 2014.

Walls, Elena, Caitlin Tulloch, and Christine Harris-Van Keuren. 2021. [Cost Analysis Guidance for USAID-Funded Education Activities, second edition](#). Washington, D.C.: United States Agency for International Development.

Williams, Elizabeth Nutt, and Susan L. Morrow. "[Achieving trustworthiness in qualitative research: A pan-paradigmatic perspective](#)." *Psychotherapy research* 19, no. 4-5 (2009): 576–582.

# ANNEX A: QUESTION DESCRIPTION AND SOURCE

Detailed descriptions for each question in the ASQ Tool are included in this table, under the “Descriptor” column. References for each question are linked in the “Source” column to provide more in-depth guidance (see the “Reference” section for full citations). Researchers and evaluators may refer to this table during study design, implementation, and report writing for detailed guidance about how to address each question in the ASQ Tool. Those reviewing study reports may use this table for guidance on what to look for under each ASQ question when assessing whether a study met the criteria under each principle of quality. Where applicable, we use **bold font to note special considerations** for studies using specific designs (i.e., experimental, quantitative, qualitative) or systems thinking approaches.

Principle of Quality	Question	Score	Descriptor	Source
Conceptual Framing: Appropriate study questions included	[1] Are clear study questions that are appropriate to the stated purpose of the study included in the report?  [All study types]	Yes / No / Partial	<p>The study’s purpose shapes the research questions. All research/evaluation questions must be phrased as questions; it is not enough that they be inferable from the stated objectives of the study. Questions must be clearly stated and be answerable through the reported research methods. All research/evaluation questions should be relevant to the purpose of the study, as described in the report.</p> <p>For studies using a <b>systems thinking approach</b>, there must be a research/evaluation question about the inter-relationships between sub-systems or agents in systems. There must also be research/evaluation questions about the boundaries of the system being studied and the key stakeholders involved in the issue or intervention being studied.</p> <p>“Partial” score could be given if some but not all of the questions are clear or if some but not all of the questions are relevant to the purpose of the study. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">ADS 201maa; BE2 Guidance Note on Qualitative Research</a>, page 9 (figure 1) and pages 61–62.</p> <p><i>Additional source:</i> <a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, page 16</p>
Conceptual Framing: Study hypotheses included	[2] If applicable, are study hypotheses included in the report?  [All study types]	Yes / No / Partial / N/A	<p>For studies requiring hypotheses, research/evaluation hypotheses must be explicitly described; it is not enough that they be inferable from the stated objectives of the study. <b>Quantitative</b> and <b>qualitative</b> studies may require hypotheses, depending on the study design and purpose.</p> <p>For studies using a <b>systems thinking approach</b> which require hypotheses, the hypotheses must be based upon the three foundational concepts: boundaries, inter-relationships, and perspectives.</p> <p>“Partial” score could be given if some but not all hypotheses are explicitly described.</p> <p>“N/A” score should be given for quantitative and qualitative study designs which do not require hypotheses.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, Checklist (page 28); <a href="#">Barroga and Matanguihan, 2022</a>, page 7.</p> <p><i>Additional source:</i> <a href="#">Lamont and White, 2005, page 10</a>.</p>

Principle of Quality	Question	Score	Descriptor	Source
Conceptual Framing: Study questions appropriate to the conceptual/theoretical framework	[3] Are the study questions appropriate to the conceptual/theoretical framework or theory of change? <b>[All study types]</b>	Yes / No / Partial	<p>Study questions should be appropriate to the study purpose. Conceptual or theoretical frameworks, including theories of change, should inform the study questions and the remainder of the study design.</p> <p>Conceptual or theoretical frameworks shed light on how an issue is being framed and the major assumptions made in a study. High quality studies explicitly detail the conceptual or theoretical frameworks used, including clearly stating the assumptions.</p> <p>For <b>evaluation studies</b> or studies addressing an intervention, this means that questions should be based on the intervention's theory of change or results framework.</p> <p>For studies using a <b>systems thinking approach</b>, the conceptual or theoretical framework should address boundaries, perspectives, and inter-relationships. The assumptions should address the inter-relationships between system agents/sub-systems.</p> <p>“Partial” score could be given when a framework is described but the assumptions embedded within the framework are not described, or when some, but not all, listed questions correspond to the framework or the intervention’s theory of change. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, page 16 and Checklist (page 28).</p> <p>Additional source: <a href="#">USAID Evaluation Policy</a>, page 8</p>
Conceptual Framing: Study acknowledges/draws upon existing country-specific research	[4] Does the report acknowledge/draw upon existing relevant research? <b>[All study types]</b>	Yes / No / Partial	<p>Studies should build on existing research, both local and funded by international donors. The report should specify how questions, methodology, tools, and analysis plans are informed by prior research.</p> <p>“Partial” score could be given when only some of the questions are informed by existing knowledge. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, Checklist (page 28)</p>
Conceptual Framing: Local context provided allows non-experts appreciate relevance of the study	[5] Does the report explain the local context in sufficient detail as it relates to the study purpose and questions? <b>[All study types]</b>	Yes / No / Partial	<p>The local context should be explained in enough detail for a general audience to be able to appreciate the relevance of the study or the relevance of the intervention being evaluated.</p> <p>“Partial” score could be given when some, but not all, elements of the study and/or intervention have corresponding contextual information detailed in the report. Other reasons to give a “partial” score are possible and up to the judgement of the reviewer.</p>	<p><a href="#">USAID Evaluation Policy</a>, page 8;</p> <p><a href="#">BE2 Guidance Note on Qualitative Research</a>, page 8</p>
Robustness of Methodology: Methodology appropriate for answering posed study questions	[6] Is the methodology appropriate for answering posed study questions? <b>[All study types]</b>	Yes / No / Partial	<p>The assessment of the methodology must include the review of the study design vis-à-vis each stated study question. USAID recognizes that the methodology used to address the posed questions may be defined in the issued scope of work for the evaluation. USAID also recognizes that different designs are more or less appropriate to answering different research questions, and that the selection of a method (or methods) for a particular study also balances cost, timelines, feasibility, and the level of rigor needed to inform specific decisions. Assessing the appropriateness of the chosen methodology may be further complicated when the study includes a variety of questions that require a mixed-method approach.</p> <p>“Partial” score could be given if the methodology proposed is appropriate for some, but not all posed questions. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">USAID Evaluation Policy</a>, page 9.</p>

