USAID
FROM THE AMERICAN PEOPLE

THE STATE OF QUALITY IN USAID-FUNDED EDUCATION RESEARCH AND EVALUATIONS:

Results from the 2023 Study Quality Review

This publication was produced for review by the United States Agency for International Development (USAID). It was prepared by Michelle Solorio and Gaëlle Simon for EnCompass LLC and its partner MSI, a Tetratech 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.

CONTENTS

LIST OF EXHIBITS

ACKNOWLEDGEMENTS

This report is made possible by the support of the American People through the United States Agency for International Development (USAID). The contents of this report are the sole responsibility of EnCompass LLC and do not necessarily reflect the views of USAID or the United States Government.

ACRONYMS AND ABBREVIATIONS

ADS	Automated Directives System
ASQ Tool	Assessment of Study Quality Tool
BE2	Building Evidence in Education
CASP	Critical Appraisal Skills Programme
DEC	Development Experience Clearinghouse
DEEP	Data and Evidence in Education Program
DQA	data quality assessment
E3/ED	Bureau for Economic Growth, Education, and Environment, Office of Education
EE	Eastern Europe
EGRA	Early Grade Reading Assessment
EiCC	education in conflict and crisis
ERC	Ethical Review Committee
GAP	Gender and Power Analysis
ICR	inter-coder reliability
IES	Institute of Education Sciences
IRB	Institutional Review Board
IRR	inter-rater reliability
LAC	Latin America and the Caribbean
MDES	Minimum Detectable Effect Size
MDI	Minimum Detectable Impact
MENA	Middle East and North Africa
MSI	Management Systems International
NA	not applicable
RERA	Rapid Education and Risk Analysis
TOC	table of contents
TWP	Technical Work Plan
UNEG	United Nations Evaluation Group
USAID	United States Agency for International Development
YWFD	Youth Workforce Development

EXECUTIVE SUMMARY

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 develop and apply a tool to appraise the quality of USAID-funded evaluations in the education sector. In 2022, USAID commissioned a team under the Data and Evidence in Education Program (DEEP) to revise the tool and conduct a second review. To support USAID's quality assessment of education evaluations and research commissioned by the Agency since 2017 and get feedback about the revised Assessment of Study Quality (ASQ) Tool, this study seeks to answer the following questions:

1. What is the overall quality of study reports published on the Development Experience Clearinghouse (DEC) since 2017?
    a. What are the strongest and weakest principles of quality of study reports published on the DEC since 2017?
    b. How has the quality of study reports changed since the previous review?
2. What is the quality of study reports published since 2017 by principle of quality?
3. To what extent do researchers and evaluators find the ASQ Tool relevant and user-friendly?

METHODS AND LIMITATIONS

163 REVIEWERS BASED IN 39 COUNTRIES SIGNED UP TO PARTICIPATE.

In collaboration with the USAID Center for Education, DEEP revised the ASQ Tool. The revised tool is composed of 35 items under eight principles of quality (conceptual framing, robustness of methodology, cultural appropriateness, ethics, validity, reliability, openness and transparency, and cogency) grouped by study phase (study design, study implementation, and report writing). From January 23 to February 6, 2023, a group of education researchers and evaluators used an online version of the tool to review 122 of 214 study reports published on the DEC between January 2018 and December 2022. Reports were scored as meeting "Minimum Adequacy" or "Not Adequate" for each principle based on reviewer responses to items in the tool. DEEP calculated the percentage of reports that met "Minimum Adequacy" in each principle and the percentage of reports that received Yes, No, and Partial responses for each item. DEEP compared the 2023 principle-level results to the 2017 results.

The reviewer composition likely biased the results given that the reviewers were researchers, evaluators, and other experts who understand the technical jargon. In addition, reviewers used a draft version of the ASQ Tool which potentially skewed the data or led to non-response for items that were not clearly defined. Although the study was designed as a census of study reports published between January 2018 and December 2022, not all volunteers completed their reviews. Thus, the final sample may not be truly representative of all reports published on the DEC despite the random assignment to reviewers. The results of this study can only serve to identify the relative strengths and weaknesses of USAID-funded education studies included in this review since the determination of "Minimum Adequacy" is not based on theory. Finally, the comparison of the 2017 and 2023 results cannot be considered a true comparison due to differences in tools and methodology.

FINDINGS

Eighty-six percent of reports reviewed (n = 105) met "Minimum Adequacy" for at least two principles of quality and over half of the reports (n = 64) met "Minimum Adequacy" for five or more principles. Report quality has improved since 2017. Cogency was the strongest principle of quality in both reviews. Cultural appropriateness remained among the weakest, but experienced gains in the percentage of reports meeting minimum adequacy. The biggest gains were in validity (14.7 percentage point increase), openness and transparency (11.7 percentage point increase), and cultural appropriateness (11.2 percentage point increase). Robustness of methodology decreased by 24.8 percentage points, followed by conceptual framing (3.9 percentage point decrease).

The strongest principle of quality, met by 86.1 percent of reports, is cogency, and the third-strongest principle, met by 60.7 percent of reports, is openness and transparency. Both principles are part of the report writing phase. The weakest principles are part of the study design phase: ethics (met by 36.9 percent of reports), cultural appropriateness (met by 40.2 percent of reports), and robustness of methodology (met by 40.2 percent of reports). At the item level, there was low adherence to items related to biases in all study phases.

Reviewers found the tool relevant and easy to use for designing and implementing studies. Reviewers are likely to use the tool in their own work and recommend it to their colleagues.

CONCLUSIONS

The principles that need the most support are ethics, cultural appropriateness, and robustness of methodology. While cultural appropriateness remains weak since the previous review, it has improved. This may indicate that more attention is being paid to cultural appropriateness by those designing, implementing, and writing about education studies. The two principles related to study implementation, validity and reliability, are stronger than robustness of methodology, cultural appropriateness, and ethics, three of the four principles under study design. This could mean that study implementation is stronger than study design, or that report writing is not accurately capturing the items in all principles of quality.

Across each study phase, items related to biases were among the weakest, based on evidence provided in the reports. Items related to biases received lower scores in design, implementation, and report writing. This indicates that planning for and mitigating biases is challenging for those conducting education studies, as is reporting on bias. It also highlights how connected all the study phases are.

Although cogency is the strongest principle of quality, each finding may point to issues of quality in the report writing phase. Those who designed and implemented the studies may have addressed all of the items under each principle of quality, but if the report provided no evidence, the reviewers could not conclude that the criteria were met. This could reflect reviewer participation bias because those who participated in the study are experts whose experience may have allowed them to make connections and assumptions about the items in the ASQ Tool that are not clearly reported.

The reviewers' feedback indicates that the tool is relevant and user-friendly, although the audience is an important factor when deciding whether to recommend the tool for use. Reviewers suggested that the tool is more useful for designing and writing about research and evaluation than for implementing research and evaluation. Most reviewer feedback was positive, though there were contradictions

regarding length. Reviewers indicated that the explanations and examples included in the tool made it more user-friendly, but also mentioned that the tool was too long and technical.

RECOMMENDATIONS

As part of this review, DEEP developed a separate Internal Action Plan which includes a comprehensive set of recommendations for USAID. For the sake of transparency and accountability, a summary of these recommendations follows.

DEEP proposes targeted socialization of the ASQ Tool with USAID Missions, Regional Bureaus, implementing partners (IPs) and other partners in the next six months. This includes developing materials to help Missions and Regional Bureaus communicate the tool and discuss expectations for its use with IPs, pinning the tool to the EducationLinks home page to increase awareness and enable access, and developing materials to support understanding of the tool. In 6 to 12 months, DEEP proposes developing and holding a series of targeted training sessions for each study phase that address the applicable principles of quality for the study phase. These training sessions would focus on addressing weaknesses identified during this review and supporting overall understanding about how to design, implement, and write about high-quality studies. After one year, DEEP proposes conducting a rapid study of the ASQ Tool socialization efforts to provide insight into additional requirements to support continued use of the tool. DEEP also proposes engaging with researchers and evaluators working in education subsectors through virtual interactive sessions in which participants discuss their approaches to and challenges with applying practices under each principle of quality. In addition, DEEP proposes developing additional materials to support ASQ Tool understanding, such as annotated examples of high-quality report sections.

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 (MSI) to conduct an evaluation synthesis, the first step of which was to assess the quality of USAID-funded evaluations in the education sector. Through this request, the Office of Education intended to curate, analyze, and disseminate the robust evidence generated by USAID related to the objectives laid out in the Agency's 2011 Education Strategy. One key result of this study was a tool used 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.[1] Another result was the selection of a subset of evaluations that passed the minimum quality criteria for inclusion in the synthesis study.

The 2016–2017 study reviewed impact and performance evaluations[2] and five Rapid Education and Risk Analysis (RERA) reports. The study included evaluations that spanned the Agency's three Education Strategy Goals and all six USAID regions, and those conducted in countries ranging from low- to upper-middle-income and countries that were and were not in crisis and conflict. The review used a participatory approach to involve evaluation practitioners in the process. The Office of Education used findings from the study to identify specific gaps in the quality of evaluations.

In 2022, USAID commissioned a team under the Data and Evidence in Education Program (DEEP) to revise the evaluation quality tool to capture updated international best practices in research and evaluation and updated USAID guidelines and policies. The commission also included a mandate to revise the tool to be more broadly applicable to research and evaluation in education and the social sciences. The revised tool was renamed the Assessment of Study Quality (ASQ) Tool.

The 2022 commission included a second review of the quality of USAID-funded studies in education. In January 2023, the DEEP team conducted this second review using the ASQ Tool. The purpose of the review is to support USAID's assessment of the quality of education evaluations and research commissioned by the Agency since 2017, and to get feedback about the tool, which will help address any outstanding issues and continue to improve the tool for future use. The review seeks to answer the following research questions:

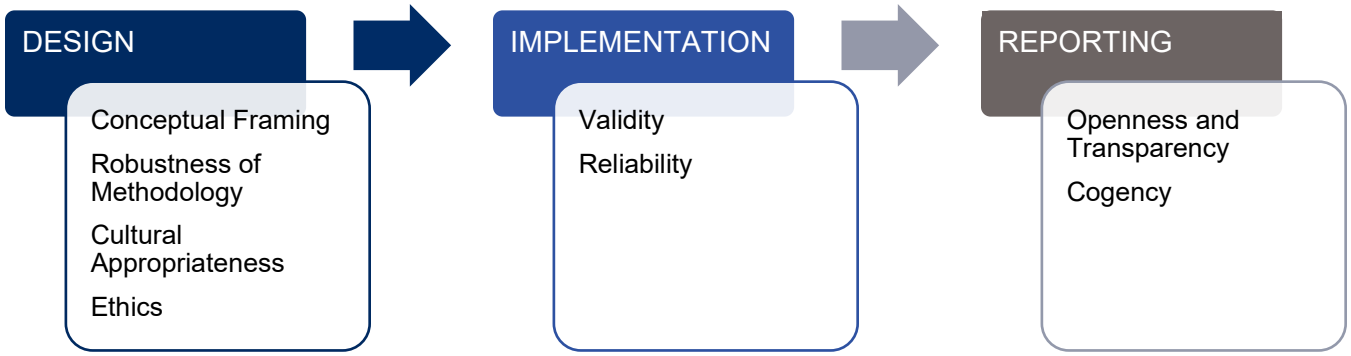
1. What is the overall quality of study reports published on the Development Experience Clearinghouse (DEC) since 2017?
    a. What are the strongest and weakest principles of quality of study reports published on the DEC since 2017?
    b. How has the quality of study reports changed since the previous review?
2. What is the quality of study reports published since 2017 by principle of quality?
3. To what extent do researchers and evaluators find the ASQ Tool relevant and user-friendly?

[1] The original Assessing the Quality of Education Evaluations Tool is published on the EducationLinks website and on the Learning Lab website.
[2] "Evaluations" were defined in accordance with the USAID Evaluation Policy.

METHODS

In collaboration with the Center for Education, the DEEP team revised the ASQ Tool using the same Building Evidence in Education (BE2) framework to assess the principles of quality used for the previous tool. The revised ASQ Tool composed of 35 items under eight principles of quality grouped by study phase. These principles are conceptual framing, robustness of methodology, cultural appropriateness, and ethics under the study design phase; validity, and reliability under the study implementation phase; and openness and transparency, and cogency under the report writing phase (see Exhibit 1). The draft version of the ASQ Tool used during the review is included in Annex A.[3]

Exhibit 1: Principles of quality by study phase



A group of researchers, evaluators, international partners, consultants, and academic partners were recruited to participate in the voluntary collaborative review process using an online version of the revised ASQ Tool (Annex B). The review ran from January 23 to February 6, 2023.

USAID invited potential participants based on the collective network of USAID, MSI, and EnCompass, and asked them to participate in the review and forward the invitation to their own networks, paying special attention to local researchers and evaluators in countries where USAID operates. A total of 163 reviewers signed up and many offered to review multiple reports. Reviewers were asked to indicate the education topic they would be most comfortable reviewing, and the team used this information to randomly assign reports—checking each assignment to ensure there would be no conflicts of interest.

163 REVIEWERS BASED IN 39 COUNTRIES SIGNED UP TO PARTICIPATE IN THE REVIEW.

Reviewers received an asynchronous orientation module, were invited to participate in virtual orientation Q&A sessions and were provided ongoing support from the DEEP team. They used an online version of the tool, which was reordered to follow the USAID Evaluation Report Template table of contents (TOC) as closely as possible.

[3] The ASQ Tool has since been revised, taking into consideration feedback from reviewers who participated in this study.

DATA

The data comprised study reports published on the DEC between January 2018—after the previous review was complete and the tool was available to the public—and December 2022. The material included evaluation reports, research reports, special evaluations, other USAID supported studies and documents, and assessment studies about education. Reports about studies that did not collect primary data, such as desk reviews, were not included to match the scope of the tool. This resulted in a list of 214 reports. Given the number of volunteers who signed up to participate in the review and the number of reviewers who volunteered to review multiple reports, the team determined that all reports could be assigned for review as a census. The list of reports assigned for review, including an indication of which reviews were completed, is included in Annex C.

ANALYSIS

To analyze the review data, the team first identified a centrally important question for each principle of quality. The questions were based upon the core meaning of the principles. Not all principles had a single question that could serve as a centrally important question; for some principles, two questions served as the centrally important question while for two principles no centrally important questions could be identified. If a reviewer scored a report as "No" for the centrally important question (or at least one of the centrally important questions), meaning it did not meet the criteria for that item, the report automatically received a "Not Adequate" score for that principle of quality.

Exhibit 2: Centrally important questions for each principle of quality

PRINCIPLE OF QUALITY	CENTRALLY IMPORTANT QUESTION	RATIONALE
STUDY DESIGN PHASE		
Conceptual Framing	Are clear study questions that are appropriate to the stated purpose of the study included in the report?	Clear questions that are appropriate to the purpose of the study are foundational to guiding each study. If questions are not appropriate to the purpose or clearly stated, the quality of the study is compromised from the start.
Robustness of Methodology	Is the methodology appropriate for answering posed study questions?	If the methodology is not appropriate to the study questions, the remainder of the study is no longer appropriate. This also follows BE2 guidance (p. 18) in assessing robustness of methodology, which points to identifying whether the design and methods are appropriate to the study questions.
	Does the counterfactual meet standards of rigor? (for experimental/quasi-experimental studies only)	If the counterfactual is not set up correctly, all the other questions on the design in the experimental studies are moot.
Cultural Appropriateness	Does the report list steps taken to ensure that study questions and methodology are informed by local stakeholders, culturally relevant, contextually appropriate, gender-sensitive, and inclusive as appropriate?	Study questions and methodology are foundational to study design and implementation. If the foundation is not culturally appropriate, it will be reflected in the implementation of the study and in the findings.
Ethics	Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report?	The heart of this principle of quality is ensuring that research is conducted ethically.
STUDY IMPLEMENTATION PHASE		
Validity	N/A	No questions were determined to be central to the principle.
Reliability	N/A	No questions were determined to be central to the principle.
STUDY IMPLEMENTATION PHASE		
Openness and Transparency	Is the report open and clear about limitations inherent to the study design and with its implementation?	This is a defining feature of transparency and addresses limitations. It follows BE2 guidance (p. 17), which points to limitations as the first step in assessing openness and transparency.
Cogency	Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations?	This item gets to the core of the principle of quality, which is that a high-quality study provides a clear and logical thread through the entire report.

For each question, responses were scored using the following point system:

- "Yes" response = 2 points;
- "Partial" response = 1 point; and
- "No" response = 0 points.

When a question was marked as "Not Applicable," this flagged that the total points possible for the appropriate principle of quality would be different for that report.

The team then summed the points earned for each report on each principle of quality, using the following formula:

$$\text{Total Points Earned for } Principle_i = \sum(\text{points scored per question in } Principle_i)$$

where $Principle_i$ = the principle of quality in question.

The team calculated the score for each report using the following formula:

$$Principle_i \text{ score} = (\text{Total Points Earned for } Principle_i \div \text{Total Points Possible for } Principle_i) \times 100\%$$

where $Principle_i$ = the principle of quality in question.

Based on this calculation, the report was classified either as meeting "Minimum Adequacy" or as "Not Adequate." For a report to meet "Minimum Adequacy" in a given principle, the report must (1) receive a "Yes" or "Partial" response for the critically important question(s), if applicable, and (2) score 66 percent or above[4] for the relevant principle.

The team calculated the percentage of all reports reviewed that earned "Minimum Adequacy" in each principle and looked at the item-level data for the 2023 review by calculating the percentage of reports that received Yes, No, and Partial responses for each item. Where the data allowed, the team disaggregated the 2023 results by the primary USAID Education Policy Priority of the study, the year the study was published, the region of study, and the methodological approach used for the study.

The team also compared the strongest and weakest principles in the 2023 review to the strongest and weakest principles in the 2017 review and calculated percentage point changes for the percent of reports meeting "Minimum Adequacy" in each principle of quality between 2017 and 2023. However, all comparisons to the 2017 results are limited due to the different methodological and sampling approaches used, the difference between data collection instruments, and other limitations discussed in the "Limitations" section below.

LIMITATIONS

The reviewer composition likely biased the results of the study given that the volunteers were researchers, evaluators, and other experts who understand the relevant technical jargon. Consumers of reports and other end users were not recruited to participate as reviewers, thus limiting the ability to capture the extent to which non-experts understand reports and find them useful. This limits the ability to fully assess the cogency principle of quality, especially the items that ask whether the report is written so that the intended audience can understand it and whether visualizations support comprehension for non-technical audiences. Similarly, the selection criteria for the reports included in the review limit the possibility of assessing the extent to which all documents produced from a study are accessible to various audiences, thus limiting the team's assessment of a study's cogency to the technical report rather than the full suite of study documents.

The ASQ Tool used during the review was still in a draft form. For this reason, some items in the data collection instrument may not have been clear or fully explained. This could limit reviewers' ability to assess the extent to

[4] The cut point for "Minimum Adequacy" was determined based upon internal conversations between USAID and the DEEP team about the meaning of "Minimum Adequacy" followed by an internal assessment of the item-level response data. The cut point was originally set at 50 percent, to acknowledge that a report which scores at least half of the points possible is at least partially addressing most items in the ASQ Tool. Upon a closer review of the item-level data, however, this cut point did not make it clear that a low percentage of reports adhered to the items in the ASQ Tool.

which each item was addressed in the reports they reviewed, potentially skewing the data or leading to non-response for certain items.

The review was designed to collect a census of the quality of study reports published on the DEC between January 2018 and December 2022. However, the achieved sample size was not a census because not all volunteers completed their reviews. Thus, the final sample may not be truly representative of all reports published on the DEC despite the random assignment to reviewers. This likely biased the findings of this review, especially the disaggregated findings.

An additional limitation of the review is the determination of the cut point, which was not determined based on previous research. Therefore, the designation of "Minimum Adequacy" cannot be taken as a final determination of the overall quality of USAID-funded education studies. Instead, "Minimum Adequacy" can serve to identify the relative strengths and weaknesses of the USAID-funded education studies included in this review.