Principle of Quality	Question	Score	Descriptor	Source
Robustness of Methodology: Counterfactual meet standards of rigor	[7] Does the counterfactual meet standards of rigor? <b>[Exp./Quasi]</b>	Yes / No / Partial / N/A	<p>Measuring what would have happened in the absence of an intervention is a requirement for establishing a causal relationship. A counterfactual can be created in a number of ways, from simply using respondents from a geographically close unit as comparison group to using statistical analysis to compensate for the potential selection biases of non-randomization to randomly assigning subjects to treatment(s) and control groups. A regression discontinuity design is another acceptable form of establishing a counterfactual. Considerations about its rigor may include a review of information in the report about baseline equivalence, differential attrition, etc.</p> <p>"Partial score" could be given if the counterfactual is mentioned but not fully described. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.</p> <p>"N/A" score should be given if the study is not an impact evaluation or a study using an experimental/quasi-experimental design.</p>	<a href="#">USAID Evaluation Policy</a> , page 2.
Robustness of Methodology: Data triangulation described as part of methodology	[8] Does the analysis include triangulation of data from different sources? <b>[All study types]</b>	Yes / No / Partial / N/A	<p>Typically, stronger bodies of evidence are likely to emerge if similar findings are obtained from different types of data (e.g., tests, interviews, or observations) and respondent types (e.g., students, parents, or teachers). It is important that contradictory data be considered when discussing the findings.</p> <p>"Partial" score could be given if data from different sources are presented but the findings don't connect them into a coherent narrative. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.</p> <p>"N/A" score should be given if the study does not use multiple data sources.</p>	<p><a href="#">CASP</a>, Qualitative Checklist, page 4.</p> <p>Additional sources: <a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, page 26; <a href="#">BE2 Guidance Note on Qualitative Research</a>, page 61</p>
Robustness of Methodology: Addressed internal validity, either threats to inference or common biases	[9] Does the report mention steps to mitigate common biases or threats to the integrity of the study? <b>[All study types]</b>	Yes / No / Partial	<p>USAID Evaluation Policy requires that evaluation reports address methodologically common limitations, such as when there is a disjunction between the treatment that is assigned and the treatment that is received (non-compliance). Research and other non-evaluation studies should follow the same guidance.</p> <p>Some common threats to the integrity of <b>quantitative studies</b> may include non-equivalence at baseline, non-compliance, spillover, and systematic attrition. Some common biases for quantitative studies may include confounding bias, selection bias, and experimenter bias.</p> <p>Some common threats to the integrity of <b>qualitative studies</b> may include threats to trustworthiness such as participant non-availability. Some common biases for qualitative studies may include selection bias and researcher bias.</p> <p>Other threats to the integrity/trustworthiness and other common biases may be discussed in the report as well.</p> <p>"Partial" score could be given if some, but not all threats or biases identified are discussed. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">USAID Evaluation Policy</a>, page 11.</p> <p>Additional sources: <a href="#">What Works Clearinghouse Procedures and Standards Handbook Version 5.0</a> page 18</p>



Principle of Quality	Question	Score	Descriptor	Source
<p>Robustness of Methodology: Described sampling approach and parameters used to compute sample size</p>	<p>[10] Are the sampling approach and size appropriate to the study objectives, calculated to sufficiently accommodate necessary disaggregations, designed to be generalizable/transferable or sufficiently representative of the target population(s), and presented in sufficient detail?</p> <p><b>[All study types]</b></p>	<p>Yes / No / Partial</p>	<p>For <b>quantitative studies</b>, a number of characteristics of the study design, such as timing of the assessment and absence of sampling weights, may affect the interpretation and/or calculation of population estimates. The evaluator/research should provide information about the timing of the assessment (e.g., pre-test and post-test being conducted at comparable time points in a cross-sectional design), construction and use of sampling weights in the analysis (when different observations in a random selection process may have different probabilities of selection). Sampling details should include, at a minimum, sample size calculations, documentation of intended and achieved sample size, type of analysis, and power calculations. Details of power calculation should be included in either the main body of the report or in an annex. This should include the parameters used in the power function that relates power (beta) to its determinants: (1) level of significance (alpha), (2) minimum detectable effect size (MDES) or minimum detectable impact (MDI), and (3) the sample size. Sampling for two or more levels must include intraclass correlation coefficient (ICC) as part of the power analysis. Where the calculations are included should be based upon the intended audience of the report. For example, if the report is intended for a research audience, the calculations may be included in the body of the report or as a footnote. If the report is intended for decision-makers, the calculations may be included in an annex. Evidence that necessary disaggregations were included in the sample size calculation such as through the selected design effect should be presented in the report. This may be documented in an annex or in the body of the report.</p> <p>For <b>qualitative studies</b>, a number of characteristics such as the timing of the study, the stakeholders targeted to be included in the study, the characteristics of the stakeholders included, the characteristics of focus group members, and the reason why the stakeholders were selected may be described in the report. Participants should be selected because they are likely to generate useful data for the study. Researchers/evaluators should provide a description of the sampling frame and potential issues with it, if any. This should include an explanation of how the participants were selected, whether these participants were the most appropriate to provide access to the type of knowledge sought by the study, whether there was a point at which incoming data produced little or no new information (saturation) as well as any discussions around recruitment, such as why some people might have chosen not to take part in the study. Where applicable, there should also be a discussion around the intended sample size with justification as well as discussion of the achieved sample size. Evidence that the sample was designed to be sufficiently representative of the target populations should be presented in the report. This may be documented in an annex or in the body of the report.</p> <p>A study using a <b>systems thinking approach</b> requires accounting for the interrelationships of all variables and the perspectives of all relevant stakeholders within the bounded systems, including those that are not statistically significant or are found to be outliers.</p> <p>"Partial" score should be given if only some of these details were discussed or presented.</p>	<p><a href="#">JPAL's Running Randomized Evaluations</a>, page 271; <a href="#">CASP</a>, Qualitative Checklist, page 3; <a href="#">EGRA Toolkit, 2<sup>nd</sup> Edition</a>, pages 117; <a href="#">StataCorp's Survey Data Reference Manual</a>, page 3; <a href="#">BE2 Guidance Note on Qualitative Research</a>, pages 40-41.</p> <p><i>Additional sources:</i> <a href="#">EGRA Toolkit, 2<sup>nd</sup> Edition</a>, pages 120 and 175; <a href="#">UIS Handbook on Measuring Equity in Education</a>, page 74</p>

Principle of Quality	Question	Score	Descriptor	Source
Cultural Appropriateness: Study questions informed by local stakeholders	[11] Does the report list the steps taken to ensure that study questions and methodology are informed by local stakeholders, culturally relevant, contextually appropriate, gender-sensitive, and inclusive as appropriate? <b>[All study types]</b>	Yes / No / Partial	The study questions and methodology selection should be informed by relevant local stakeholders. This could be done during in-country design workshops as well as through meetings with the government officials and/or other relevant stakeholders. Where appropriate to the study purpose and the context, the study should use inclusive, gender-sensitive, and participatory research methods.  Studies using <b>systems thinking approaches</b> should ensure that all possible perspectives from within each bounded system are reflected in the report.  “Partial” score could be given if the methodology is contextually appropriate but was not informed by relevant local stakeholders. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">ADS 201 sae; Save the Children’s Gender and Power (GAP) Analysis Guidance.</a>
Cultural Appropriateness: Data collection tools developed with participation of local stakeholders	[12] Does the report demonstrate that data collection tools were developed/adapted with participation of relevant local stakeholders, were piloted with representatives of the target populations and revised as needed, are culturally appropriate, gender-sensitive, and inclusive, as appropriate? <b>[All study types]</b>	Yes / No / Partial	The report should describe whether tools have been developed to suit the local context, such as whether the tool was developed by international experts and then merely translated into a local language or whether local knowledge has been used effectively in the adaptation of the tool to reflect resources relevant to the context, such as including support from host country experts. Quality control of translators is recommended. Where appropriate, the study should use inclusive, gender-sensitive, and participatory methods.  Researchers/evaluators should describe if respondents used to pilot the data collection tools were similar to the target population of the study. Researchers/evaluators should describe whether they used the results of the pilot to revise data collection tools prior to data collection. While piloting and revising the tools is a step to achieving validity, it is included as a question under the cultural appropriateness principle of quality since a tool cannot be valid if it is not first culturally appropriate.  “Partial” score could be given if some but not all tools suit the local context or if the report mentions that piloting was done but not with whom or how the results were used. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , page 20; <a href="#">EGRA Toolkit, 2nd Edition</a> , page 92.
Cultural Appropriateness: Findings/conclusions/recommendations validated with local stakeholders	[13] Does the report list steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders and incorporate stakeholder feedback in the report? <b>[All study types]</b>	Yes / No / Partial	Findings, conclusions, and recommendations must be communicated to the appropriate audiences in a culturally and contextually suitable way prior to finalization of the report, in order to validate accuracy of conclusions and inform recommendations. Stakeholders should have an opportunity to provide feedback on the findings before they are finalized in the report, and this feedback should be accounted for in the report. Steps to validate these with local stakeholders may include in-country presentations and workshops conducted during the study (instead of as dissemination workshops after the study was concluded).  “Partial” score could be given if the steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders are described but the stakeholder feedback was not incorporated into the report. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">EGRA Toolkit, 2nd edition</a> , page 122.  Additional source: <a href="#">BE2 Guidance Note on Qualitative Research</a> , page 61