Finally, the comparison of the 2017 and 2023 reviews cannot be considered a true comparison. The reviews used different sample approaches, had different sample sizes, used different tools, and had different scoring procedures. The 2017 review and tool only included evaluations, while the 2023 study expanded both to be inclusive of research more broadly. The team provided details on percentage point changes in reports meeting "Minimum Adequacy" for each principle, but this information should be viewed in light of these limitations.

FINDINGS

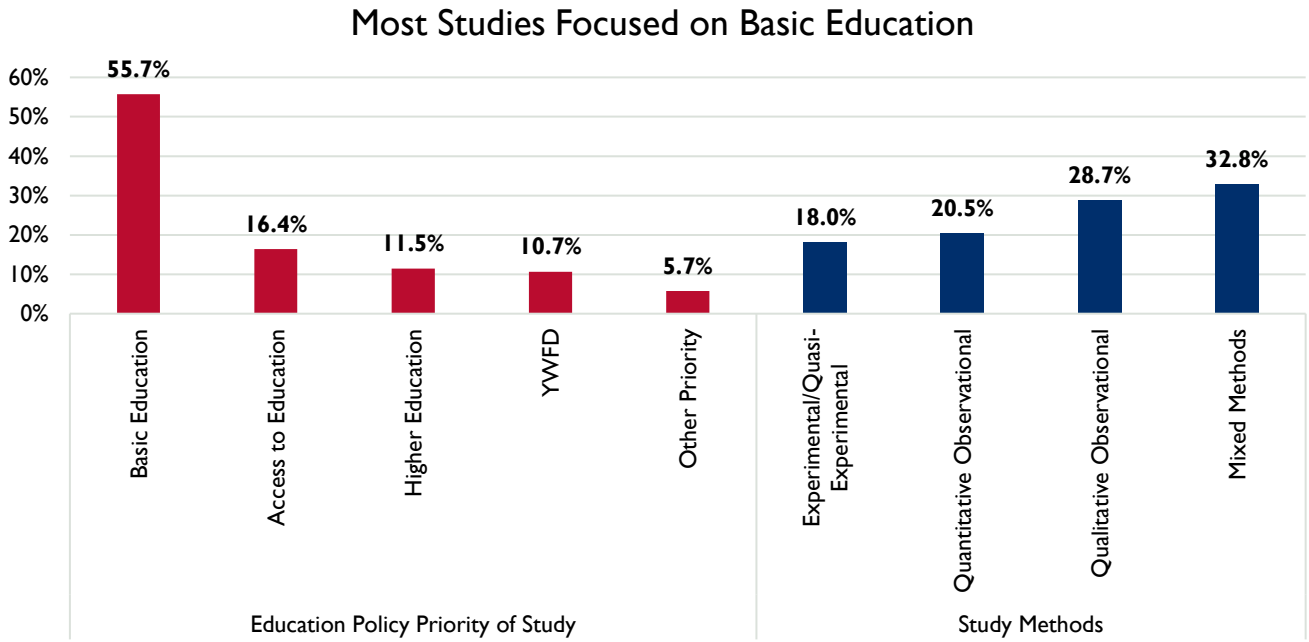
REVIEWER CHARACTERISTICS

A global group of researchers, evaluators, international partners, consultants, and academic partners volunteered to review one or more study reports following a broad solicitation by USAID. Ninety-two (56.4 percent) of the 163 volunteers completed their reviews. Volunteers who completed their reviews were based in 26 countries, representing a reviewer reach across six global regions. More reviewer details are in Annex D.

REPORT CHARACTERISTICS

While all 214 reports were assigned with the goal of conducting a census, the volunteers completed reviews of 122 (57 percent) reports. Most reports were about basic education, and mixed methods studies were more prevalent than other methods used for the reports reviewed. Exhibit 3 shows the distribution of reports by the primary USAID Education Policy[5] priority and methods used.

Exhibit 3: Distribution of reports by primary USAID Education Policy priority and method



Most reports were about studies conducted in the Africa region and in conflict and crisis settings. Exhibit 4 provides more details about the context of the reports reviewed.

[5] https://www.usaid.gov/sites/default/files/2022-05/2018_Education_Policy_FINAL_WEB.pdf

Exhibit 4: Context of reports reviewed

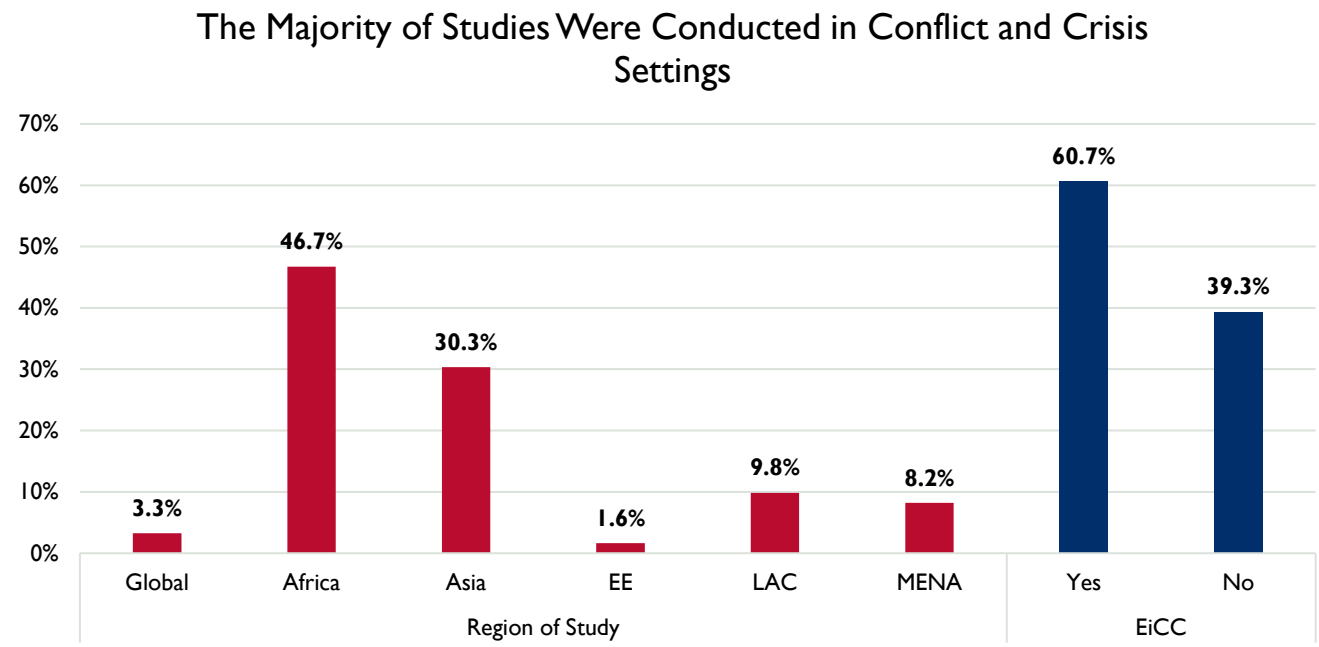
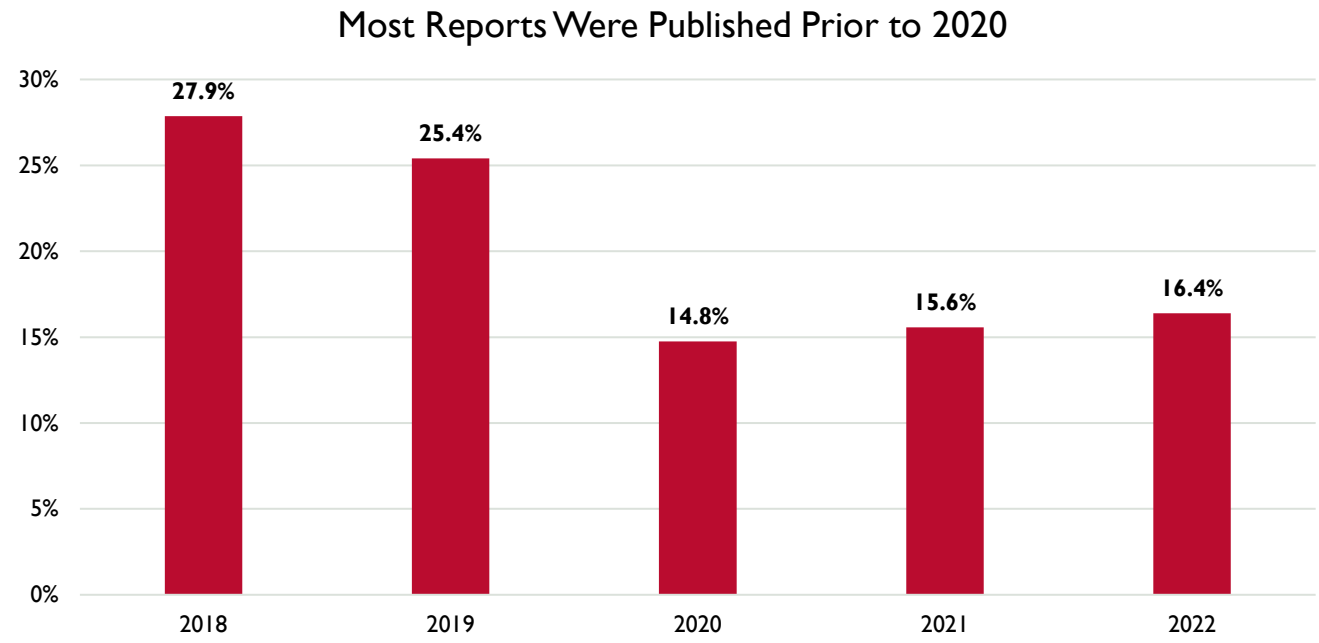


Exhibit 5 shows that most of the studies reviewed were published on the DEC in 2018 and 2019.

Exhibit 5: Distribution of reports reviewed by year of publication



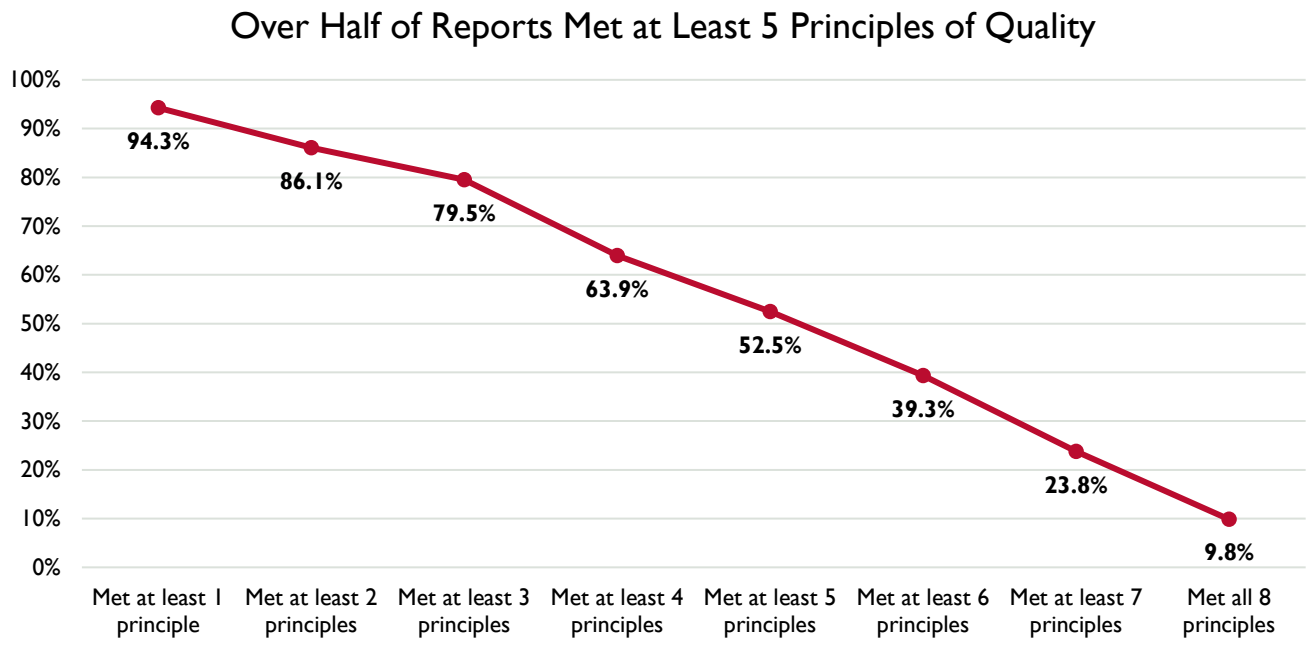
WHAT IS THE OVERALL QUALITY OF STUDY REPORTS PUBLISHED ON THE DEC SINCE 2017?

The overall quality of reports demonstrates that most met "Minimum Adequacy" for at least one of the eight principles of quality, and most met "Minimum Adequacy" for multiple principles of quality.

FINDING 1: More than half (52.5 percent) of reports met "Minimum Adequacy" for five or more of the eight principles of quality.

Exhibit 6 shows that 86.1 percent of the reports reviewed (n = 105) met "Minimum Adequacy" for at least two principles of quality and over half of the reports reviewed (52.5 percent, n = 64) met "Minimum Adequacy" for five or more of the eight principles.

Exhibit 6: Percent of reports meeting multiple principles of quality



WHAT ARE THE STRONGEST AND WEAKEST PRINCIPLES OF QUALITY AMONG STUDY REPORTS PUBLISHED SINCE 2017?

The results by principle of quality provide insight into the strengths and weaknesses of the reports reviewed.

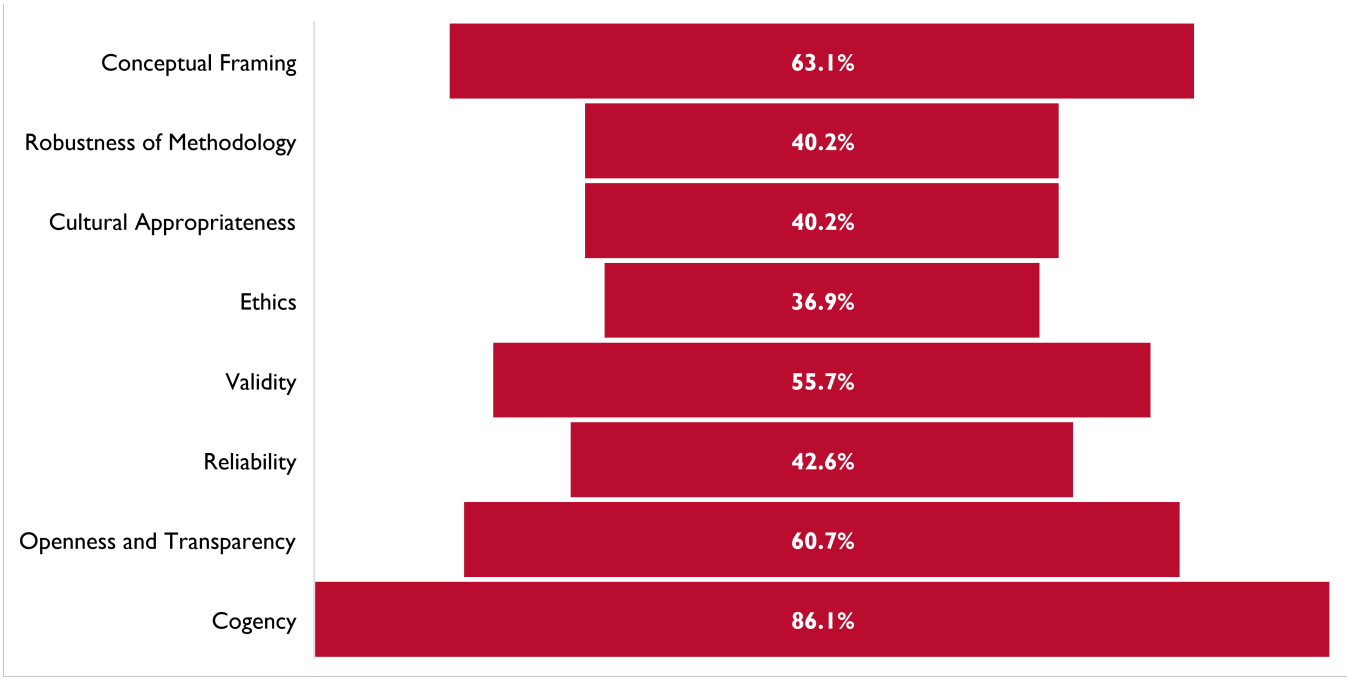
FINDING 2: The strongest principle of quality, openness and transparency, was met by 60.7 percent of reports.

As seen in Exhibit 7, the strongest principle of quality, met by 86.1 percent of reports, is cogency, and the third-strongest principle of quality, met by 60.7 percent of reports, is openness and transparency. Since these two principles are part of the report writing phase, this indicates strong overall scores for report writing.

FINDING 3: The weakest principle of quality, ethics, was met by 36.9 percent of reports.

The weakest principles are ethics (met by 36.9 percent of reports), cultural appropriateness (met by 40.2 percent of reports), and robustness of methodology (met by 40.2 percent of reports). These three principles are part of the study design phase, suggesting that study design may be a challenge for those planning USAID-funded education research and evaluation.

Exhibit 7: Percentage of reports meeting "Minimum Adequacy" by principle of quality



The two principles of quality that fall under study implementation, validity and reliability, are among the middle principles in terms of reports meeting "Minimum Adequacy." Validity is the fourth strongest principle, with 55.7 percent of reports meeting "Minimum Adequacy," followed by reliability, with 42.6 percent.

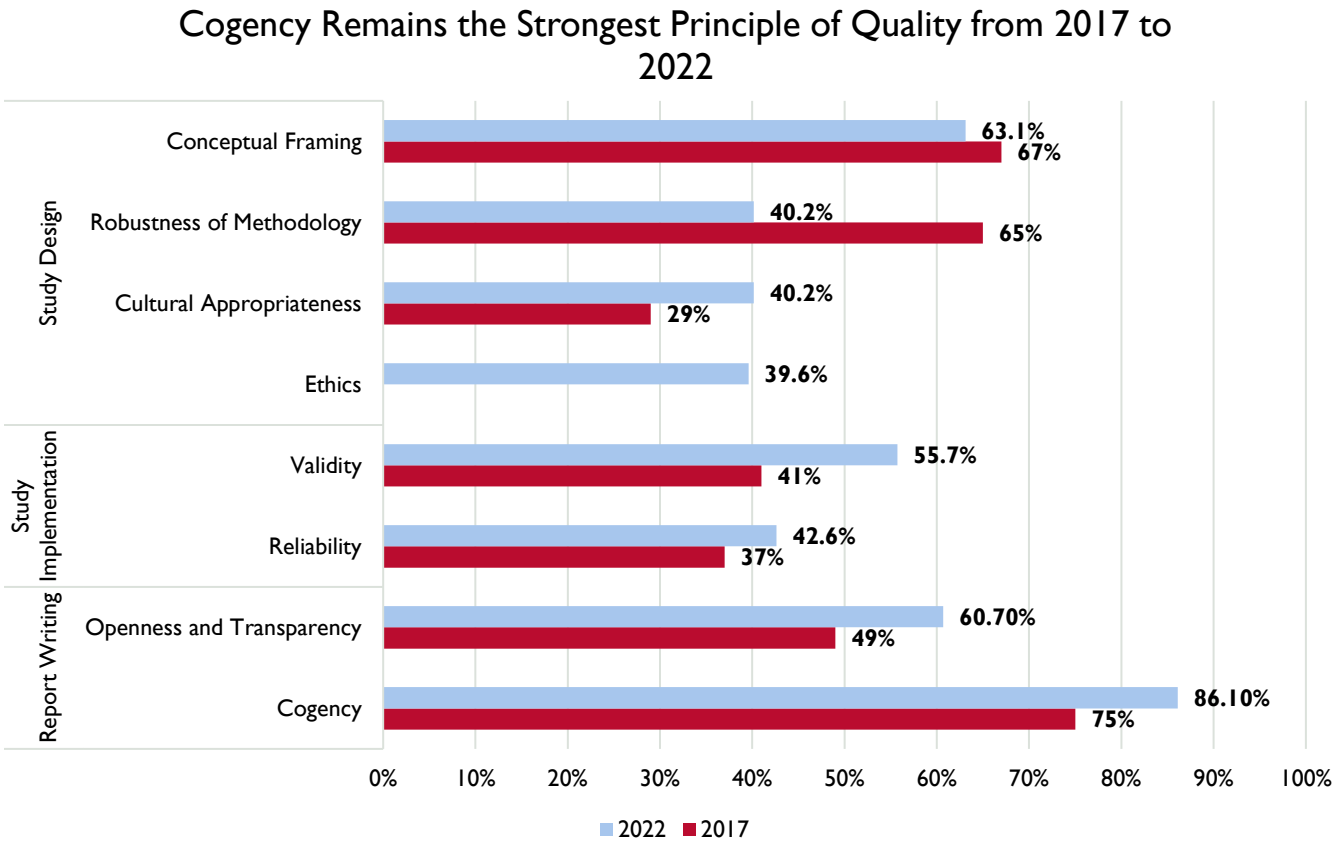
HOW HAS THE QUALITY OF STUDY REPORTS CHANGED SINCE THE PREVIOUS REVIEW?