Principle of Quality	Question	Score	Descriptor	Source
Cultural Appropriateness: Data collection and analysis allows for disaggregation by locally relevant stratifiers	<p>[14] Was the study designed to consider locally relevant stratifiers, such as political, social, ethnic, religious, geographical, sex/gender, disability status, displacement status, socio-economic status, and/or other relevant phenomena, during data collection and analysis?</p> <p><b>[All study types]</b></p>	Yes / No / Partial / N/A	<p>The extent to which a study takes into account locally relevant stratifiers has considerable bearing on the study's design, its analytical strategy and the interpretation of its findings. Designing a study to take into account locally relevant stratifiers might include a sample design which includes different groups. The analysis being informed by locally relevant stratifiers might include making cross-cultural or cross-linguistic comparisons part of the analytical strategy or ensuring that knowledge of the local context is used in the interpretation of differential effects between groups. The choice of stratifiers included in the study depends upon the study purpose and study questions, the study population and target sample, and the context in which the study takes place.</p> <p>“Partial” score should be assigned when the study is purposeful with considering one stratifier, such as gender, in data collection or considering variable impacts on one stratifier, such as gender, but not any other stratifiers. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, page 20.</p>
Ethics: Protection of human subjects is integrated in the study	<p>[15] Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report?</p> <p><b>[All study types]</b></p>	Yes / No / Partial	<p>It is vital that from the inception of a study to the dissemination of a report, all studies adhere to the highest ethical standards and protect the human subjects involved. USAID requires that evaluations are conducted to the highest ethical standards. As such, research and evaluations must be ethical, fair, and take into consideration cultural and contextual factors that may influence findings or how the findings are used. Research and evaluations with human subjects must adhere to the three key principles of the Common Rule: “respect for persons,” which refers to informed consent, privacy, confidentiality, and special protections for those with diminished autonomy, “beneficence,” which refers to “do no harm,” maximizing possible benefits and minimizing potential harm, and “justice,” which refers to the equitable distribution of research/evaluation burdens and benefits. USAID-funded studies must receive IRB approval (see the next question) and include “a meaningful informed consent procedure for research subjects.” All members of the study team are responsible for knowing and understanding the foundations of ethical research and ensuring that risks to human subjects are mitigated and that no harm is done to children, vulnerable populations, or the study participants as a result of the study. Additionally, all USAID-funded research must adhere to USAID policies regarding protection of human subjects. Reporting and referral protocols should be developed and used to ensure the study team knows how to report issues and/or refer children and adults for further support if the need emerges during the study. The steps taken to integrate ethical principles of protection of human subjects, to mitigate risks, and to ensure no harm should be documented in the report or in an annex. Informed consent/assent protocols should be included in an annex.</p> <p>“Partial” score could be given if the report mentions informed consent/assent but does not mention the steps taken to mitigate risks to participants. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">USAID Evaluation Policy</a>, page 9; <a href="#">USAID Ethics in Research and Evaluation in the Education Sector</a>, pages 1-2.</p> <p>Additional sources: <a href="#">BE2 Guidance Note on Qualitative Research</a> pages 56, 58-59; <a href="#">Save the Children's Gender and Power Analysis Guidance</a>, step 4</p>

Principle of Quality	Question	Score	Descriptor	Source
Ethics: Research clearances were obtained	[16] Was/were research clearance(s) appropriate to the study obtained and documented prior to starting data collection? <b>[All study types]</b>	Yes / No / Partial	USAID-funded studies which involve human subjects must consult an IRB and receive IRB approval. Studies funded by other donors must follow relevant requirements. Studies being conducted in other countries must follow the local research clearance and IRB/ERC requirements in the country of the study. This includes seeking and documenting “exempt” status as applicable. IRB/ERC approval or “exempt” status and local research clearance should be documented either in the report or in an annex.  “Partial” score could be given if the study mentions IRB approval but does not provide evidence that IRB approval or exemption was obtained. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">EGRA Toolkit, 2<sup>nd</sup> Edition</a> , page 13; <a href="#">USAID Ethics in Research and Evaluation in the Education Sector</a> , pages 1-5.  Additional source: <a href="#">BE2 Guidance Note on Qualitative Research</a> , page 57
Validity: Addressed the construct validity of the data collection tools	[17] Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated? <b>[All study types]</b>	Yes / No / Partial	In order to assess the validity of the measurement, it is important to consider whether or not the chosen indicators adequately capture the concepts being measured or whether there are other dimensions central to the concepts that are being ignored, such as a labor market condition index that ignores underemployment. “Partial” scores could be given if some, but not all key constructs or indicators, adequately captured the concepts being measured.  “Partial” score could be given if some but not all key constructs or indicators adequately capture the concepts being measured. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , page 24.
Validity: Addressed ecological validity of findings	[18] Is the report open and clear about how the act of doing the study may have biased the findings? <b>[All study types]</b>	Yes / No / Partial / N/A	Evaluators/researchers might discuss in the report whether findings could have been influenced by the process of research itself (ecological validity) or whether participants may have changed their behavior in response to their perception of the evaluators’ objective (response bias), such as when the treatment group works harder than normal in response to being part of an evaluation (Hawthorne effects). This might include discussions about whether the implementer may have brought in irreproducible energies that account for the success of a pilot but that might be absent in a scale-up. Note that the tendency of participants to give an answer to a question that is in line with social norms even if this does not accurately reflect their experience (social desirability bias) is not relevant for this question.  “Partial” score could be given if the report mentions that the study may have biased the findings but does not explain how. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , page 25.

Principle of Quality	Question	Score	Descriptor	Source
Validity: Address the credibility of the findings	<p>[19] Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections?</p> <p><b>[All study types]</b></p>	Yes / No / Partial	<p>The report should balance the presentation of the findings with a discussion contextualizing them and/or addressing how they might be affected by methodological decisions. This discussion might include broaching alternative explanations for the findings. If some findings yield inconsistencies with others, this should be discussed as well.</p> <p>For <b>qualitative studies</b>, credibility establishes that the data, analyses, and interpretation are truthful. Approaches to establishing credibility include triangulation, referential adequacy (such as collecting materials to check interpretation against official materials), member checking, peer debriefing, and structural corroboration (such as negative case analysis to test alternative interpretations).</p> <p>“Partial” score could be given if the report provides evidence that some but not all the findings are credible. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, page 17.</p> <p>Additional sources: <a href="#">BE2 Guidance Note on Qualitative Research</a>, page 61; <a href="#">Chilisa, 2020</a>; <a href="#">Lincoln and Guba, 1985</a>)</p>
Validity: Addressed the external validity or transferability of findings to other contexts	<p>[20] Does the report address the external validity (for quantitative studies) or transferability (for qualitative studies) of findings?</p> <p><b>[All study types]</b></p>	Yes / No / Partial / N/A	<p><b>Quantitative findings</b> are externally valid when they are valid in contexts other than those the evaluation was conducted in. Thus, researchers/evaluators may discuss the local conditions that would make it replicable in a different context. <b>Qualitative findings</b> are transferable if the findings are situated in their specific context so that readers may extrapolate or relate the findings within one context to possibilities in other contexts. The report should balance the presentation of the findings with a discussion contextualizing them.</p> <p>“Partial” score could be given if the external validity or transferability of some, but not all key findings, are discussed in the report. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p> <p>“N/A” score could be given in case this study did not intend to have data from a sample extrapolated to a population.</p>	<p><a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a>, Checklist (p.29); <a href="#">Chilisa, 2019</a>, page 216.</p> <p>Additional sources related to transferability in qualitative research: <a href="#">Lincoln and Guba, 1985</a>; <a href="#">Shenton, 2004</a>; <a href="#">Williams and Morrow, 2009</a></p>
Validity: Confidence intervals reported around point estimates	<p>[21] If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates?</p> <p><b>[Quantitative]</b></p>	Yes / No / Partial / N/A	<p>Statistical data should be presented to be useful, which includes providing sufficient detail so that a reader may evaluate the precision and certainty of the results. USAID recommends that the confidence interval (margin of error) around point estimates be reported when presenting the findings from statistical samples. Reporting standard errors is also important to evaluate the results by capturing uncertainty around the estimates.</p> <p>“Partial” score could be given if the standard errors or confidence intervals are reported for some but not all point estimates. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p> <p>“N/A” score should be given if the study does not use inferential statistical methods.</p>	<p><a href="#">ADS 201sae</a>.</p> <p>Additional sources: <a href="#">What Works Clearinghouse Procedures and Standards Handbook Version 5.0</a> page 161; <a href="#">Remler and Van Ryzin, 2021</a>, pg. 180.</p>

Principle of Quality	Question	Score	Descriptor	Source
Reliability: Steps taken to ensure that data were reliably collected	[22] Does the report document the steps taken to ensure that data were collected with a high degree of reliability?  <b>[All study types]</b>	Yes / No / Partial	USAID recommends that data collection procedures be documented in writing to ensure that the same procedures are followed each time. The report may describe the use of data quality assurance checks such as accompaniments, back-checks, and scrutiny, and these may have been conducted through spot-checking or for all questions in the data collection form. In case of paper-and-pencil data collection, double data entry report and/or double manual verification may also be mentioned in the report. Steps used in <b>qualitative studies</b> may include audio recording, videotaping, and transcribing interviews.  In studies using a <b>systems thinking approach</b> , system diagrams that are developed during analysis must be developed with stakeholder input to ensure their perspectives are included and accurately represented.  “Partial” score could be given if steps to ensure the reliability of some, but not all data collected, are described. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">ADS 201sae.</a>
Reliability: addressed internal reliability/consistency of instruments	[23] If applicable, was internal consistency of the instrument(s) established and documented?  <b>[Quantitative]</b>	Yes / No / Partial / N/A	Instruments which measure a scale comprised of a set of items or indicators must ensure that all the items go together to reflect the same thing and are internally consistent. Internal consistency of an instrument may be determined through methods such as split-half reliability or Cronbach’s alpha. The most widely used measure is Cronbach’s alpha, and a minimum alpha coefficient of 0.7 is considered acceptable.  “Partial” score could be given if a report mentions that internal consistency of an instrument is established but does not provide details about what method was used. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.  “N/A” score should be given for studies which do not use multi-item instruments where multiple items are intended to measure the same variable.	<a href="#">EGRA Toolkit, 2nd Edition, page 93-94; Remler and Van Ryzin, 2021, pg. 135.</a>
Reliability: Inter-rater reliability was established	[24a] For studies where data is collected by a team, was inter-rater reliability established and documented?  <b>[Quantitative]</b>	Yes / No / Partial / N/A	In survey or assessment studies collecting data with multiple enumerators, it is important for enumerators to agree on how they mark the data. This requires regular measurement of the rate of agreement between enumerators.  “Partial” score could be given if the report mentions that inter-rater reliability was established but does not explain how. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.  “N/A” score should be given for qualitative studies or for survey/assessment studies in which data was not collected by multiple enumerators.	<a href="#">EGRA Toolkit, 2nd Edition, page 89.</a>