Since 2017, when the last review was conducted, report quality has notably improved, although not consistently across principles of quality.

FINDING 4: The percentage of reports meeting "Minimum Adequacy" increased for all principles of quality except conceptual framing and robustness of methodology, which both decreased.

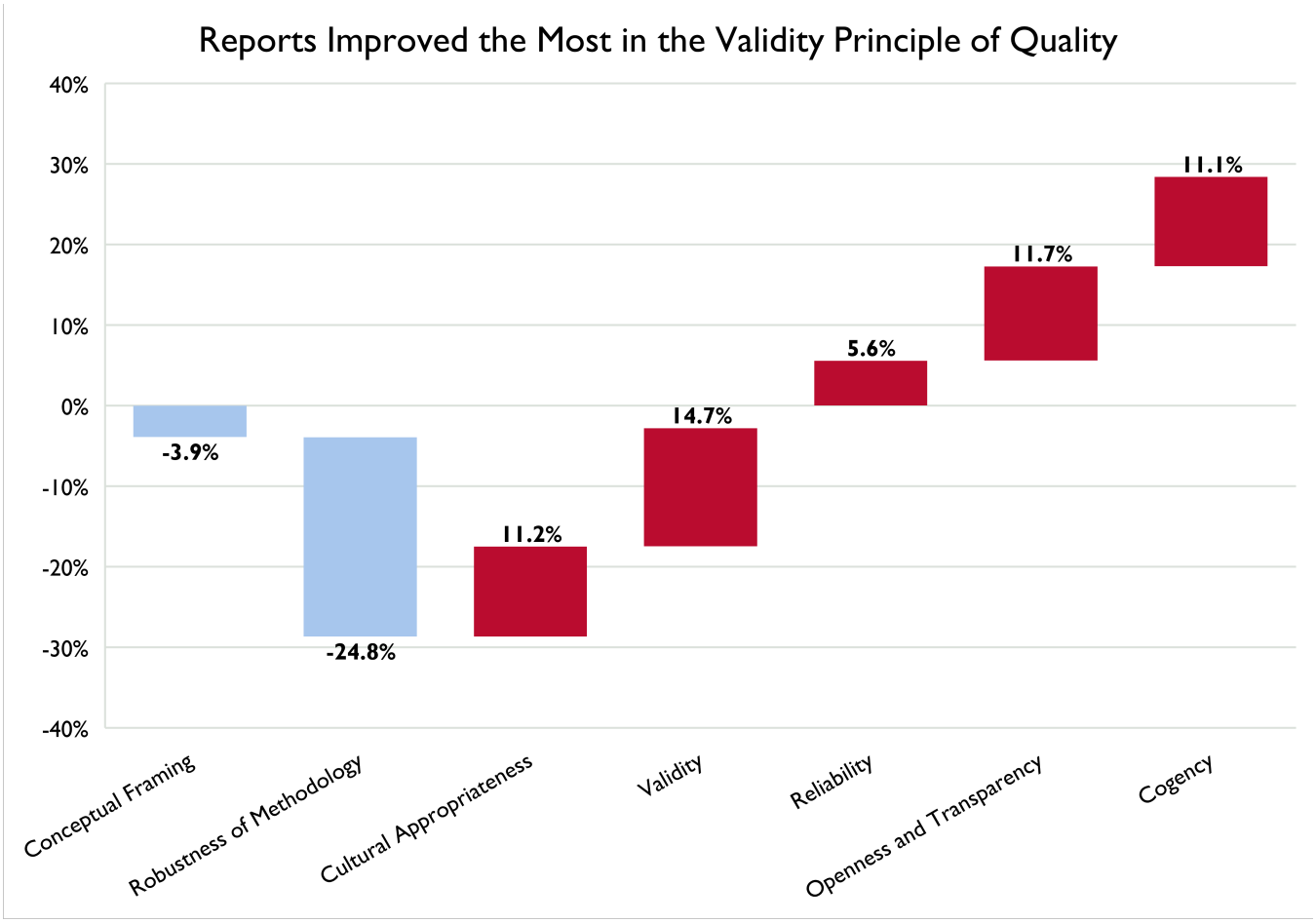
As Exhibit 8 shows, cogency remained the strongest principle of quality for both reviews. Cultural appropriateness remained among the weakest, although in 2017 it was the weakest principle and in 2023 it was tied for second weakest. Validity and reliability were among the bottom three in 2017 and are now solidly in the middle in 2022. Note that ethics was not a principle of quality included in the 2017 review.

Exhibit 8: Percent of reports meeting "Minimum Adequacy," by principle of quality and review year



Looking at changes based on percentage point change since 2017 (Exhibit 9), the biggest gains were in validity, which showed a 14.7 percentage point increase, followed by openness and transparency (11.7 percentage point increase) and cultural appropriateness (11.2 percentage point increase). From 2017 to 2022, two principles of quality experienced a decrease in the percentage of reports meeting "Minimum Adequacy." Robustness of methodology experienced the largest decrease, 24.8 percentage points. However, the two reviews used different scoring methods, tools, and sampling approaches, which limited the comparison. These limitations are described in the Limitations section above.

Exhibit 9: Percentage point change in reports meeting "Minimum Adequacy" since 2017



WHAT IS THE QUALITY OF STUDY REPORTS PUBLISHED SINCE 2017 BY PRINCIPLE OF QUALITY?

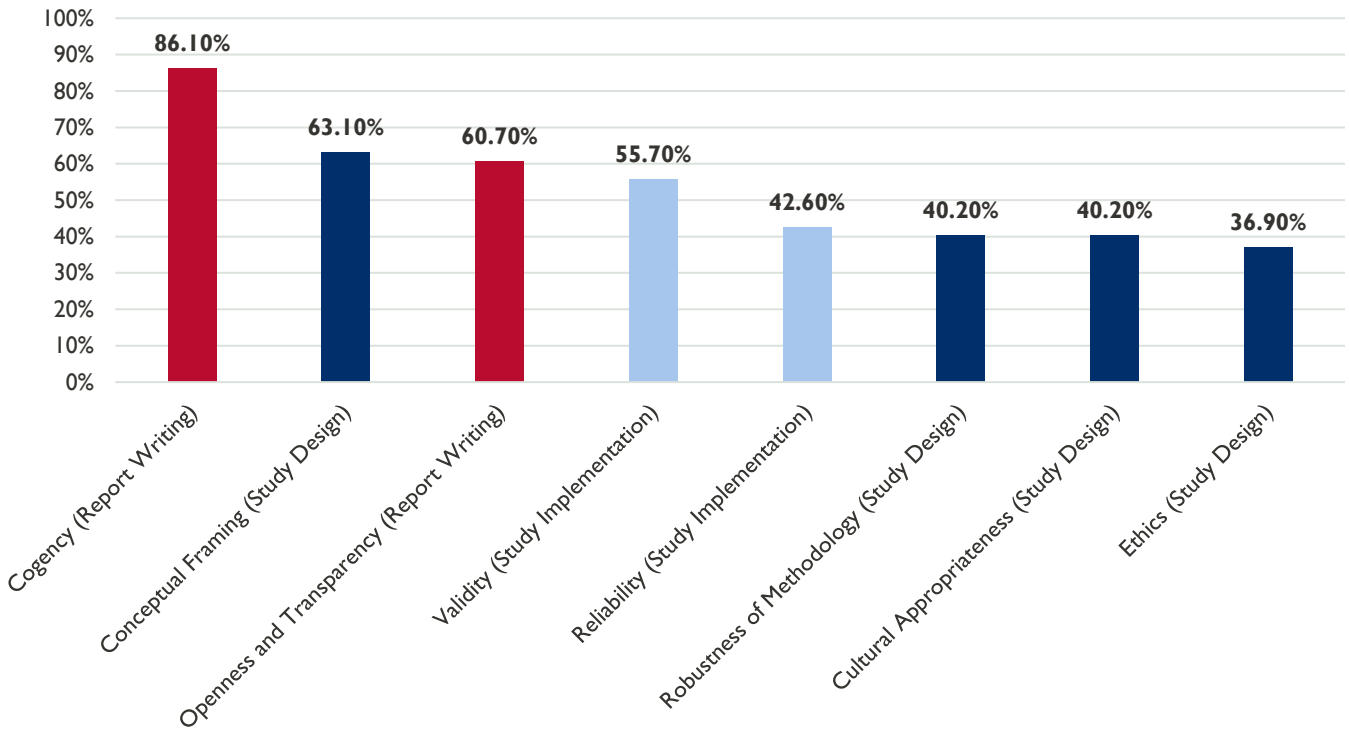
While the overall quality of reports provides insight into the strongest and weakest principles of quality as well as changes since the previous review, additional insight is gained when examining the item-level data for each principle of quality. This section is organized by study phase, beginning with the study design phase comprising the conceptual framing, robustness of methodology, cultural appropriateness, and ethics principles. The section continues with the study implementation phase comprising the validity and reliability principles, before concluding with the report writing phase that includes the openness and transparency and cogency principles of quality.

STUDY DESIGN PHASE

FINDING 5: The three weakest principles of quality were under the study design phase.

Although conceptual framing, one of the principles of quality under the study design phase, was the second-strongest principle, principles in this phase were the weakest in terms of the percentage of reports that met "Minimum Adequacy." As seen in Exhibit 10, the three weakest principles of quality were robustness of methodology, cultural appropriateness, and ethics.

Exhibit 10: Percent of reports that met "Minimum Adequacy" for each principle of quality, in descending order



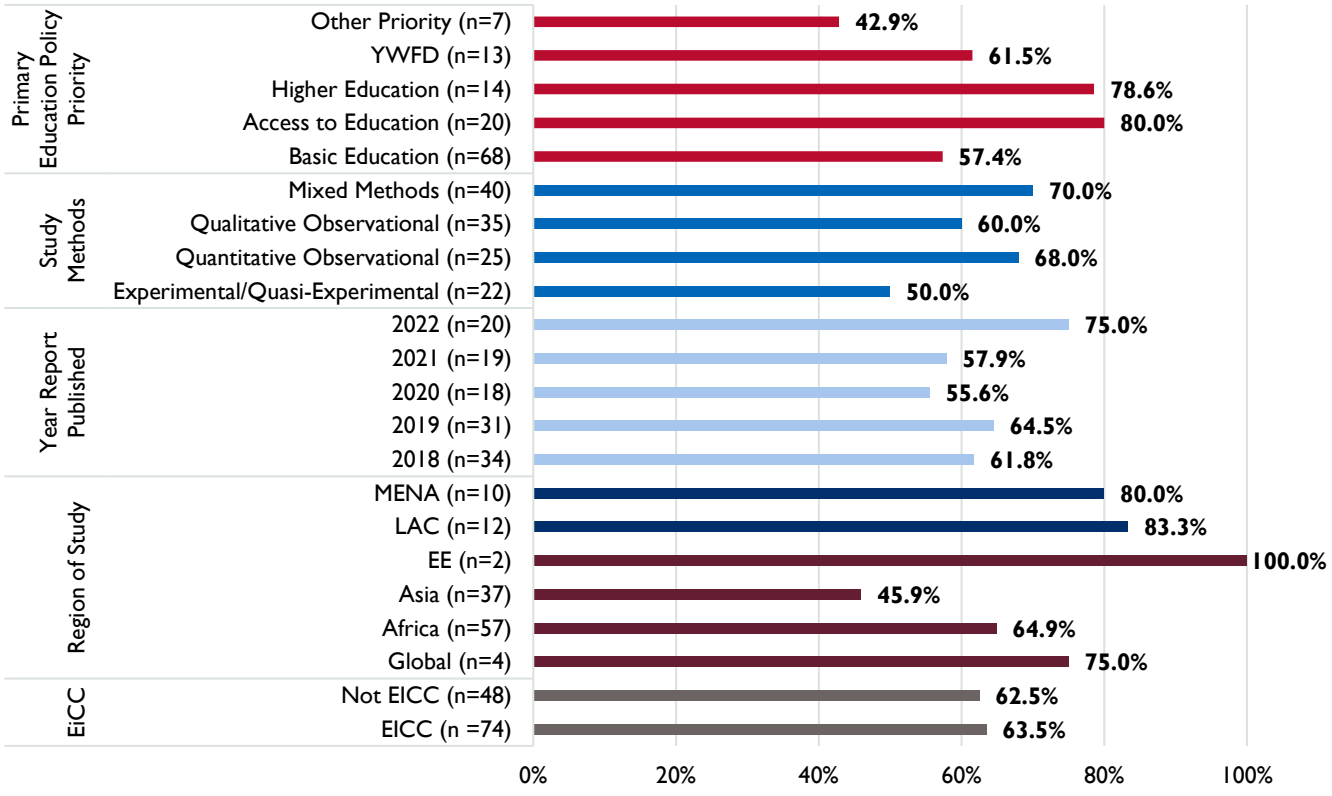
CONCEPTUAL FRAMING

FINDING 6: Conceptual framing is the second-strongest principle of quality despite the percentage of reports meeting "Minimum Adequacy" decreasing since the previous review.

Although the percentage of reports meeting "Minimum Adequacy" in conceptual framing decreased by 3.9 percentage points since the previous review, it was the second-strongest principle with 63.1 percent of reports meeting "Minimum Adequacy."

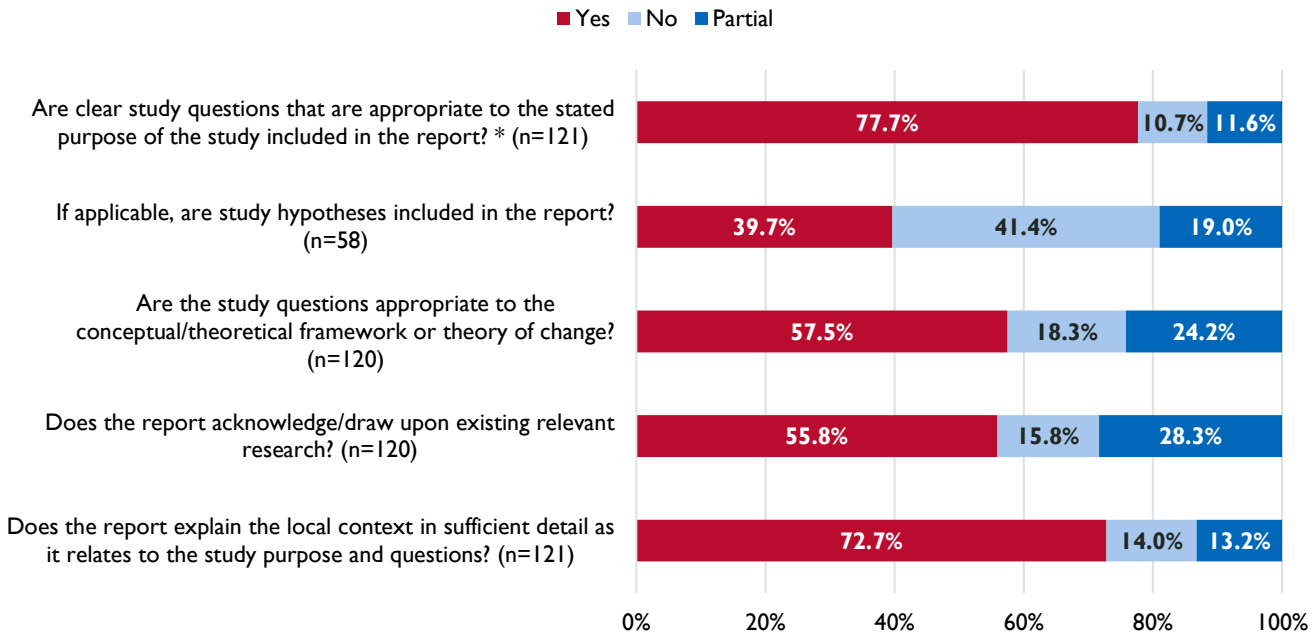
Exhibit 11 shows the percentage of reports that met minimum adequacy for conceptual framing by report characteristics.

Exhibit 11: Percentage of reports meeting "Minimum Adequacy" in conceptual framing by factor



As seen in Exhibit 12, the item with highest adherence is the critically important question, "Are clear study questions that are appropriate to the stated purpose of the study included in the report?". Explaining local context was also a strong point for conceptual framing. As one reviewer commented, "The report explains the local context in sufficient detail as it relates to the study purpose and questions, and this makes it easier for the reader to be conversant with the background to the assessment study." The Item with lowest adherence is related to hypotheses; reviewers indicated hypotheses would be appropriate for 58 studies, but only 39.7 percent of those reports included the hypotheses. Some comments from reviewers indicated there may be confusion about the difference between hypotheses and questions. For example, one reviewer commented that "the research questions can be considered as hypotheses in this study." However, most reviewer comments were clear that hypotheses were either unnecessary or should have been included and were not, leaving the reader to "infer them."

Exhibit 12: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the conceptual framing principle of quality



** Denotes the critically important question. Reports that received a "No" for that item were automatically classified as "Not Adequate" for the principle.*

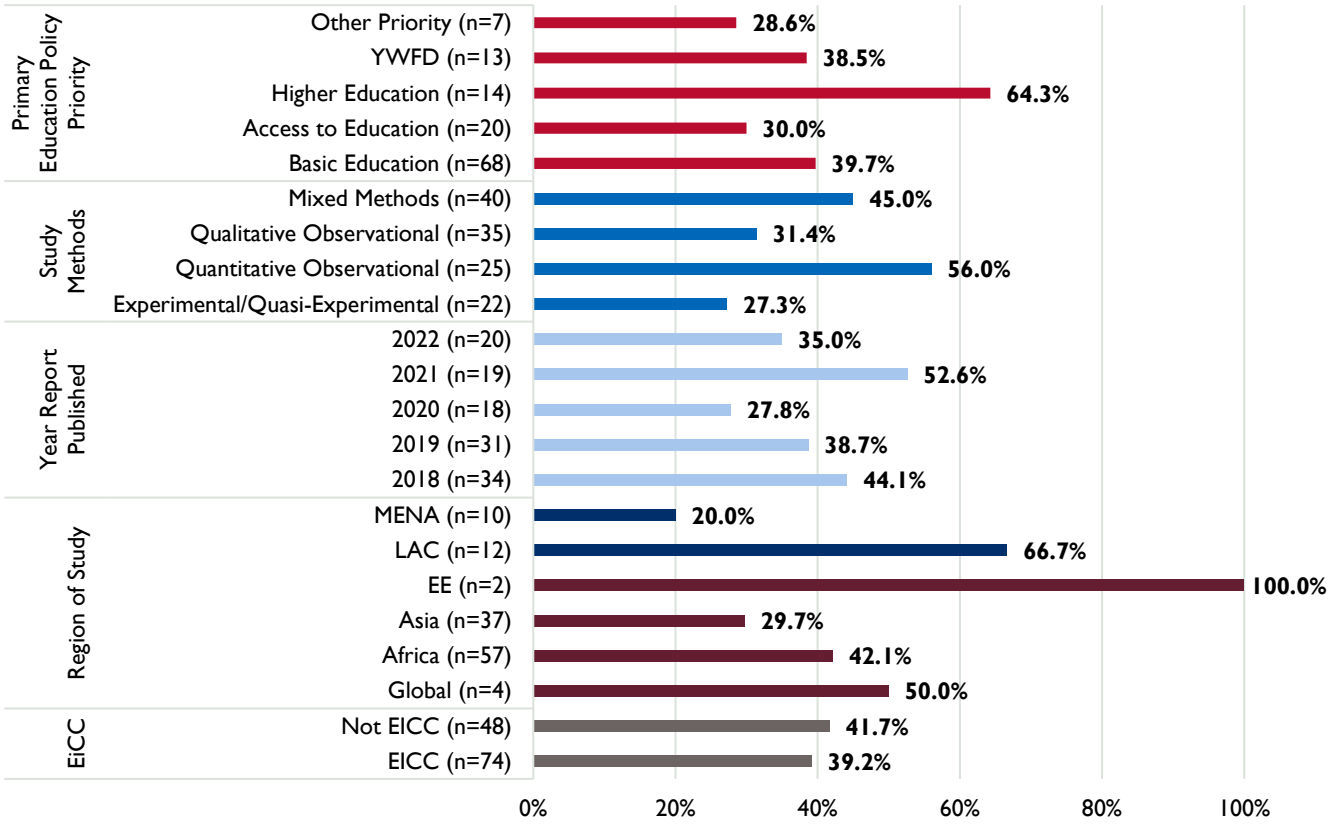
ROBUSTNESS OF METHODOLOGY

FINDING 7: Robustness of methodology is tied for the second-weakest principle of quality in terms of the percentage of reports meeting "Minimum Adequacy."

With 40.2 percent of reports meeting "Minimum Adequacy," robustness of methodology was tied with cultural appropriateness as the second-weakest principle of quality and experienced the greatest decrease—24.8 percentage points—in reports meeting "Minimum Adequacy" since the previous review.

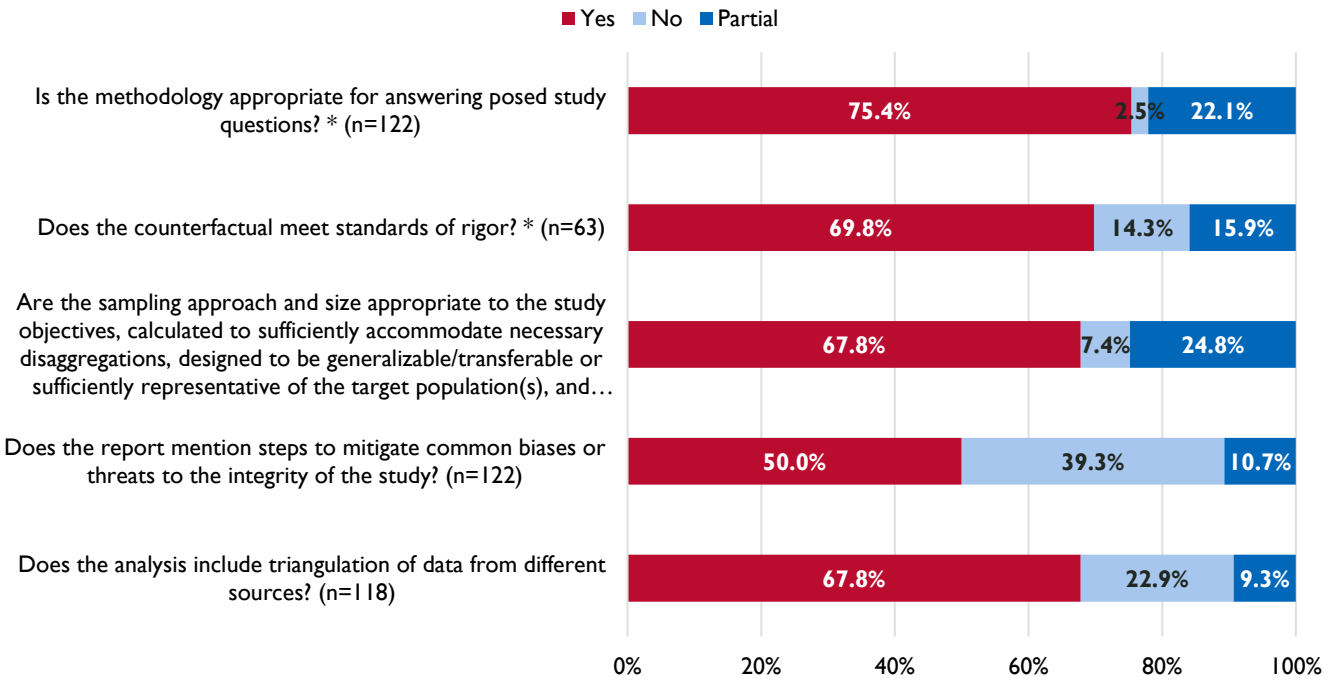
Exhibit 13 shows the percentage of reports that met minimum adequacy for robustness of methodology by report characteristics.

Exhibit 13: Percentage of reports meeting "Minimum Adequacy" in robustness of methodology by factor



The first critically important question, "Is the methodology appropriate for answering posed study questions?" had relatively high adherence, with 75.4 percent of reports meeting the criteria, so it appears that those designing education research and evaluation studies are selecting appropriate methodologies. Items related to the counterfactual, the sampling approaches and sampling sizes, and triangulation of data in the analysis were met by roughly two-thirds of the studies, while mitigation of biases or threats to the integrity of the study was only met by half of the studies. Exhibit 14 provides more details about the percentage of reports that met (fully and partially) and did not meet each of the criteria.

Exhibit 14: Percent of reports receiving "Yes," "No," and "Partial" scores for each item under the robustness of methodology principle of quality



** Denotes the critically important questions. Reports that received a "No" for one or both of those items were automatically classified as "Not Adequate" for the principle.*

Some reviewer comments indicated that robustness of methodology is lagging due to issues with other principles of quality such as cultural appropriateness and cogency. For example, one reviewer commented, "The methodology is mostly appropriate but is not robust or cogent enough. The limitations of the findings are acknowledged; however, it does not elaborate on ethics or cultural sensitivity in its design." Another reviewer said, "The stated methodology could be appropriate for the stated questions. However, there are gaps. The report states that the evaluation uses a quasi-experimental design and that comparison schools were selected based on the degree to which they matched the characteristics of the intervention schools. It's unclear how this matching was conducted." Thus, the low percentage of reports meeting "Minimum Adequacy" in robustness of methodology seems to overlap with challenges in other principles of quality, which is logical because designing a study necessarily requires attending to cultural appropriateness and methodological design, and if a report is not clearly written or is missing details, a reader cannot know whether the methodology is adequately robust.

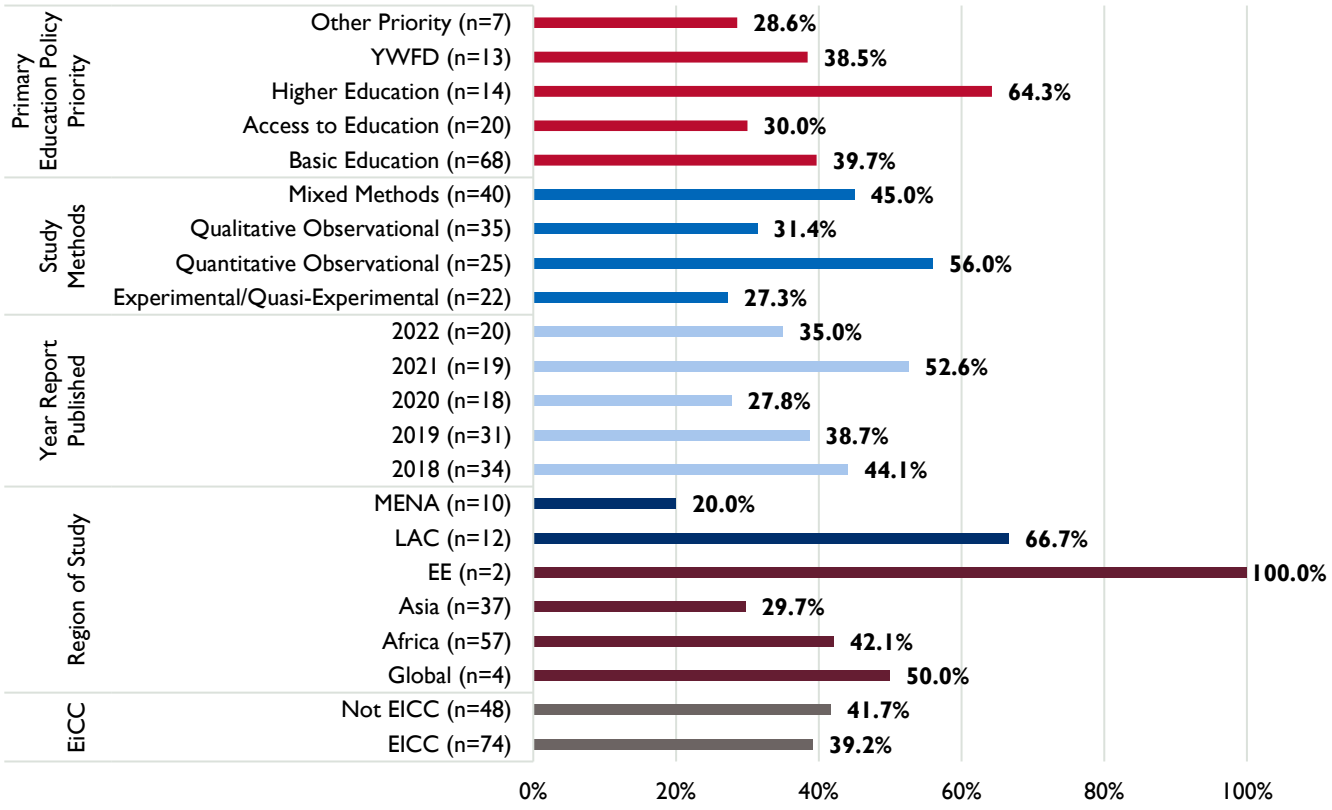
CULTURAL APPROPRIATENESS

FINDING 8: Cultural appropriateness is tied for the second-weakest principle of quality in terms of the percentage of reports meeting "Minimum Adequacy."

The item-level assessment reveals why cultural appropriateness is among the bottom three principles, with 40.2 percent of reports meeting "Minimum Adequacy."

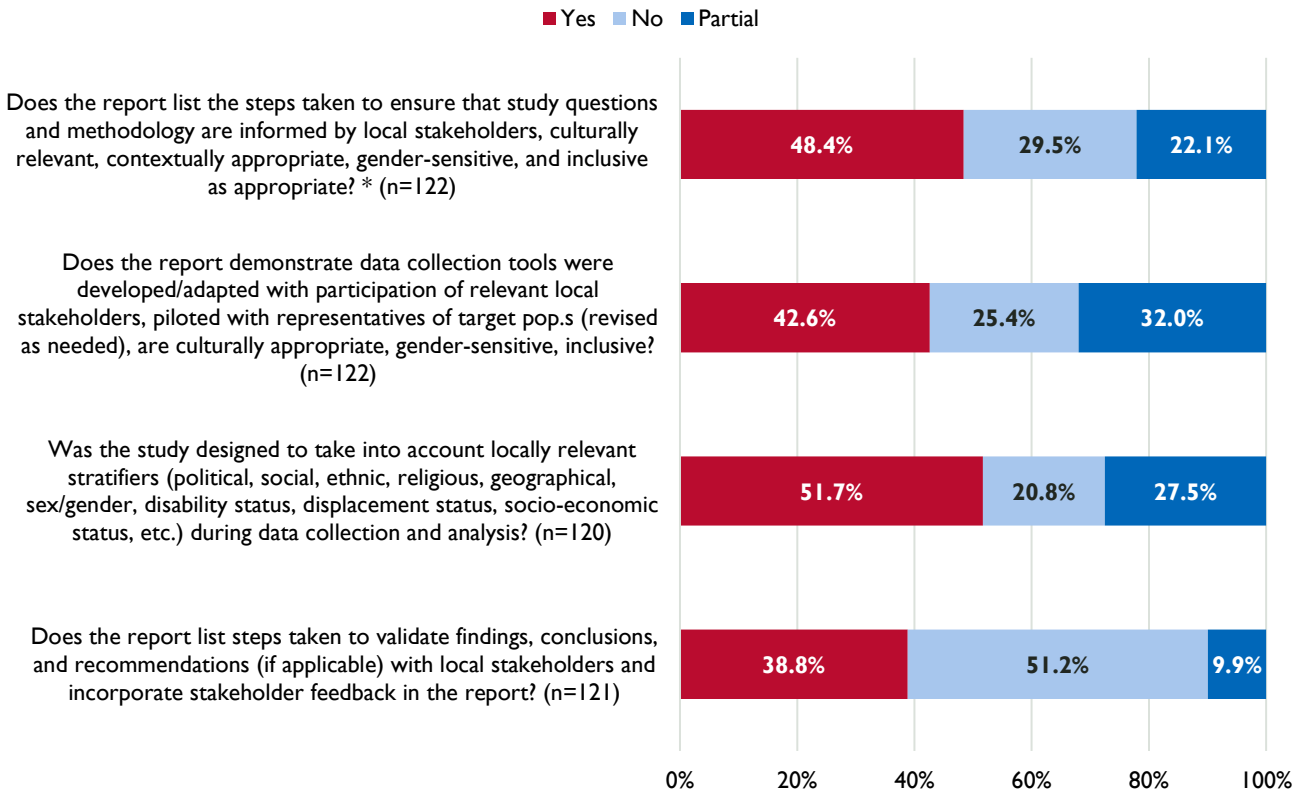
Exhibit 15 shows the percentage of reports that met "Minimum Adequacy" for cultural appropriateness by report characteristics.

Exhibit 15: Percentage of reports meeting "Minimum Adequacy" in cultural appropriateness by factor



While reports have improved since 2017 in this principle of quality, up from 29 percent, the low item-level scores show that studies still struggle with culturally appropriate design and implementation (see Exhibit 16). The item adhered to the most, "Was the study designed to take into account locally relevant stratifiers during data collection and analysis?" was only evidenced in 51.7 percent of the reports, with 27.5 percent partially meeting this criterion.

Exhibit 16: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the cultural appropriateness principle of quality[6]



** Denotes the critically important question. Reports that received a "No" this item were automatically classified as "Not Adequate" for the principle.*

Some of the reviewers' comments provide deeper insight into both the improvement and the consistent struggle with cultural appropriateness. One reviewer stated that "stakeholders were discussed in the report in a transactional and instrumental way." Another said, "Lots of work was done to ensure the tools were relevant to the local context—through desk review and then through review by their local research partner. There was little work to engage local stakeholders outside of the research partner and little work to ensure the gender sensitivity and inclusion of tools and methods." A third reviewer commented, "[The report] talks about piloting the tool for translation purposes. This explanation seems weak to me. I want to know more about the piloting—what was changed and why? The translation process and a cognitive pretesting should have been done in addition to the piloting." These comments and the percentage of reports with "Partial" scores for each item suggest that researchers and evaluators are attending to some aspects of cultural appropriateness but are often not taking all the elements of context and culture into consideration. The many reviewer comments on these questions indicate that studies are incorporating quite a few cultural and contextual elements but are not going far enough. For example, reports are not documenting how the results of a pilot were used to adjust data collection tools or how locally relevant stratifiers other than gender were considered in the study.

[6] Some questions were shortened in this chart due to space. The unabridged questions are included in Annex A.

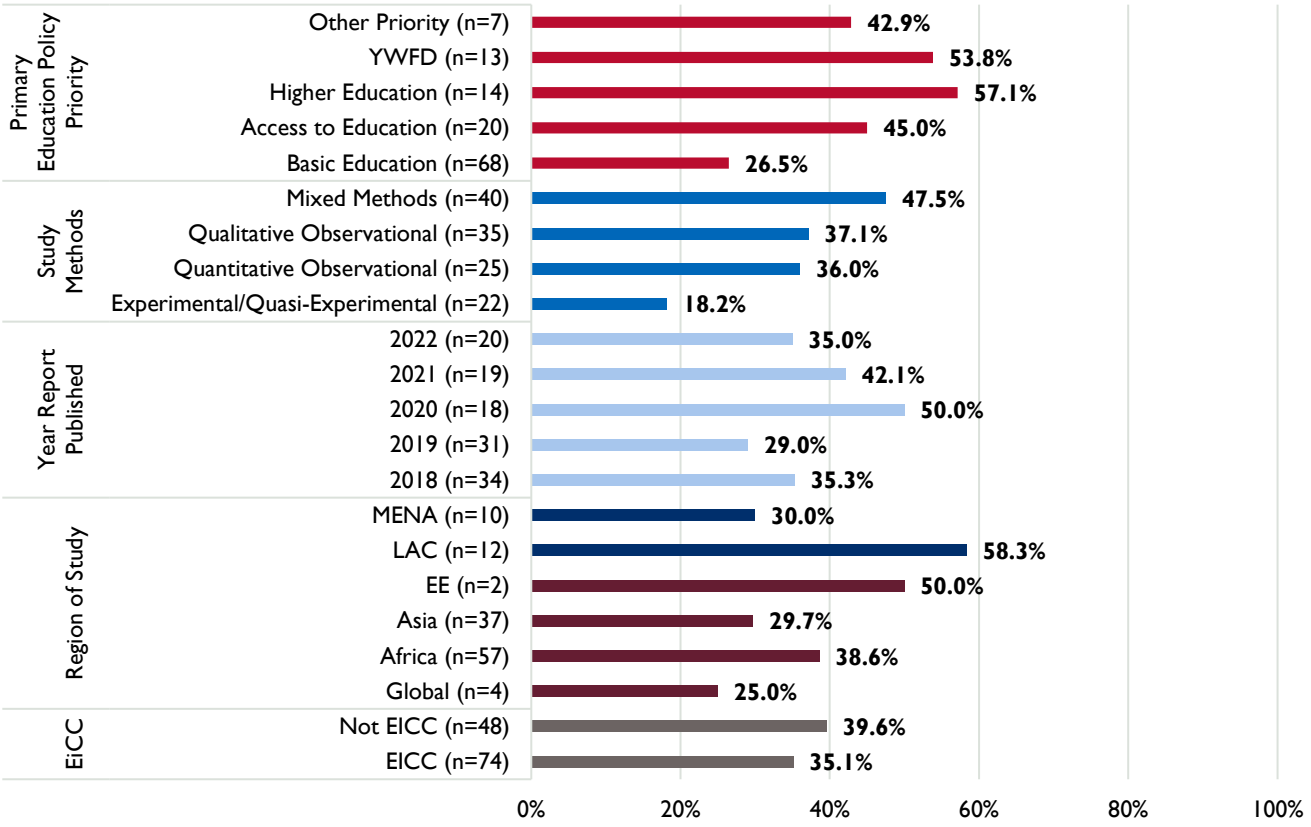
ETHICS

FINDING 9: Ethics is the weakest principle of quality, with the lowest percentage of reports meeting "Minimum Adequacy" relative to the other principles of quality.

Ethics was introduced as a principle of quality for this review because of its importance and the growing attention it is receiving in the field.

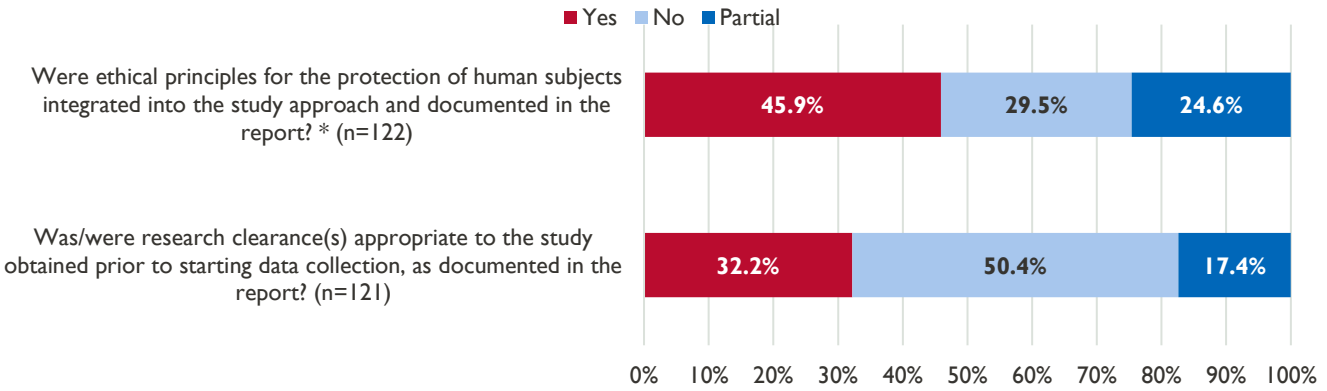
Exhibit 17 shows the percentage of reports that met minimum adequacy for ethics by report characteristics.