Principle of Quality	Question	Score	Descriptor	Source
Reliability: Inter-coder reliability was established	[24b] If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team? <b>[Qualitative]</b>	Yes / No / Partial / N/A	In <b>qualitative studies</b> analyzing data through a team effort, it is important for all team members to agree upon how data is coded. The study should describe how inter-coder disagreement was measured and addressed.  "Partial" score could be given if the report mentions that inter-coder reliability was established but does not explain how. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.  "N/A" score should be given for quantitative studies, for qualitative studies which do not incorporate coding in the method, or for qualitative studies which were not coded by multiple team members.	<a href="#">Saldaña, 2021</a> , page 52.
Reliability: Target and actual sample sizes reported and non-responses bias discussed	[25] Does the report adequately address missing data/non-response? <b>[All study types]</b>	Yes / No / Partial	Researchers/evaluators should report the target number of respondents, the number of respondents reached, and the number of respondents who were included in the data analysis. This includes addressing non-response in <b>qualitative studies</b> . For <b>quantitative research</b> , the report may also mention using post-stratification to adjust weights for non-response or other strategies for addressing missing data.  "Partial" score could be given if information about valid responses is provided to some, but not all data used in the findings. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.	<a href="#">What Works Clearinghouse Procedures and Standards Handbook Version 3.0</a> , page D.4.  Additional sources: <a href="#">What Works Clearinghouse Standards Handbook Version 4.0</a> page 65; <a href="#">BE2 Guidance Note on Qualitative Research</a> page 46
Openness and Transparency: Open about limitations to the study design and to implementing the study	[26] Is the report open and clear about limitations <b>inherent to the study design and with its implementation?</b> <b>[All study types]</b>	Yes / No / Partial	Limitations to the implementation of the study should be clearly presented. Clarity around study limitations is particularly important if they directly affect the evaluator's/researcher's ability to credibly and effectively answer a study question or affect validity, reliability, or generalizability of the findings (i.e., if data collection was successful but more expensive or inconvenient than anticipated, it is not a limitation). An example of limitations inherent to the study design is a design which cannot produce generalizable results. An example of limitations due to the implementation of the study could be issues faced during data collection.  USAID Evaluation Policy requires that evaluation reports address methodologically common limitations, such as methods that do not allow for generalizability.  "Partial" score could be given if the report mentions limitations without discussing them in detail. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , page 17.  Additional sources: <a href="#">ADS 201 mah; Blaikie and Priest, 2019</a> , page 15; <a href="#">Greener, 2018</a> , page 568



Principle of Quality	Question	Score	Descriptor	Source
Openness and Transparency: Open about how the intervention affects the study	[27] For evaluations, is the report open and clear about study limitations due to issues with the <b>implementation of the intervention</b> being evaluated? <b>[Evaluations]</b>	Yes / No / Partial / N/A	In <b>evaluation research</b> , limitations to the implementation of the intervention being evaluated should be clearly presented, such as delays or changes that may compromise the integrity of the evaluation design.  “Partial” score could be given if the report mentions limitations without discussing them in detail. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.  “N/A” score should be given to studies that do not evaluate a specific intervention.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , page 17.
Openness and Transparency: Open about potential biases due to the study team composition	[28] Is the report open about potential biases due to the study team composition? <b>[All study types]</b>	Yes / No / Partial	In evaluation research, USAID encourages study teams to include at least one research/evaluation specialist, partner country team member, and a team leader who is external to USAID. USAID also requires that evaluation team members certify their independence by signing statements <b>disclosing any conflict of interest or fiduciary involvement with the project or program</b> they will evaluate. It is expected that an evaluation will indicate that such forms, or their equivalent, are on file and available or are provided in an evaluation annex. Research and other non-evaluation studies should follow the same guidance.  “Partial” score could be given if some, but not all, these recommendations are followed. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">BE2 Assessing the Strength of Evidence in the Education Sector</a> , Checklist (page 29-29).  Additional source: <a href="#">USAID Evaluation Policy</a> , page 8
Openness and Transparency: Methodology explained in detail	[29] Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made? <b>[All study types]</b>	Yes / No / Partial	USAID requires that a study report identify the study design, data collection methods, and data analysis techniques used. It is common to include the methodology description in the body of the report under a methodology section with a longer and more detailed methods annex.  The description of methods must indicate how respondents were selected, how data was collected from the participants, who participated in the study, and detailed information on the kinds of analyses that were conducted (e.g., correlations, regressions, content analysis, or pattern analysis).  Researchers/evaluators using a <b>systems thinking approach</b> must determine the boundaries of a study and the key actors (agents) within the system boundaries. Researchers/evaluators should explain how those boundaries are determined, who made those decisions, and the implications for the study.  “Partial” score could be given if some, but not all elements mentioned (design, data collection methods, and data analysis techniques) were described in sufficient detail. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.	<a href="#">ADS 201 maa</a> .  Additional sources: <a href="#">USAID Evaluation Policy</a> page 8