Exhibit 17: Percentage of reports meeting "Minimum Adequacy" in ethics by factor



Ethics was the weakest of the principles of quality, with 36.9 percent of reports meeting "Minimum Adequacy." While this may partly be due to the low number of items under this principle, the item-level data show that reports do not provide sufficient evidence that the protection of human subjects was accounted for or that research clearances were obtained (see Exhibit 18). This does not necessarily mean that the design and implementation of studies do not include ethical considerations but may point to a reporting issue. One reviewer indicated that they assumed the protection of human subjects was addressed, stating "ethical considerations are not explicitly expressed but since data were collected in a such study, this took place." Other reviewers noted that reports mentioned that ethical principles were applied, and consent was obtained, but did not explain how this was accomplished or provide details in the report narrative or annexes.

Exhibit 18: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the ethics principle of quality



** Denotes the critically important question. Reports that received a "No" this item were automatically classified as "Not Adequate" for the principle.*

STUDY IMPLEMENTATION PHASE

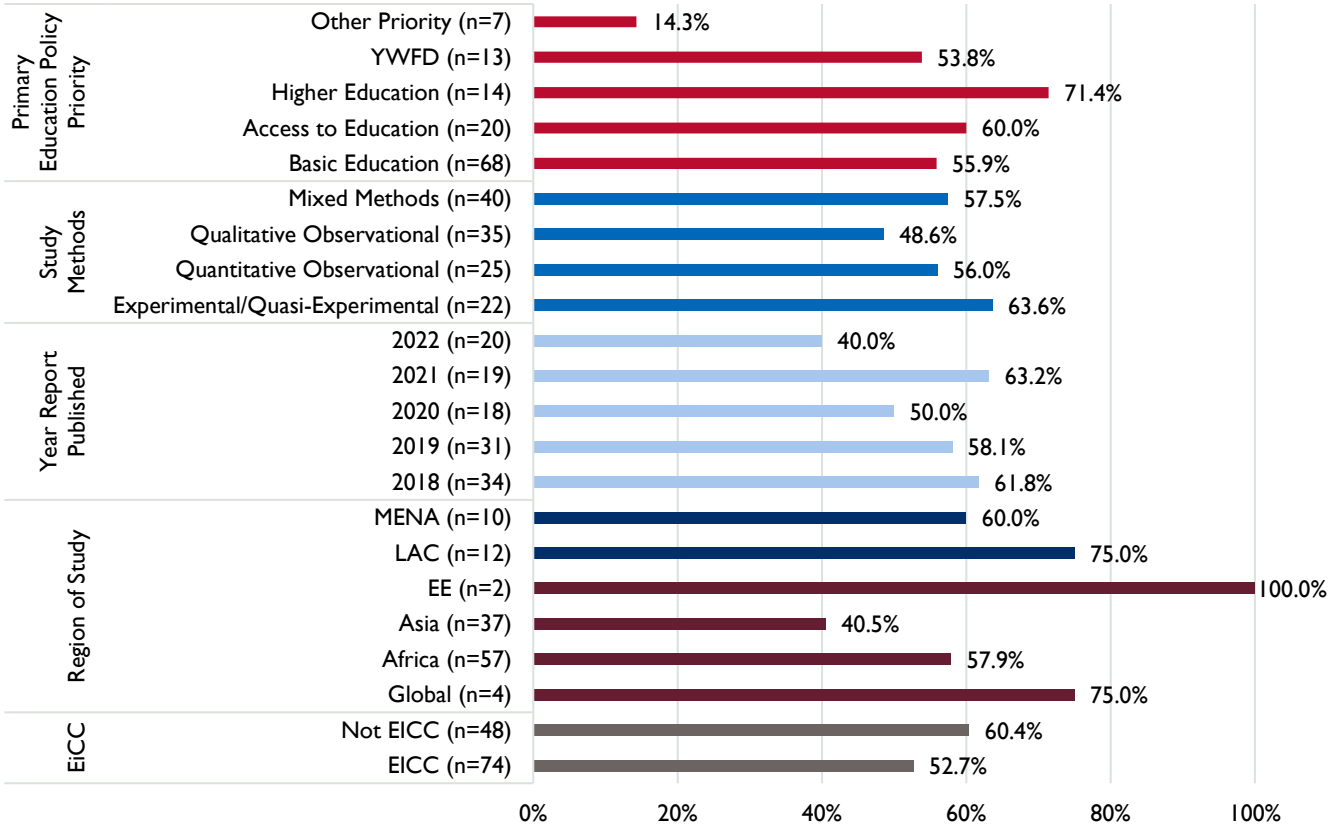
VALIDITY

FINDING 10: The validity principle of quality experienced the greatest improvement since the previous review to become the fourth-strongest principle in terms of percentage of reports meeting "Minimum Adequacy."

The validity principle of quality is the fourth-strongest principle, with 55.7 percent of reports meeting "Minimum Adequacy." Validity experienced the greatest gains since the previous review, with a 14.7 percentage point increase in reports meeting "Minimum Adequacy."

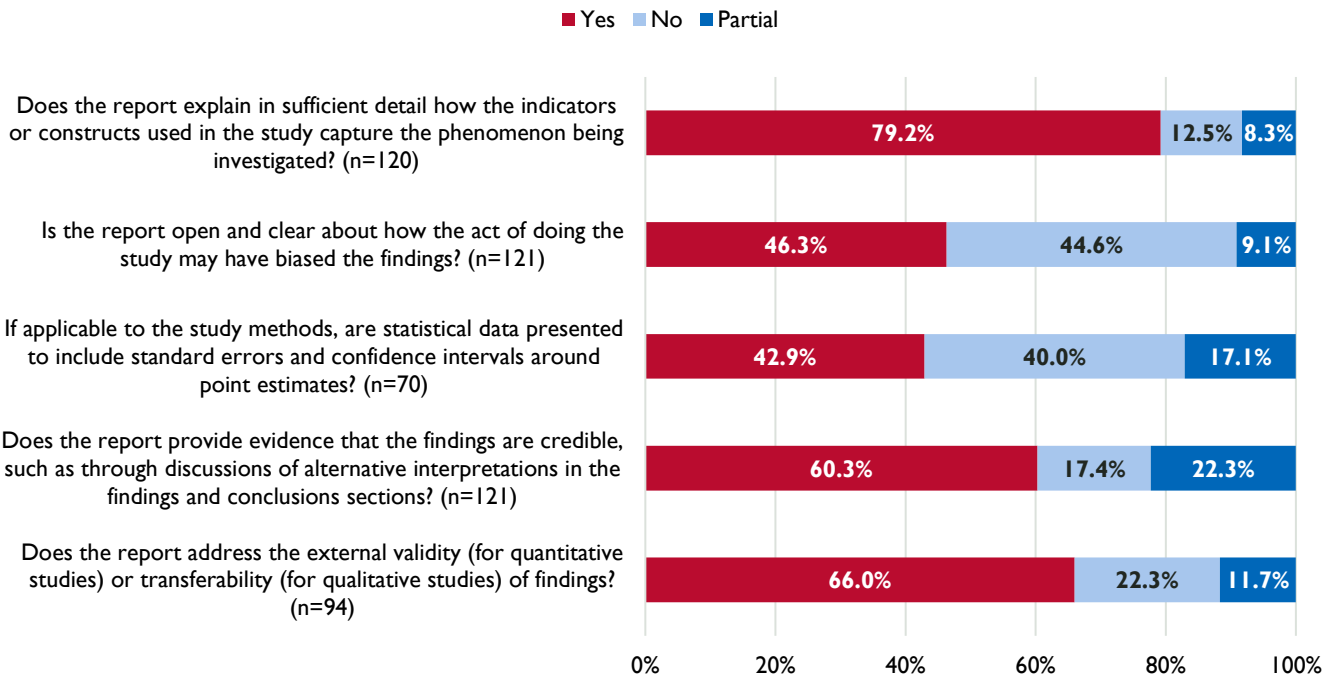
Exhibit 19 shows the percentage of reports that met "Minimum Adequacy" for validity by report characteristics.

Exhibit 19: Percent of reports meeting "Minimum Adequacy" in validity by factor



As Exhibit 20 shows, most of the items under validity were addressed by a majority of the reports reviewed, supporting the high percentage of reports that met "Minimum Adequacy" and demonstrating that attention is paid to validity and transferability. Among the items with lowest adherence are reports being open and clear about how conducting the study may bias the findings.

Exhibit 20: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the validity principle of quality



There was no critically important question for this principle.

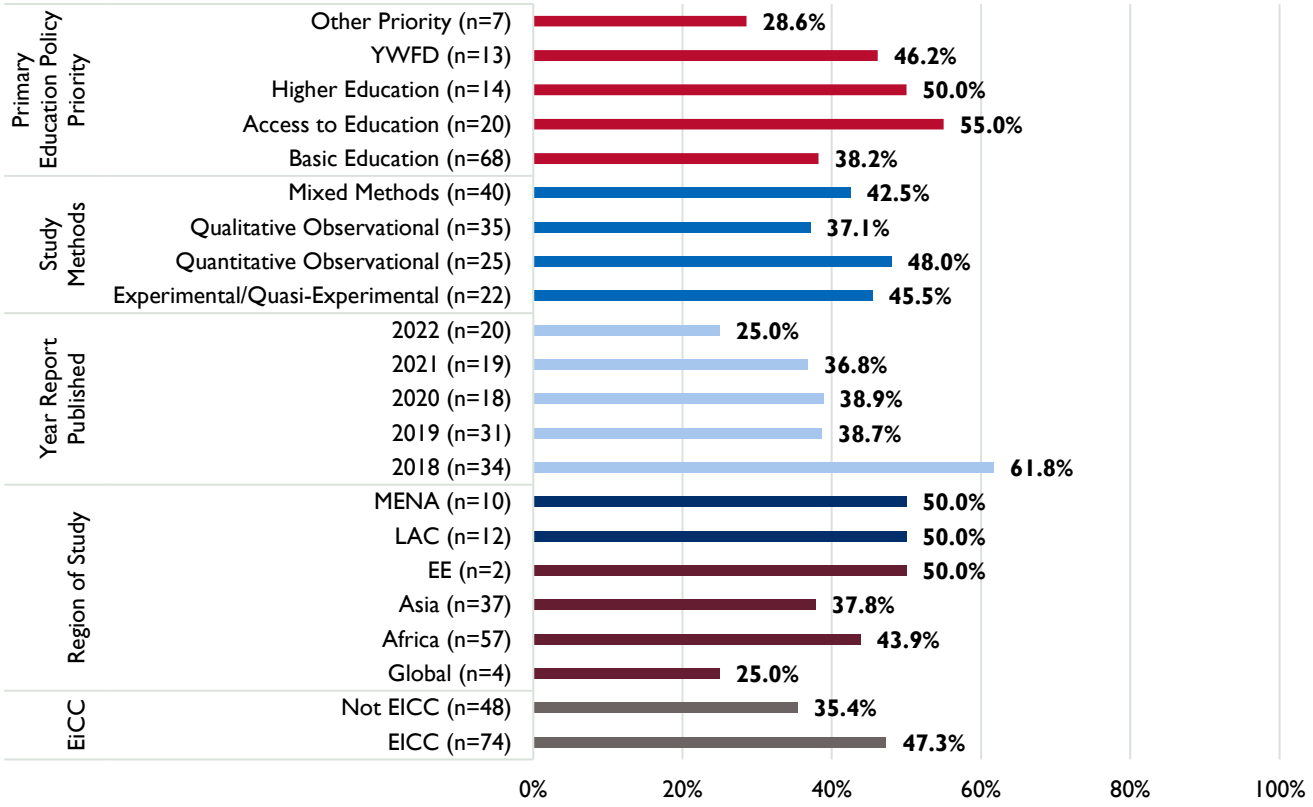
RELIABILITY

FINDING 11: Reliability is the fifth-strongest principle of quality.

Reliability also improved since the previous review, by 5.6 percentage points, and is the fifth-strongest principle of quality, with 42.6 percent of reports meeting "Minimum Adequacy."

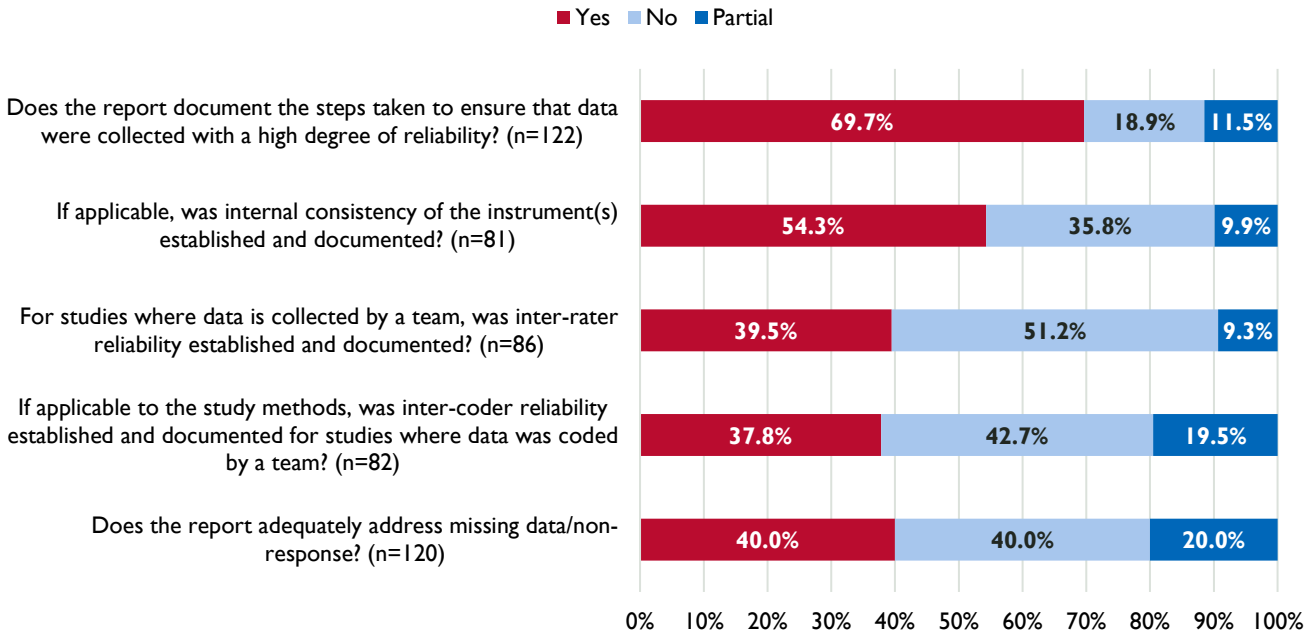
Exhibit 21 shows the percentage of reports that met "Minimum Adequacy" for reliability by report characteristics.

Exhibit 21: Percentage of reports meeting "Minimum Adequacy" in reliability by factor



The item-level data (see Exhibit 22), indicate that those implementing studies are paying attention to reliable data collection and including documentation of these steps in reports. However, based on what is documented in reports, implementers are not adequately attending to inter-rater reliability (IRR) and inter-coder reliability (ICR). This is not applicable to all studies; for example, not all qualitative studies use coding as a strategy, and not all rely on a team to code the data. Many of the reviewers' comments were along the lines of "it was unclear if this was necessary." This may indicate a few issues. First, it could signal limited reviewer understanding of IRR and ICR. It could also indicate that study teams are not implementing IRR or ICR or are uncertain about when these steps are necessary. There may also be gaps in report writing, with writers either not reporting on IRR or ICR, or being uncertain about how to indicate the necessity of these steps in the report. Finally, IRR or ICR may not have been necessary, but report writers did not note that in the methods section.

Exhibit 22: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the reliability principle of quality



There was no critically important question for this principle.

REPORT WRITING PHASE

FINDING 12: The report writing phase is the strongest of the three study phases in terms of the percentage of reports meeting "Minimum Adequacy" for the relevant principles of quality.

The two principles of quality under the report writing phase were ranked in the top three strongest principles of quality in terms of the percentage of reports meeting "Minimum Adequacy."

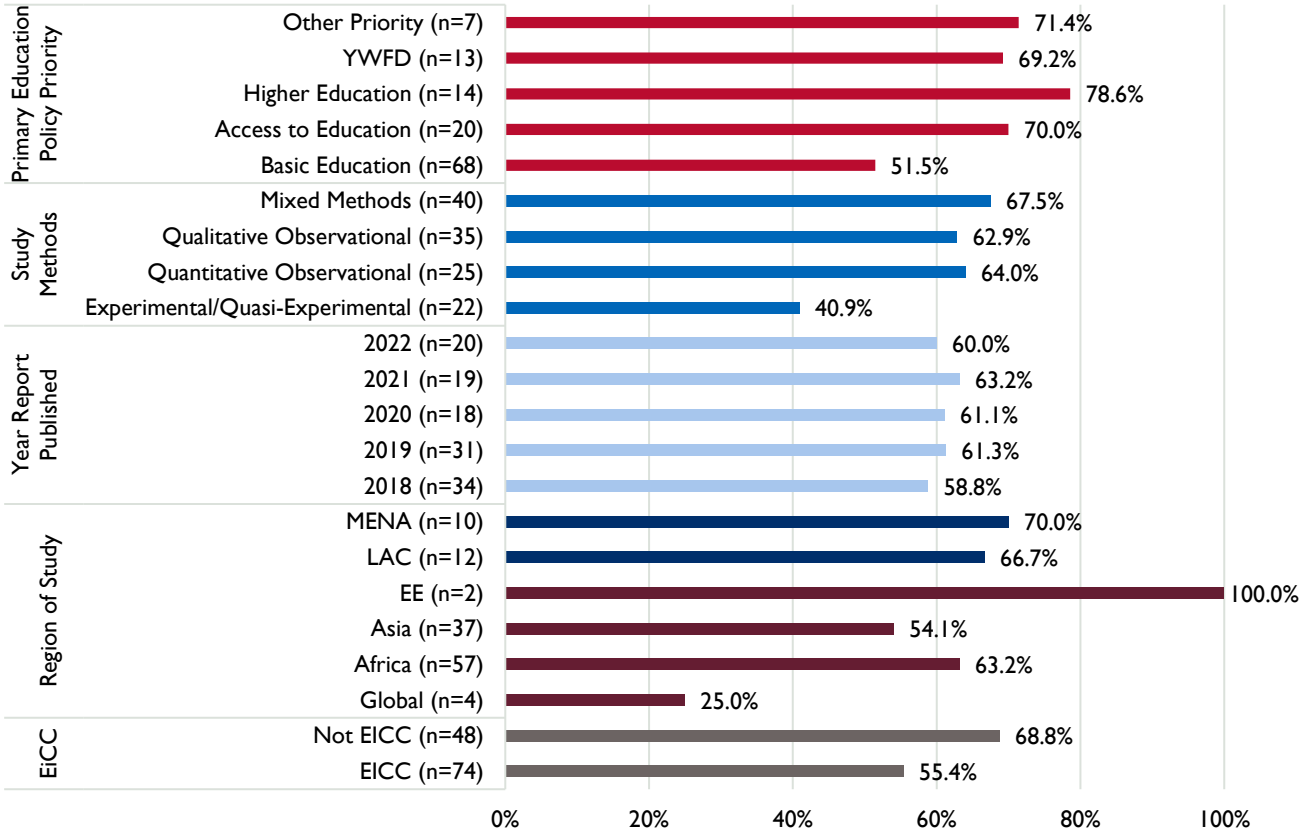
OPENNESS AND TRANSPARENCY

FINDING 13: Openness and transparency is the third-strongest principle of quality in terms of the percentage of reports meeting "Minimum Adequacy."

Openness and transparency is the third-strongest principle of quality, with 60.7 percent of reports meeting "Minimum Adequacy."