Principle of Quality	Question	Score	Descriptor	Source
Openness and Transparency: Cost analysis	<p>[30] For impact evaluations, is a cost analysis of the intervention being evaluated included in the report?</p> <p><b>[Impact evaluations]</b></p>	Yes / No / Partial / N/A	<p>USAID requires all <b>impact evaluations</b> to include a cost analysis of the evaluated intervention(s). The findings of the cost analysis should be included in the findings section of the impact evaluation report and should include elements that are useful for decision-making. Required details that must be included in the report include:</p> <ul style="list-style-type: none"> <li>• Details about the intervention, such as the theory of change, the model implemented, dosage details (contact time), critical components of the intervention, sequence of activities (if important to the intervention), when and by whom the intervention was implemented, and the funder.</li> <li>• Details about the beneficiaries: who and where they are, including geography, age and sex, marginalization status, and other relevant details.</li> <li>• Cost estimates: what is included/not included and why, whether recurrent and non-recurrent costs are separated, whether contributions were costed out or listed alongside final estimates, whether shared costs were included, and from whose perspective the costs are estimated (e.g., USAID, partner government, implementer, or beneficiary). Assumptions and data limitations must be noted.</li> <li>• Computation: how cost estimates were computed, including assumptions and the computations in an annex in sufficient detail to allow for replication of calculations.</li> </ul> <p>"Partial" score could be given if some but not all of these elements are included in the report. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.</p> <p>"N/A" score should be given for studies that do not include an impact evaluation.</p>	<p><a href="#">USAID Evaluation Policy</a>, page 2; <a href="#">USAID Cost Analysis Guidance for USAID-Funded Education Activities</a>, page 79.</p>
Cogency: Answers to all study questions, including sub-questions, included	<p>[31] Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings?</p> <p><b>[All study types]</b></p>	Yes / No / Partial	<p>USAID expects that the answers to all study questions, including any sub-questions, will be provided in the report. The executive summary must provide an accurate representation of the main elements of the report without adding any new material information or contradicting the report in any way. As such, it is recommended that all study questions/issues, including any sub-questions/issues, will be provided in the executive summary. Study findings should relate to the questions to ensure the findings are applicable to the study.</p> <p>"Partial" score could be given if the answers are provided in the report but not the executive summary. Other reasons to give a "partial" score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">ADS 201mah; E3 Sectoral Synthesis</a>, Checklist, question 17, page 145.</p> <p>Additional source: <a href="#">USAID Evaluation Report Template</a></p>

Principle of Quality	Question	Score	Descriptor	Source
Cogency: Written in a language adequate to its stated audience	[32] Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)? <b>[All study types]</b>	Yes / No / Partial	<p>Reports should be written in an accessible way to non-experts. Excessive use of research terminology is undesirable. The report should favor terminology that its intended audience is expected to be familiar with, and the language used in the report should be simple and straightforward. The report should be written so that it is accessible to non-native English speakers. Reports should make effective use of footnotes and annexes, such as to include technical details that are not relevant for the intended audience but necessary to demonstrate the quality of the study or to provide technical definitions and descriptions without interrupting the flow of the report narrative.</p> <p>“Partial” score could be given if some but not all sections of the report are written in a style that non-experts can understand or if some but not all technical jargon is explained. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<a href="#">USAID Evaluation Policy</a> , page 11.
Cogency: Recommendations are relevant, actionable, and based on findings	[33] If recommendations are made, are they specific, relevant, actionable, and based on the findings? <b>[All study types]</b>	Yes / No / Partial / N/A	<p>It is important that recommendations be practical for the intended audience, action-oriented, and specific as well as relevant to the study.</p> <p>“Partial” score could be given if some but not all recommendations meet these criteria. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p> <p>“N/A” score should be given to studies which are not intended to produce recommendations.</p>	<a href="#">USAID Evaluation Policy</a> , p11.  <i>Additional source: <a href="#">UNEG 2010</a>, page 6</i>
Cogency: The report is logically connected from start to finish	[34] Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations? <b>[All study types]</b>	Yes / No / Partial	<p>Well-designed studies make a logical connection between the study objective, questions, framework, methodology, and findings. In order to strengthen the study’s conclusion validity, USAID requires that findings be based on reliable quantitative and/or qualitative data, and that conclusions and recommendations should be based on these findings.</p> <p>USAID also encourages evaluators to present a clear progression from study questions to findings to conclusions to recommendations (if any) in their reports, such that none of a report’s conclusions and recommendations appear to lack grounding.</p> <p>Studies using a <b>systems thinking approach</b> should discuss findings, conclusions, and recommendations in such a way that their inter-relationships are clear. For studies using a systems thinking approach, the findings should address the role of the sub-systems and the agents in the system, their relationships, differences in perspectives, barriers as well as leverage points.</p> <p>The study team should present a clear progression from study questions to methodology (including methods decisions comprising data collection and analysis) to findings to conclusions to recommendations (if any) in their reports.</p> <p>“Partial” score could be given if there is a clear progression from the study questions to the methodology but not from the findings to the conclusions. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<a href="#">E3 Sectoral Synthesis</a> , Checklist, question 32, page 145.