Exhibit 23 shows the percentage of reports that met minimum adequacy for openness and transparency by report characteristics.

Exhibit 23: Percentage of reports meeting "Minimum Adequacy" in openness and transparency by factor

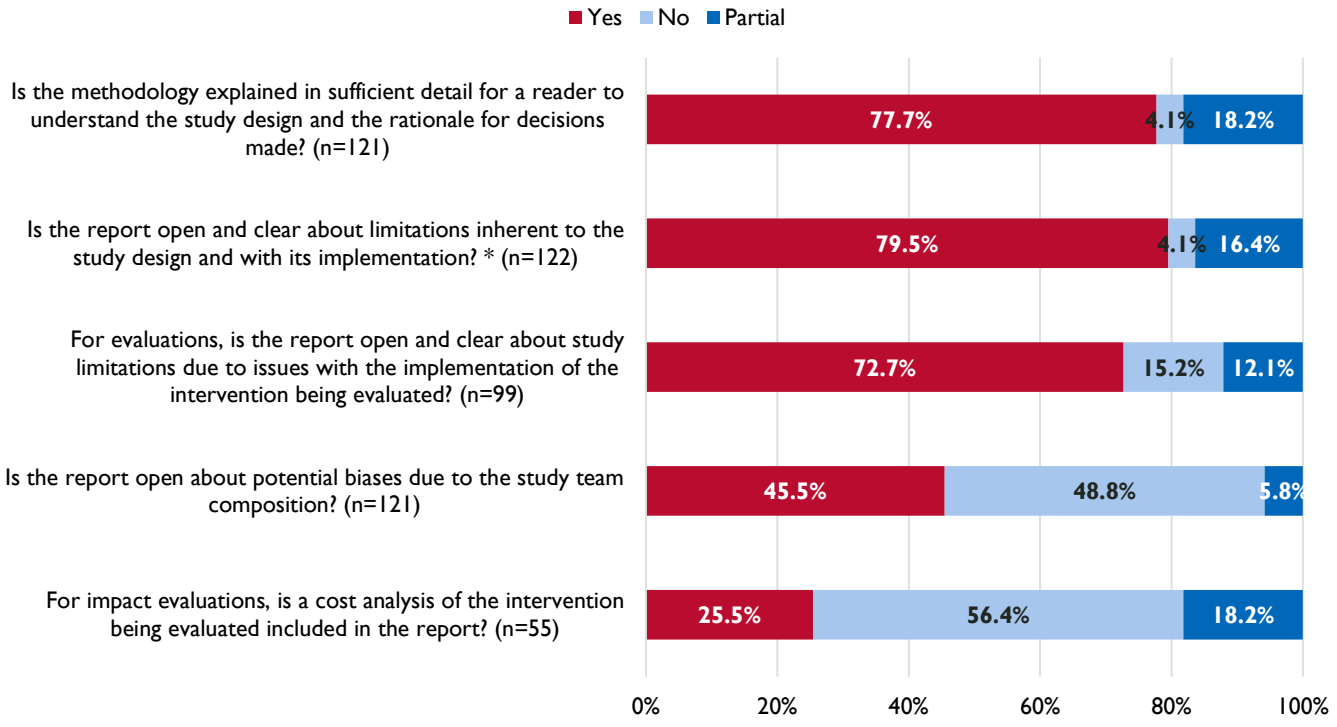


As Exhibit 24 shows, the item with the highest adherence was the critically important question, "Is the report open and clear about limitations inherent to the study design and with its implementation?" There was also high

adherence to explaining the methodology sufficiently for readers to understand the design and decisions made. This is interesting in light of the results related to IRR/ICR (see the "Reliability" section above); based upon these adherence results, reviewers should have been able to determine whether IRR and/or ICR were necessary for the study they reviewed. However, the reviewers' comments highlighted their uncertainty about the necessity of IRR/ICR. The discrepancy supports the possibility that reviewers may have limited familiarity with the two concepts (IRR and ICR).

Low adherence to the item "Is the report open about potential biases due to the study team composition?" may be due to uncertainty about what this means, as some reviewer comments indicated. One reviewer said, "The report acknowledges biases of participants (teachers towards students) but does not acknowledge the author's and/or researcher's [biases]" and, tellingly, "I've never seen a report do this." This suggests that the practice may not be well known among those conducting USAID-funded educational research and evaluations. However, awareness of the practice may be spreading, because openness and transparency experienced the second-greatest improvement since 2017, with an increase of 11.7 percentage points.

Exhibit 24: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the openness and transparency principle of quality



** Denotes the critically important question. Reports that received a "No" this item were automatically classified as "Not Adequate" for the principle.*

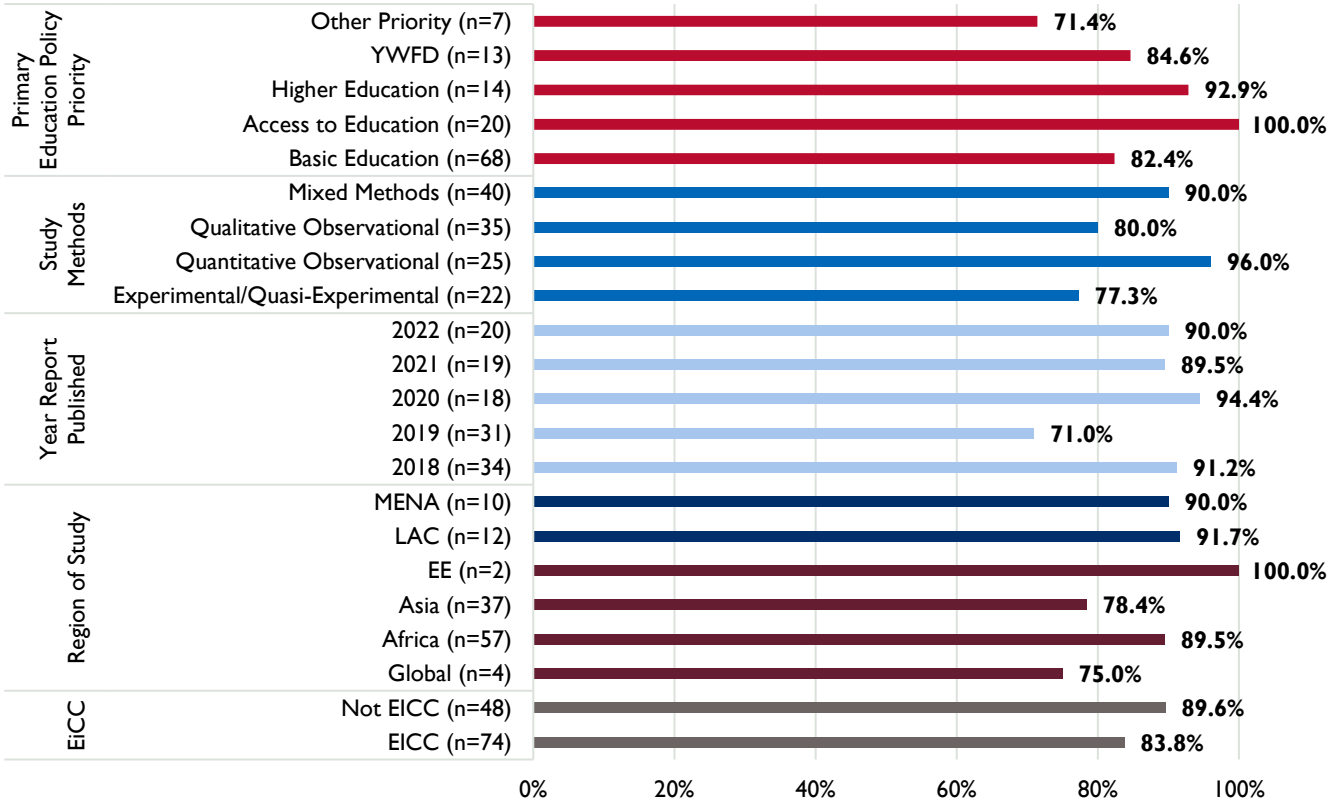
COGENCY

FINDING 14: Cogency is the strongest principle of quality in terms of the percentage of reports meeting "Minimum Adequacy."

The principle of quality with the highest percentage of reports meeting "Minimum Adequacy," 86.1 percent, was cogency.

The percentage of reports that met "Minimum Adequacy" for cogency by report characteristics is provided in Exhibit 25.

Exhibit 25: Percentage of reports meeting "Minimum Adequacy" in cogency by factor

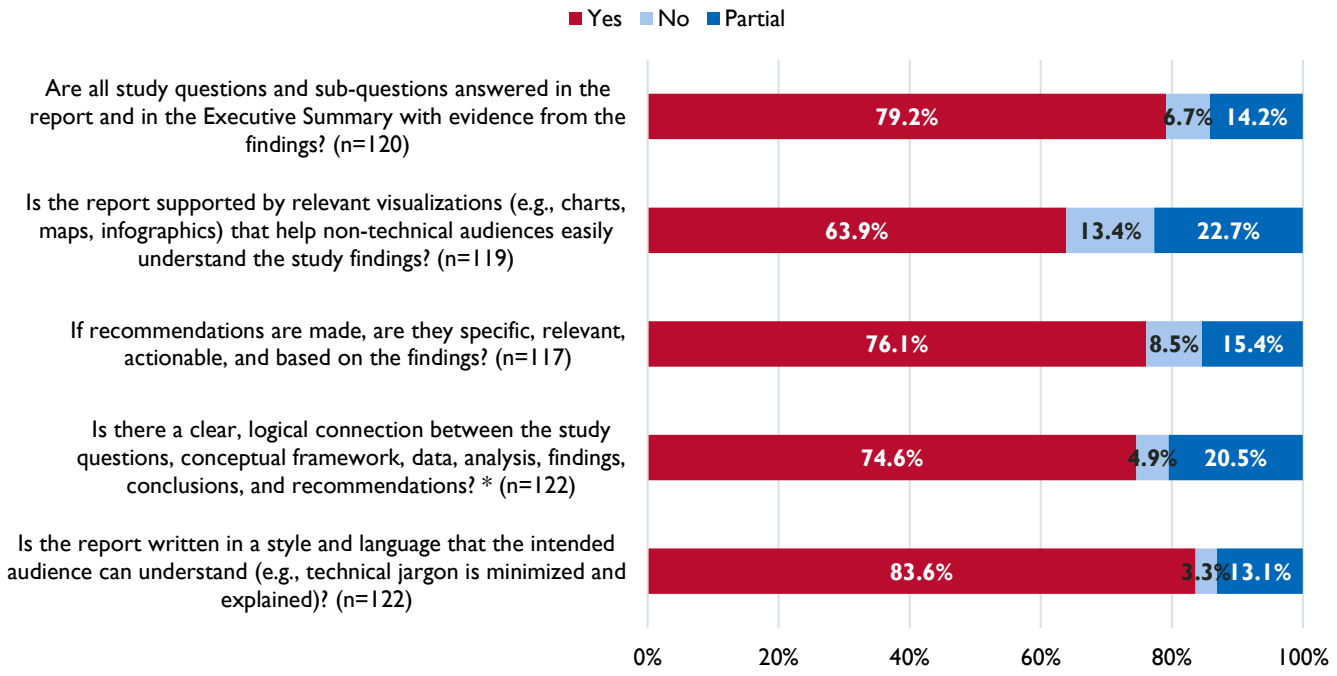


At the item level, the highest percentage of reports met the criterion for being written in a style and language that the intended audience can understand (see Exhibit 26). However, these results do not match anecdotal evidence. Decision makers and other consumers of reports often express frustration that reports are not easy to understand or usable. The mismatch may be due to the expertise and experience of the reviewers, who were researchers, evaluators, technical experts, and academics. One reviewer suggested this possibility, writing: "If the intended audience are technical experts at USAID, then yes, it read logically and I (as a technical expert with higher education level) was able to understand just fine. However, if these evaluations are meant to involve stakeholders, then these reports are extremely long, cumbersome, very technical, and difficult to pull out the key nuggets that they would need. I am imagining a primary school teacher or principal trying to determine from this report what the best intervention package/approach would be based on the data or what professional development they should pursue for the best cost effectiveness. It would be overwhelming without a guide."

This mismatch also highlights an important limitation of this study: the intended audience. This study is limited by the selection of documents reviewed using the ASQ Tool. The reports included in the study are technical documents whose intended audience includes technical experts, other researchers/evaluators, and donors with technical and research/evaluation backgrounds. Other dissemination products that distill the information, whether for policymakers or beneficiaries, were not included in this study. These types of documents should be written in a different style to reach these audiences. Therefore, the items in the ASQ Tool related to intended audience and limiting jargon must be considered based upon the intended audience for the report, in this case, researchers, evaluators, and technical experts.

The discrepancy between anecdotal evidence and these findings may indicate challenges outside the scope of this study, namely that dissemination products from USAID-funded education studies that are shared with decision makers use the same writing style as technical reports or that technical reports are being used as dissemination products in lieu of teams developing additional documents to share with various audiences.

Exhibit 26: Percentage of reports receiving "Yes," "No," and "Partial" scores for each item under the cogency principle of quality



** Denotes the critically important question. Reports that received a "No" this item were automatically classified as "Not Adequate" for the principle.*

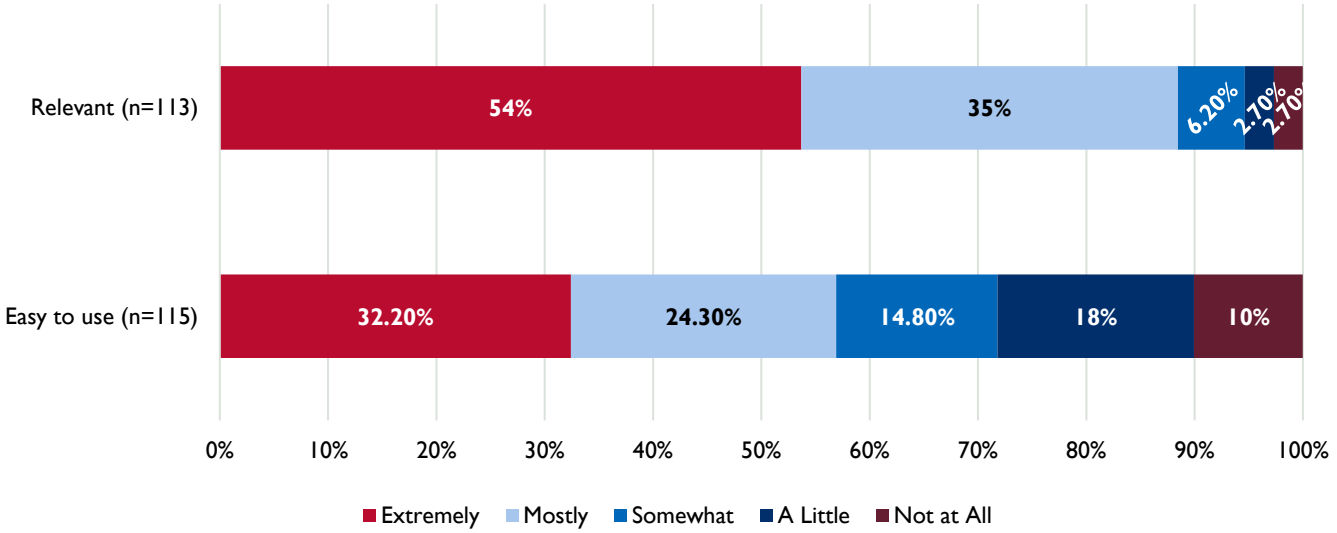
REVIEWER FEEDBACK

Following the review of the reports, reviewers were asked to provide feedback on the ASQ Tool. Their feedback was solicited to get a better understanding of the future of the tool and inform final revisions to the ASQ Tool to ensure it is user-friendly and appropriate.

FINDING 15: Most reviewers found the ASQ Tool relevant and easy to use to inform the research and evaluation process.

As shown in Exhibit 27, most reviewers indicated that the ASQ Tool is relevant to research and evaluation. Most reviewers also indicated that it is extremely or mostly easy to use to inform the research and evaluation process.

Exhibit 27: Reviewers' perspective on the relevance and ease of use of the ASQ Tool



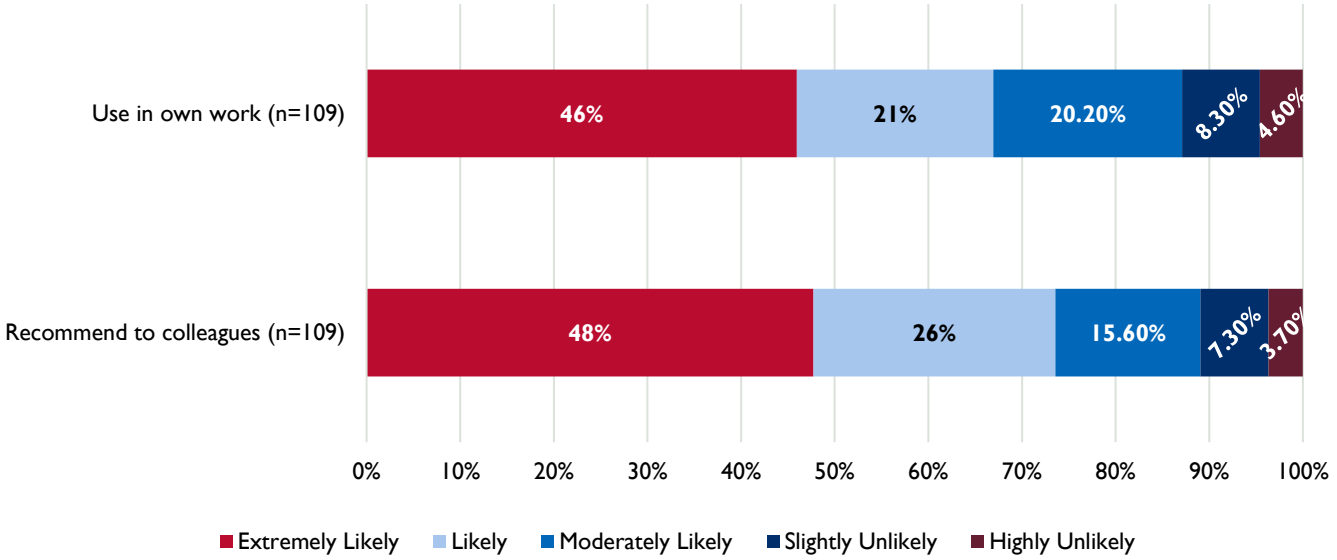
The reviewers' comments provided additional insight into the variation among responses about relevance and ease of use, indicating that the tool is user-friendly but might be too long for those implementing research or evaluation activities. For example, a reviewer commented that, "the tool is easy but long." Based on the reviewers' comments, the tool's strengths include extensive explanations of each item and the use of examples for "Partial" scores. However, both of those strengths also contributed to the length of the tool. Interestingly, reviewers also indicated that the tool may not be long enough, stating that it is not entirely relevant to the research and evaluation process because it does not go "sufficiently in-depth."

A key takeaway from these findings is that the audience for the tool affects its perceived relevance and use. One reviewer stated that the relevance "depends on the audience." Another clarified that for "someone who wants to put more energy into understanding the use of the report rather than whether the report ticks all the boxes, the tool could be frustrating."

FINDING 16: Most reviewers are likely to use the ASQ Tool and recommend it to their colleagues.

Most reviewers indicated that they are extremely likely or likely to use the ASQ Tool in their work, as shown in Exhibit 28. In addition, most reviewers indicated they are extremely likely or likely to recommend the tool to their colleagues.

Exhibit 28: Likelihood that reviewers will use and recommend the ASQ Tool



Similar to their perspectives on the relevance and ease of use of the ASQ Tool, reviewers indicated that the tool's audience will largely dictate whether it is used or recommended to colleagues. For example, one reviewer said, "I am not sure I would use this tool to assess quality—I think it is too focused on the technical aspects of statistical analysis and quality and less focused on whether these reports are in fact useful to anyone or used in a way to help in strategic decision-making." However, other reviewers indicated that they had already recommended that colleagues use the tool as a reference during the report writing phase. One reviewer wrote, "If a colleague is looking for a guide on how to write a study, this can be helpful." This reinforces the importance of audience, because reviewers indicated they would use the tool or recommend it to those who are writing about research and evaluations. Other comments indicated that the tool will be used during the design phase of research and evaluation activities but could be too "cumbersome and technical" during the implementation phase.