Principle of Quality	Question	Score	Descriptor	Source
<p>Cogency:            Visuals are helpful for a non-technical audience to understand the findings</p>	<p>[35] Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings?   <b>[All study types]</b></p>	<p>Yes / No / Partial</p>	<p>Visuals must be used to facilitate understanding of the findings by intended audiences. As appropriate, visuals should be standalone, such that they are interpretable without the audience needing to read extra text. The visuals included should clearly support the findings.</p> <p>“Partial score” could be given if the report uses visuals to an insufficient extent. Other reasons to give a “partial” score are possible and up to the judgment of the reviewer.</p>	<p><a href="#">EGRA Toolkit 2<sup>nd</sup> Edition</a>, page 120.</p>

## ENDNOTES

- 
- <sup>i</sup> Norman Blaikie, *Designing Social Research*, 2<sup>nd</sup> Edition (Malde, MA: Polity, 2009), 40, 70, 74.
- <sup>ii</sup> Dahlia K. Remler and Gregg G. Van Ryzin, *Research Methods in Practice: Strategies for Description and Causation*, 2<sup>nd</sup> Edition (Los Angeles: Sage Publications, 2015), 5.
- <sup>iii</sup> Building Evidence in Education, “Assessing the Strength of Evidence in the Education Sector,” September 22, 2015, [https://www.edu-links.org/sites/default/files/media/file/BE2\\_Guidance\\_Note\\_ASE\\_0.pdf](https://www.edu-links.org/sites/default/files/media/file/BE2_Guidance_Note_ASE_0.pdf).
- <sup>iv</sup> Since many elements of doing a study overlap and must be addressed during all phases, this is a loose alignment.
- <sup>v</sup> <https://usaidlearninglab.org/resources/evaluation-report-template>
- <sup>vi</sup> These are not meant to be comprehensive descriptions of the different methodological groups. For more information about each design type, please refer to the BE2 “Assessing the Strength of Evidence in the Education Sector” guidance note or research manuals and other resources referenced throughout this document.
- <sup>vii</sup> The original evaluation quality assessment tool, developed as part of the Assessment of the Quality of USAID-Funded Evaluations in the Education Sector, 2013–2016, was released in 2017. This version updates the tool to be inclusive of research and reflect updates in USAID policy and international best practices.
- <sup>viii</sup> Arnold, Ross D., and Jon P. Wade. “A definition of systems thinking: A systems approach.” *Procedia computer science* 44 (2015): 669–678. <https://doi.org/10.1016/j.procs.2015.03.050>
- <sup>ix</sup> USAID. “Local Systems: A Framework for Supporting Sustained Development.” April 2014. <https://www.edu-links.org/sites/default/files/media/file/LocalSystemsFramework.pdf>
- <sup>x</sup> Kim, Daniel. “Introduction to Systems Thinking,” *The Systems Thinker*. 1999, <https://thesystemsthinker.com/introduction-to-systems-thinking/>
- <sup>xi</sup> Allen, Will. “Systems thinking,” *Learning for Sustainability*. Accessed 28 October 2022, <https://learningforsustainability.net/systems-thinking/>
- <sup>xii</sup> Coghlan, David, and Mary Brydon-Miller. “Systems Thinking.” In *The SAGE Encyclopedia of Action Research*, edited by Coghlan, David, and Mary Brydon-Miller, 753–754. London: SAGE Publications Ltd, 2014. <https://dx.doi.org/10.4135/9781446294406.n333>
- <sup>xiii</sup> Kim, Daniel. “Introduction to Systems Thinking,” *The Systems Thinker*. 1999, <https://thesystemsthinker.com/introduction-to-systems-thinking/>
- <sup>xiv</sup> Goodman, Michael. “Systems Thinking: What, Why, When, Where, and How?,” *The Systems Thinker*. Accessed 31 October 2022, <https://thesystemsthinker.com/systems-thinking-what-why-when-where-and-how/>
- <sup>xv</sup> Allen, Will. “Systems thinking,” *Learning for Sustainability*. Accessed 28 October 2022, <https://learningforsustainability.net/systems-thinking/>
- <sup>xvi</sup> “Research must do no harm: new guidance addresses all studies relating to people.” *Nature* 606 (2022): 434. <https://doi.org/10.1038/d41586-022-01607-0>