CONCLUSIONS

The principles that need the most support are ethics, cultural appropriateness, and robustness of methodology. These are the weakest principles of quality in terms of the percentage of reports meeting "Minimum Adequacy" and the item-level data demonstrate low adherence for most of the items. As seen in Exhibit 29, the item-level data show that less than 40 percent of studies adhered to items regarding cost analysis, research clearances, validating findings with local stakeholders, and study hypotheses. Of these, cost analysis and study hypotheses are not applicable to all studies. The items under cultural appropriateness and ethics were the weakest overall; no more than half of the studies reviewed adhered to even the strongest of these items.

Exhibit 29: Strongest and weakest items

PRINCIPLE OF QUALITY	STRONGEST ITEM(S)	WEAKEST ITEM(S)
STUDY DESIGN PHASE		
Conceptual Framing	Are clear study questions that are appropriate to the stated purpose of the study included in the report? (77.7%)	If applicable, are study hypotheses included in the report? (39.7%) Does the report acknowledge/draw upon existing relevant research? (59.8%)
Robustness of Methodology	Is the methodology appropriate for answering posed study questions? (75.4%)	Does the report mention steps to mitigate common biases or threats to the integrity of the study? (50.0%)
Cultural Appropriateness	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? (51.7%)	Does the report list steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders and incorporate stakeholder feedback? (38.8%)
Ethics	Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report? (45.9%)	Was/were research clearance(s) appropriate to the study obtained prior to starting data collection, as documented in the report? (32.2%)
STUDY IMPLEMENTATION PHASE		
Validity	Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated? (79.2%)	If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates? (42.9%) Is the report open and clear about how the act of doing the study may have biased the findings? (46.3%)
Reliability	Does the report document the steps taken to ensure that data were collected with a high degree of reliability? (69.7%)	For studies where data are collected by a team, was inter-rater reliability established and documented? (39.5%) If applicable to the study methods, was inter-coder reliability established and documented for studies where data were coded by a team? (37.8%)
STUDY IMPLEMENTATION PHASE		
Openness and Transparency	Is the report open and clear about limitations inherent to the study design and with its implementation? (79.5%)	For impact evaluations, is a cost analysis of the intervention being evaluated included in the report? (25.5%) Is the report open about potential biases due to the study team composition? (45.5%)
Cogency	Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)? (83.6%)	Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings? (63.9%)

Across each study phase, items related to biases were among the weakest, based on evidence provided in the reports. Items related to biases received lower scores in design, implementation, and report writing. Under study design, the lowest scoring item under robustness of methodology was the mitigation of biases or threats

to the integrity of the study. In study implementation, the item with the second-lowest adherence in the validity principle was "Is the report open and clear about how the act of doing the study may have biased the findings?" In the report writing phase, the lowest-scoring item under the openness and transparency principle was "Is the report open to potential biases due to the study team composition?" These results indicate that planning for and mitigating biases is challenging for those conducting education studies, as is reporting on bias. They also highlight how connected the study phases are: perhaps if biases are not planned for, it makes sense that mitigating them in implementation will be a challenge and it will be hard for a report writer to write about bias.

While cultural appropriateness remains a weakly adhered to principle since the previous review, it has improved, which may indicate that those designing, implementing, and writing about education studies are paying it more attention. This is not the case for robustness of methodology, which declined since the previous review. In fact, robustness of methodology was stronger than validity and reliability in the previous review.[7] In the current review, this dynamic switched; the two principles related to study implementation, validity and reliability, are stronger than robustness of methodology, cultural appropriateness, and ethics, three of the four principles under study design. This could mean that study implementation is stronger than study design. However, given that study design dictates many elements of study implementation, it is more likely that report writing is not accurately capturing the items under all the principles of quality.

Each of the findings may point to an issue with report writing. All the items under each principle of quality may have been addressed by those who designed and implemented the studies, but if there is no evidence provided in the report, the reviewers' judgment may not reflect the study design and implementation. This is notable because cogency is the strongest principle of quality. There could also be an issue of reviewer participation bias, because those who participated in the study are experts, researchers, evaluators, and individuals who understand technical jargon. The reviewers' experience may have allowed them to make connections and assumptions about the items in the ASQ Tool that are not clearly reported, raising the possibility that reviewer participation bias exists for all principles of quality.

Reviewers' feedback indicated that the ASQ Tool is relevant and user-friendly, although the audience is an important factor to consider when deciding whether to recommend the tool for use. Reviewers suggested that the tool is more useful for designing and writing about research and evaluation than for implementing them. Most reviewer feedback was positive, although there were contradictions in reviewers' perceptions of the tool length. Some reviewers indicated that the explanations and examples included in the tool made it more user-friendly; others mentioned that the tool was too long and overly technical. The reviewers did not have access to the one-page checklist version of the tool, which was developed after the review. The checklist was originally designed to be a stand-alone resource but based on reviewer feedback it will be included in the full ASQ Tool document to give the audience access to multiple versions of the tool in one location.[8] This will enable users to select different versions of the tool as needed and will be useful to address the needs of different researchers and evaluators. For example, those who need a quick reminder may find the checklist most appropriate, while those still developing their research and evaluation skills may find the full version of the tool more appropriate.

---

[7] Management Systems International (MSI). *Assessment of the Quality of USAID-Funded Evaluations. Education Sector, 2013-2016.* USAID, January 2018. https://pdf.usaid.gov/pdf_docs/pa00srw1.pdf

[8] The ASQ Tool and Guidance document, including the one-page checklist, is undergoing final clearance by USAID. It will be available on EducationLinks when the final version is approved.

RECOMMENDATIONS

As part of this review, DEEP developed a separate Internal Action Plan that includes a comprehensive set of recommendations for USAID. For the sake of transparency and accountability, a summary of these recommendations follows.

IMMEDIATE ACTIONS

In the next six months, DEEP proposes targeted socialization of the ASQ Tool with USAID Missions and Regional Bureaus. This includes developing materials to help Missions and Regional Bureaus communicate the tool and discuss expectations for its use with IPs. DEEP also proposes socializing the tool with IPs and other partners, which includes pinning the tool to the EducationLinks home page to increase awareness and enable access. DEEP also recommends developing materials to support understanding of the tool, such as infographics about how and when to use it and asynchronous training materials similar to the orientation materials employed during this review.

SHORT-TERM ACTIONS

In 6 to 12 months, DEEP proposes focusing on developing and holding a series of targeted training sessions about the study process, based upon the findings of this review and the ASQ Tool. This includes conducting separate training sessions for each study phase to address the applicable principles of quality for the study phase. These sessions would focus on addressing weaknesses identified during this review and supporting overall understanding about how to design, implement, and write about high-quality studies.

LONG-TERM ACTIONS

DEEP proposes conducting a rapid study of the ASQ Tool socialization efforts after one year. The results of this study will provide insight into additional requirements to support continued use of the ASQ Tool. DEEP also proposes engaging with researchers and evaluators working in education subsectors through virtual interactive sessions in which participants discuss their approaches to and challenges with applying practices under each principle of quality. In addition, DEEP proposes developing additional materials to support ASQ Tool understanding, such as annotated examples of high-quality report sections.

ANNEXES

ANNEX A: ASSESSMENT OF STUDY QUALITY (ASQ) TOOL USED IN THE REVIEW (DECEMBER 2022 DRAFT)

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 as well as 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 Study Quality Assessment Tool (also referred to as "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 **intended for social science researchers and evaluators, commissioners of social science research and evaluations, and users of social science research and evaluation**.

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. Next, we provide a detailed item description table, which includes detailed descriptions and a rubric for each item included in the tool, followed by the complete list of sources we referenced to develop the tool.

A NOTE ON TERMINOLOGY

Following the tradition set forth in the social sciences[i,ii], 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)[iii] working group. For the updated version, we added an additional principle, ethics, to the framework to address the importance of protecting human subjects. From

these principles of quality, 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 eight principles are defined in Exhibit 1.**

Exhibit 1: Principles of Quality

The *Principles of Quality*

Conceptual Framing: High quality studies are situated within a theory, acknowledge existing research, and pose specific questions.

Openness and Transparency: High quality studies are transparent about the design, methods, data, and limitations.

Cultural Appropriateness: High quality studies consider the local context when designing the study and the data collection tools.

Robustness of Methodology: High quality studies use designs and methods that are appropriate to the stated purpose and questions.

Validity: High quality studies produce credible and accurate results.

Reliability: High quality studies use consistent approaches and produce consistent results.

Cogency: High quality studies provide a clear, logical thread linking the purpose to the methods and data to the conclusions.

Ethics: 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.

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.

HOW TO USE THE TOOL

The tool can be used (1) when commissioning a study, (2) when designing and conducting a study, and (3) when reading or reviewing a study report. **Not all items 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 requirements to be included when developing 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.

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 item 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 implementing partners, donor agency staff, practitioners, graduate 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 items 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 activities. 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.

2022 STUDY QUALITY ASSESSMENT TOOL[iv]

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?		
OPENNESS AND TRANSPARENCY	[6] Is the report open and clear about limitations **inherent to the study design** and **with its implementation**?		
	[7] For evaluations, is the report open and clear about study limitations due to issues with the **implementation of the intervention** being evaluated?		
	[8] Is the report open and clear about potential biases due to the study team composition?		
	[9] Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made?		
	[10] For impact evaluations, is a cost analysis of the intervention being evaluated included?		
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?		

PRINCIPLE OF QUALITY	EXPERIMENTAL/QUASI EXPERIMENTAL, INCLUDING IMPACT EVALUATION	OBSERVATIONAL, INCLUDING PERFORMANCE EVALUATION	
		QUANTITATIVE	QUALITATIVE
	[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?		
ROBUSTNESS OF METHODOLOGY	[15] Is the methodology appropriate for answering posed study questions?		
	[16] Does the counterfactual meet standards of rigor?		
	[17] Does the analysis include triangulation of data from different sources?		
	[18] Does the report mention steps to mitigate common biases or threats to the integrity of the study?		
	[19] 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?		
VALIDITY	[20] Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated?		
	[21] Is the report open and clear about how the act of doing the study may have biased the findings?		
	[22] Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections?		
	[23] Does the report address the external validity (for quantitative studies) or the transferability (for qualitative studies) of findings?		
	[24] If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates?		
RELIABILITY	[25] Does the report document the steps taken to ensure that data were collected with a high degree of reliability?		
	[26] If applicable, was internal consistency of the instrument(s) established and documented?		

PRINCIPLE OF QUALITY	EXPERIMENTAL/QUASI EXPERIMENTAL, INCLUDING IMPACT EVALUATION	OBSERVATIONAL, INCLUDING PERFORMANCE EVALUATION	
		QUANTITATIVE	QUALITATIVE
	[27a] For studies where data is collected by a team, was inter-rater reliability established and documented?		[27b] If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team?
	[28] Does the report adequately address missing data/non-response?		
COGENCY	[29] Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings?		
	[30] Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)?		
	[31] If recommendations are made, are they specific, relevant, actionable, and based on the findings?		
	[32] Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations?		
	[33] Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings?		
ETHICS	[34] Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report?		
	[35] Was/were research clearance(s) appropriate to the study obtained and documented prior to starting data collection?		

USE CASE: SYSTEMS THINKING APPROACHES

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 outcome.[v,vi,vii] 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.[viii, ix,x, xi] Foundational to systems thinking approaches is the exploration of the **boundaries** of the system and subject being studied, the **perspectives** of various stakeholders about subject being studied, and the **inter-relationships** between sub-systems and stakeholders that impact the subject being studied.[xii] Systems thinking is called out as a use case for two reasons:

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.

(1) to respond to a growing interest in using systems thinking approaches in international development and the social sciences, and

(2) to acknowledge the unique terminology of the systems thinking body of inquiry.

This use case presents the eight principles of quality through the lens of systems thinking, to provide a frame for how the items in the tool will shift when applying a systems thinking approach. The items in the tool are still applicable, based upon the method selected.

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.
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
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 should be designed to examine the inter-relationships between different stakeholder groups.
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.
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

PRINCIPLE OF QUALITY	SYSTEMS THINKING APPROACH
	systems/sub-systems. Alternative interpretations about the inter-relationships between sub-systems are 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.
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 are 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.
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". 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.

ITEM DESCRIPTION AND SOURCE

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	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. For studies using a systems thinking approach, 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.	ADS 201maa; BE2 Guidance Note on Qualitative Research, page 9 (figure 1) and pages 61-62. *Additional source: BE2, page 16*
Conceptual Framing: Study hypotheses included	[2] If applicable, are study hypotheses included in the report? **[All study types]**	yes/no/NA	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. Quantitative and qualitative studies may require hypotheses, depending on the study design and purpose. For studies using a systems thinking approach which requires hypotheses, the hypotheses must be based upon the three foundational concepts: boundaries, inter-relationships, and perspectives. "NA" score should be given for quantitative and qualitative study designs which do not require hypotheses.	BE2, Checklist (page 28); Barroga and Matanguihan, 2022, page 7. *Additional source: Lamont and White, 2005, page 10.*

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? **[All study types]**	yes/partial/ no	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. 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. For evaluation studies or studies addressing an intervention, this means that questions should be based on the intervention's theory of change or results framework. For studies using a systems thinking approach, the conceptual or theoretical framework should address boundaries, perspectives, and inter-relationships. The assumptions should address the inter-relationships between system agents/sub-systems. "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.	BE2, page 16 and Checklist (page 28). *Additional source: USAID Evaluation Policy, page 8*
Conceptual Framing: Study acknowledges/draws upon existing country-specific research	[4] Does the report acknowledge/draw upon existing relevant research? **[All study types]**	yes/partial/ no	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. "Partial" score could be given when only some of the questions are informed by existing knowledge.	BE2, Checklist (page 28)
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? **[All study types]**	yes/partial/ no	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. "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.	USAID Evaluation Policy, page 8; BE2 Guidance Note on Qualitative Research, page 8
Conceptual Framing: Conclusion	Conceptual framing: Conclusion **[All]**	adequate/ not	**Adequate:** Overall, this study demonstrates adherence to principles of conceptual framing **Not Adequate:** This study contains major deficiencies in demonstrating adherence to principles of conceptual framing or provides insufficient information for determining this	

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Conceptual Framing: Justification	Conceptual framing: Notes/Justification **[All]**		For instance: *"The authors acknowledge existing research and make clear how their analyses sit within the context of existing work. They provide a theoretical framework in the report, where they outline their major assumptions. The study also poses specific research questions."*	
Openness and Transparency: Open about limitations to the study design and to implementing the study	[6] Is the report open and clear about limitations **inherent to the study design** and **with its implementation**? **[All study types]**	yes/partial/ no	Limitations to the implementation of the study should be clearly presented. Clarity around study limitations is particularly important if they directly impact the evaluator's/researcher's ability to credibly and effectively answer a study question or impact 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.	BE2, page 17. *Additional sources: ADS 201 mah; Blaikie and Priest, 2019, page 15; Greener, 2018, page 568*
Openness and Transparency: Open about how the intervention impacts the study	[7] For evaluations, is the report open and clear about study limitations due to issues with the **implementation of the intervention** being evaluated? **[Evaluations]**	Yes/partial/ no/NA	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. "NA" score should be given to studies that do not evaluate a specific intervention.	BE2, page 17.
Openness and Transparency: Open about potential biases due to the study team composition	[8] Is the report open about potential biases due to the study team composition? **[All study types]**	yes/partial/ no	USAID encourages study teams to include at least one evaluation specialist, host country team members, and a team leader who is external to USAID. USAID also requires that evaluation team members certify their independence by signing statements **disclosing any conflict of interest or fiduciary involvement with the project or program** 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.	BE2, Checklist (page 29-29). *Additional source: USAID Evaluation Policy, page 8*

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Openness and Transparency: Methodology explained in detail	[9] Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made? **[All study types]**	yes/partial/ no	USAID requires that an evaluation report identifies 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; what types of interviews were conducted; with whom they were conducted (e.g., key informant interviews, individual interviews with beneficiaries, group interviews) and; detailed information on the kinds of analyses that were conducted (e.g., correlations, regressions, content analysis, pattern analysis). Researchers/evaluators using a systems thinking approach 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 on 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.	ADS 201maa. *Additional sources: USAID Evaluation Policy page 8*

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Openness and Transparency: Cost analysis	[10] For impact evaluations, is a cost analysis of the intervention being evaluated included in the report? **[Impact evaluations]**	Yes/no/NA	Reporting on the findings from a cost analysis should be clear on all elements that may be useful for making decisions. USAID requires all impact evaluations to include a cost analysis of the 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: ● Details about the intervention, such as the ToC, 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 ● Details about the beneficiaries: who and where they are, including geography, age and sex, marginalization status, and other relevant details ● Cost estimates: what is included/not included and why, whether recurrent and non-recurrent costs are separate, whether contributions were costed out or listed alongside final estimates ● Perspectives: whose perspective(s) are reflected in the cost estimates ● Cost modeling: prospective or respective, assumptions, and data limitations ● Major cost drivers: what they are, what factors have greatest influence on cost estimates (i.e., contextual factors, beneficiary characteristics, intervention features, etc.) ● Computation: how cost estimates were computed, including assumptions and the computations in an annex "NA" score should be given for studies that do not include an impact evaluation.	USAID Evaluation Policy, page 2; USAID Cost Analysis Guidance for USAID-Funded Education Activities, page 79.
Openness and Transparency: Conclusion	Openness and transparency: Conclusion **[All]**	adequate/ not	**Adequate:** Overall, this study demonstrates adherence to principles of openness/transparency **Not Adequate:** This study contains major deficiencies in demonstrating adherence to principles of openness/transparency or provides insufficient information for determining this	
Openness and Transparency: Justification	Openness and transparency: Notes/Justification **[All]**		For instance: *"The authors are transparent about the design and methods that have been employed in the evaluation as well as the data (and resulting sample) that have been gathered and analyzed. This allows for the study to be repeated and corroborated."*	

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? **[All study types]**	yes/no	The study questions should be informed by relevant local stakeholders. This could be done during in-country design workshops as well as through meeting with the ministry or other relevant stakeholders. The study methodology should be informed by relevant local stakeholders. This could be done during in-country design workshops as well as through meeting with the ministry or other relevant stakeholders. Where appropriate, the study should use inclusive, gender-sensitive, and participatory research methods, such as a Gender and Power (GAP) Analysis. Studies using systems thinking approaches should ensure that all possible perspectives from within each bounded system are reflected in the report.	ADS 201sae; Save the Children's Gender and Power (GAP) Analysis Guidance.
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? **[All study types]**	yes/partial/ no	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 (back-translation) 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 if the results of the pilot were used 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 an item 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 who or how the results were used.	BE2, page 20; EGRA Toolkit, 2nd Edition, page 92.

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
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? **[All study types]**	yes/no	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 help 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).	EGRA Toolkit, 2nd edition, page 122. *Additional source: BE2 Guidance Note on Qualitative Research, page 61*
Cultural Appropriateness: Data collection and analysis allows for disaggregation by locally relevant stratifiers	[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? **[All study types]**	yes/partial/no	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. "Partial" score should be assigned when the study is purposeful with considering gender in data collection or considering variable impacts on gender but not any other stratifiers.	BE2, page 20.
Cultural Appropriateness: Conclusion	Cultural appropriateness: Conclusion **[All]**	adequate/not	**Adequate:** Overall, this study demonstrates adherence to principles of cultural appropriateness. **Not Adequate:** This study contains major deficiencies in demonstrating adherence to principles of cultural appropriateness or provides insufficient information for determining this.	
Cultural Appropriateness: Justification	Cultural appropriateness: Notes/Justification **[All]**		For instance: *"The evaluation describes systematic processes used to check for the cultural relevance of measurement items (for example, in the absence of lists of age-specific words for Bangla-speaking children, a list was created of words that fit two criteria: they should be known to grade 1 or 2 children but unknown to preschoolers, and they should be used in the storybooks). Thus, the instrument used is culturally sensitive. The analysis is also culturally sensitive, as it discusses the factors that undermine or promote educational outcomes within the Bangladeshi context. The study discusses the use of two supply-and-demand side interventions – a school-only grant and a school grant plus an education allowance – which the authors discuss in relevance to the context, where grants are used to provide key inputs to schools while the education allowance provides a conditional monetary incentive for out-of-school children to attend school."*	

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Robustness of Methodology: Methodology appropriate for answering posed study questions	[15] Is the methodology appropriate for answering posed study questions? **[All study types]**	yes/partial/ no	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 method (or methods) for a particular study also balances cost, 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; for such studies, the assessment of the methodology must include the review of the study design vis-à-vis each stated study questions. "Partial" score could be given if the methodology proposed is appropriate for some, but not all posed questions.	USAID Evaluation Policy, page 9.
Robustness of Methodology: Counterfactual meet standards of rigor	[16] Does the counterfactual meet standards of rigor? **[Exp./Quasi]**	yes/no/NA	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. Considerations about its rigor may include a review of information in the report about baseline equivalence, differential attrition, etc. "NA" score should be given if the study is not an Impact Evaluation or a study using an experimental/quasi experimental design.	USAID Evaluation Policy, page 2.
Robustness of Methodology: Data triangulation described as part of methodology	[17] Does the analysis include triangulation of data from different sources? **[All study types]**	yes/partial/ no	Typically, stronger bodies of evidence are likely to emerge if similar findings are obtained from different types of data (e.g., tests, interviews, observations) and respondent types (e.g., students, parents, teachers). It is important that contradictory data be taken into account when discussing the findings. "Partial" score could be given if data from different sources are presented but the findings don't connect them into a coherent narrative. "NA" score should be given if the study does not use multiple data sources.	CASP, Qualitative Checklist, page 4. *Additional sources: BE2, page 26; BE2 Guidance Note on Qualitative Research, page 61*

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Robustness of Methodology: Addressed internal validity, either threats to inference or common biases	[18] Does the report mention steps to mitigate common biases or threats to the integrity of the study? **[All study types]**	yes/partial/ no	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. Some common threats to the integrity of **quantitative studies** may include non-equivalence at baseline, non-compliance, spillover, systematic attrition. Some common biases for quantitative studies may include confounding bias, selection bias, experimenter bias. Some common threats to the integrity of **qualitative studies** may include threats to trustworthiness such as participant non-availability. Some common biases for qualitative studies may include selection bias, researcher bias. Other threats to the integrity/trustworthiness and other common biases may be discussed in the report as well. "Partial" score could be given if some, but not all threats or biases identified are discussed.	USAID Evaluation Policy, page 11.

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Robustness of Methodology: Described sampling approach and parameters used to compute sample size	[19] 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? **[All study types]**	yes/partial/ no	For **quantitative studies,** 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 may 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) or 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), (3) and the sample size. 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. For **Qualitative studies,** 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 to include, 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. A study using a systems thinking approach 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. "Partial" score should be given if only some of these details were discussed or presented.	JPAL's Running Randomized Evaluations, page 271; CASP, Qualitative Checklist, page 3; EGRA Toolkit, 2nd Edition, pages 117; StataCorp's Survey Data Reference Manual, page 3; BE2 Guidance Note on Qualitative Research, pages 40-41. *Additional sources: EGRA Toolkit, 2nd Edition, pages 120 and 175; UIS Handbook on Measuring Equity in Education, page 74*

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Robustness of Methodology: Conclusion	Robustness of methodology: Conclusion **[All]**	adequate/not	**Adequate:** Overall, this study demonstrates adherence to principles of appropriateness/rigor of chosen methodology **Not Adequate:** This study contains major issues with the appropriateness of the chosen methodology, major deficiencies in the rigor with which it was applied or provides insufficient information for determining this	
Robustness of Methodology: Justification	Robustness of methodology: Notes/Justification **[All]**		For instance: *"The study aims to identify and examine specific effects of receiving grants alone compared to receiving grants as well as training on student learning outcomes. The study clearly aims to establish a causal linkage between grants versus grants/training on student outcomes. The experimental design was, therefore, most appropriate to answer the research question. The study demonstrates rigorous application of the experimental technique within The Gambian setting. The authors clearly describe the interventions and adopt all the rigors of a well-applied randomization."*	
Validity: Addressed the construct validity of the data collection tools	[20] Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated? **[All study types]**	yes/partial/no	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.	BE2, page 24.
Validity: Addressed ecological validity of findings	[21] Is the report open and clear about how the act of doing the study may have biased the findings? **[All study types]**	yes/no	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). 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. This might include discussions about whether the implementer may have brought in irreproducible energies that accountable for the success of a pilot but that might be absent in a scale-up.	BE[2], page 25.

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Validity: Address the credibility of the findings	[22] Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections? **[All study types]**	Yes/no	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. For qualitative studies, 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)	BE2, page 17. *Additional sources: BE2 Guidance Note on Qualitative Research, page 61; Chilisa, 2020; Lincoln and Guba, 1985)*
Validity: Addressed the external validity or transferability of findings to other contexts	[23] Does the report address the external validity (for quantitative studies) or transferability (for qualitative studies) of findings? **[All study types]**	yes/partial/ no/NA	**Quantitative findings** 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. **Qualitative findings** 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. "Partial" score could be given if the external validity or transferability of some, but not all key findings, are discussed in the report. "NA" score could be given in case this study did not intend to have data from a sample extrapolated to a population.	BE2, Checklist (p.29); Chilisa, 2019, page 216. *Additional sources related to transferability in qualitative research: Lincoln and Guba, 1985; Shenton, 2004; Williams and Morrow, 2009*
Validity: Confidence intervals reported around point estimates	[24] If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates? **[Quantitative]**	yes/no/NA	USAID recommends that the margin of error be reported along with the findings from statistical samples. "NA" score should be given if the study does not use inferential statistical methods.	ADS 201sae.
Validity: Conclusion	Validity: Conclusion **[All]**	adequate/not	**Adequate:** Overall, this study demonstrates adherence to principles of validity. **Not Adequate:** This study contains major deficiencies in establishing the measurement, internal, external or ecological validity or provides insufficient information for determining this.	

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Validity: Justification	Validity: Notes/Justification **[All]**		For instance: *"The authors describe steps they took to address the validity of the study. For example, items included in the test had to relate directly to what grade 5 children would be expected to know at the start and end of the school year and statistical analyses were conducted to assess the internal consistency of questions in order to refine and adjust the assessment tools (measurement validity). In assessing learning progress of pupils in grade 5, the study included initial test scores into the estimation and controlled for background factors that may generate biases (internal validity). The study is based on longitudinal data collected from 5 provinces out of 58 in Vietnam, the generalizability of the findings is somewhat questionable (external validity), and there is no discussion of whether the findings could have been influenced by the process of research itself (ecological validity). While it could be improved, overall this study meets basic standards of scientific validity."*	
Reliability: Steps taken to ensure that data were reliably collected	[25] Does the report document the steps taken to ensure that data were collected with a high degree of reliability? **[All study types]**	yes/partial/ no	USAID recommends that data collection methods 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 qualitative studies may include audio recording, videotaping and transcribing interviews. In studies using a systems thinking approach, 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.	ADS 201sae.
Reliability: addressed internal reliability/consistency of instruments	[26] If applicable, was internal consistency of the instrument(s) established and documented?	yes/no/NA	Instruments which measure a scale comprised of a set of items or indicators must ensure that all of 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. "NA" score should be given for studies which do not use multi-item instruments where multiple items are intended to measure the same variable.	EGRA Toolkit, 2nd Edition, page 93-94; Remler and Van Ryzin, 2021, pg. 135.

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Reliability: Inter-rater reliability was established	[27a] For studies where data is collected by a team, was inter-rater reliability established and documented? **[Quantitative]**	Yes/no/NA	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. "NA" score should be given for qualitative studies or for survey/assessment studies in which data was not collected by multiple enumerators.	EGRA Toolkit, 2nd Edition, page 89.
Reliability: Inter-coder reliability was established	[27b] If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team? **[Qualitative]**	Yes/no/NA	In qualitative studies 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. "NA" 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.	Saldaña, 2021, page 52.
Reliability: Target and actual sample sizes reported and non-responses bias discussed	[28] Does the report adequately address missing data/non-response? **[All study types]**	yes/partial/ no	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 qualitative studies. For quantitative evaluations, the report may also mention using post-stratification to adjust weights for non-response. "Partial" score could be given if information about valid responses is provided to some, but not all data used in the findings.	What Works Clearinghouse Procedures and Standards Handbook Version 3.0, page D.4. *Additional sources: What Works Clearinghouse Standards Handbook Version 4.0 page 65; BE2 Guidance Note on Qualitative Research page 46*
Reliability: Conclusion	Reliability: Conclusion **[All study types]**	adequate/n ot	**Adequate:** Overall, this study demonstrates adherence to principles of reliability. **Not Adequate**: This study contains major deficiencies in establishing the reliability of the measurement or provides insufficient information for determining this.	

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Reliability: Justification	Reliability: Notes/Justification **[All]**		For instance: *"This study used multiple researchers to undertake school observations and interviews; the researchers checked their own conclusions with each other and then cross-checked them against the wider analytical team to analyze between schools. The team ensured that different types of data were collected – observations, interviews and document analysis – to triangulate findings and take into account the variety of possible contexts. The authors also provide a good example of how to enhance the reliability of qualitative analysis: interviews were videotaped and transcribed."*	
Cogency: Answers to all study questions, including sub-questions, included	[29] Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings? **[All study types]**	yes/partial/ no	The purpose of a report is to provide the evaluators' or researchers' findings and recommendations on each and every study question. Accordingly, 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. "Partial" score could be given if the answers are provided in the report but not the Executive Summary.	ADS 201mah; E3 Sectoral Synthesis, Checklist, question 17, page 145.
Cogency: Written in a language adequate to its stated audience	[30] Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)? **[All study types]**	yes/no	Reports should be written in an accessible way to non-experts. Excessive use of research terminology is also undesirable; the report should favor terminology that its intended audience is expected to be familiar with.	USAID Evaluation Policy, page 11.
Cogency: Recommendations are relevant, actionable, and based on findings	[31] If recommendations are made, are they specific, relevant, actionable, and based on the findings? **[All study types]**	Yes/no/NA	It is important that recommendations be practical, action-oriented, and specific as well as relevant to the study. "NA" score should be given to studies which are not intended to produce recommendations.	USAID Evaluation Policy, p11. *Additional source: UNEG 2010, page 6*

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Cogency: The report is logically connected from start to finish	[32] Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations? **[All study types]**	Yes/no	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. 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. Studies using a systems thinking approach 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 must address the role of the sub-systems and the agents in the system. 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.	E3 Sectoral Synthesis, Checklist, question 32, page 145.
Cogency: Visuals are helpful for a non-technical audience to understand the findings	[33] Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings? **[All study types]**	yes/partial/ no	Visuals must be used to facilitate understanding of the findings by general 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. "Partial score" could be given if the report uses visuals to an insufficient extent.	EGRA Toolkit 2nd Edition, page 120.
Cogency: Conclusion	Cogency: Conclusion **[All evaluation types]**	adequate/ not	**Adequate:** Overall, this study demonstrates adherence to principles of cogency. **Not Adequate:** This study contains major deficiencies in demonstrating adherence to principles of cogency or provides insufficient information for determining this.	
Cogency: Justification	Cogency: Notes/Justification **[All evaluation types]**		For instance: *"The evaluation contains a clear, logical argumentative thread that runs through the entire report. This links the conceptual framework for the study to the data and analysis, and, in turn, to the conclusions. The conclusions are backed up by the evaluation findings."*	

PRINCIPLE OF QUALITY	QUESTION	SCORE	DESCRIPTOR	SOURCE
Ethics: Protection of human subjects is integrated in the study	[34] Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report? **[All study types]**	Yes/no	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 project the human subjects involved. USAID requires that evaluations are conducted to the highest ethical standards. As such, evaluations must be ethical, fair, and take into consideration cultural and contextual factors that may influence findings or how the findings are used. Informed consent/assent must be received from all study participants. There are many resources to provide in-depth ethical guidelines for research and evaluation in education and studies involving children and vulnerable populations. 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. 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.	USAID Evaluation Policy, page 9. *Additional sources: BE2 Guidance Note on Qualitative Research pages 56, 58-59; Save the Children's Gender and Power Analysis Guidance, step 4*
Ethics: Research clearances were obtained	[35] Was/were research clearance(s) appropriate to the study obtained and documented prior to starting data collection? **[All study types]**	Yes/no	USAID-funded studies which involves human subjects must consult an IRB and receive IRB approval. Studies funded by other donors must follow IRB/ERC 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.	EGRA Toolkit, 2nd Edition, page 13. *Additional source: BE2 Guidance Note on Qualitative Research, page 57*
Ethics: Conclusion	Ethics: Conclusion **[All study types]**	adequate/ not	**Adequate:** Overall, this study demonstrates adherence to principles of ethics. **Not Adequate:** This study contains major deficiencies in demonstrating adherence to principles of ethics or provides insufficient information for determining this.	
Ethics: Justification	Ethics: Justification **[All study types]**		For instance: *This study clearly describes the processes for protecting the human subjects and mitigating risks to the study participants. The study also provides sufficient documentation, such as the informed consent protocols and documentation of IRB approval.*	

REFERENCES

Alcott, B., P. Rose, R. Sabates, and R. Torres. *Handbook on Measuring Equity in Education*. Montreal: UNESCO Institute for Statistics, 2018. http://uis.unesco.org/sites/default/files/documents/handbook-measuring-equity-education-2018-en.pdf.

Allen, Will. "Systems Thinking," Learning for Sustainability. Accessed October 28, 2022. https://learningforsustainability.net/systems-thinking/

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.

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). https://doi.org/10.3346/jkms.2022.37.e121.

Blaikie, Norman, and Jan Priest. *Designing Social Research: The Logic of Anticipation*. Medford, MA: Polity, 2019.

Building Evidence in Education (BE2). *Assessing the Strength of Evidence in the Education Sector*. May 2020. https://www.edu-links.org/sites/default/files/media/file/BE2_Guidance_Note_ASE_0.pdf

Chilisa, Bagele. *Indigenous Research Methodologies*. 2nd Edition. Washington, D.C.: Sage Publications, 2019.

Coghlan, David, and Mary Brydon-Miller. "Systems Thinking." In *The SAGE Encyclopedia of Action Research*, edited by David Coghlan and Mary Brydon-Miller. London: SAGE Publications, 2014, 753–54. https://dx.doi.org/10.4135/9781446294406.n333.

Critical Appraisal Skills Programme (CASP). "Qualitative Research Checklist." 2017. http://media.wix.com/ugd/dded87_25658615020e427da194a325e7773d42.pdf.

DeJaeghere, J., V. Morrow, D. Richardson, B. Schowengerdt, R. Hinton, and A. Muñoz Boudet. *Guidance Note on Qualitative Research in Education: Considerations for Best Practice.* London: United Kingdom Department for International Development, prepared for Building Evidence in Education (BE2), 2020.

Glennerster, Rachel and Kudzai Takavarasha. "Running Randomized Evaluations." In *Running Randomized Evaluations*. Princeton, NJ: Princeton University Press, 2013. http://runningres.com/.

Goodman, Michael. "Systems Thinking: What, Why, When, Where, and How?" The Systems Thinker. Accessed October 31, 2022. https://thesystemsthinker.com/systems-thinking-what-why-when-where-and-how/.

Greener, Sue. "Research Limitations: The Need for Honesty and Common Sense." *Interactive Learning Environments* 26, no. 5 (2018): 567–568. https://doi.org/10.1080/10494820.2018.1486785.

Institute of Education Sciences (IES). *What Works Clearinghouse Procedures and Standards Handbook, Version 3.0*. Washington, D.C.: U.S. Department of Education, 2014. https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf.

Institute of Education Sciences (IES). *What Works Clearinghouse Standards Handbook, Version 4.0*. Washington, D.C.: U.S. Department of Education, 2017. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.

Kim, Daniel. "Introduction to Systems Thinking." The Systems Thinker, 1999. Accessed May 9, 2023. https://thesystemsthinker.com/introduction-to-systems-thinking/.

Lamont, Michèle, and Patricia White. *Workshop on Interdisciplinary Standards for Systematic Qualitative Research*. Alexandria, VA: National Science Foundation, 2005. https://www.nsf.gov/sbe/ses/soc/ISSQR_workshop_rpt.pdf.

Lincoln, Yvonna S., and Egon G. Guba. "Establishing Trustworthiness." In *Naturalistic Inquiry.* Newbury Park, CA: SAGE Publications, 1985, 289–327.

Management Systems International (MSI). *Assessment of the Quality of USAID-Funded Evaluations. Education Sector, 2013–2016.* Washington, D.C.: USAID, 2018. https://pdf.usaid.gov/pdf_docs/pa00srw1.pdf.

RTI International. *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*. Washington, D.C.: USAID, 2015. https://www.edu-links.org/sites/default/files/media/file/EGRA%20Toolkit%20Second%20Edition.pdf.

Remler, Dahlia K., and Gregg G. Van Ryzin. *Research Methods in Practice: Strategies for Description and Causation*. Washington, D.C.: SAGE Publications, 2022.

Saldana, Johnny. *The Coding Manual for Qualitative Researchers,* Fourth Edition. Washington, D.C: SAGE Publications, 2021.

Save the Children. "Gender and Power Analysis Guidance." Child Rights Resource Center. 2021. Accessed May 9, 2023. https://resourcecentre.savethechildren.net/document/gender-power-gap-analysis/.

Shenton, Andrew K. "Strategies for Ensuring Trustworthiness in Qualitative Research Projects." *Education for Information* 22, no. 2 (2004): 63–75. https://doi.org/10.3233/EFI-2004-22201.

StataCorp. *Stata: Release 13*. Statistical Software. College Station, TX: StataCorp LP, 2013. https://www.stata.com/manuals13/svy.pdf.

United Nations Evaluation Group (UNEG). "Quality Checklist for Evaluation Reports." UNEG, June 2010. Accessed May 9, 2023. http://www.unevaluation.org/document/detail/607.

USAID. ADS Chapter 201, "Program Cycle Operational Policy." Washington, D.C.: USAID, September 2022. Accessed May 9, 2023. https://www.usaid.gov/ads/policy/200/201.

USAID. ADS Reference 201maa, "Criteria to Ensure the Quality of the Evaluation Report." Washington, D.C.: USAID, September 2016. Accessed May 9, 2023. https://www.usaid.gov/ads/policy/200/201maa.

USAID. ADS Reference 201mah, "USAID Evaluation Report Requirements." Washington, D.C.: USAID, September 2016. Accessed May 9, 2023. https://www.usaid.gov/ads/policy/200/201mah

USAID. ADS Reference 201sae, "USAID Recommended Data Quality Assessment (DQA) Checklist." Washington, D.C.: USAID, September 2016. Accessed May 9, 2023. https://www.usaid.gov/ads/policy/200/201sae.

USAID. *Evaluation Policy*. Washington, D.C.: USAID, October 2020. Accessed May 9, 2023. https://www.usaid.gov/evaluation/policy.

USAID. *Local Systems: A Framework for Supporting Sustained Development. A Mandatory Reference for ADS Chapter 220*. July 28, 2014. Accessed May 9, 2023. https://2012-2017.usaid.gov/sites/default/files/documents/1870/220maq.pdf.

Walls, Elena, Caitlin Tulloch, and Christine Harris-Van Keuren. *Cost Analysis Guidance for USAID-Funded Education Activities*, Second Edition. Washington, D.C.: USAID, 2021. https://www.edu-links.org/resources/usaid-cost-measurement.

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. https://doi.org/10.1080/10503300802702113.

ANNEX B: ONLINE VERSION OF THE ASQ TOOL USED IN THE REVIEW

Thank you for participating in the quality assessment of USAID-funded research and evaluations related to the Education Strategy Goals produced since 2018. In the following survey you will be asked to provide some details about your assigned study report, to evaluate the quality of the study as evidenced in the report based upon the Study Quality Assessment Tool, and to provide feedback on the Tool. We invite you to refer to the Study Quality Assessment Tool for guidance as you review your assigned report(s).

Quality Assessment Tool Survey

2. Study description

Please provide the following details about the study you are reviewing

1 Title of Study Report

Enter answer

2 Year of Report Publication

- ○ 2018
- ○ 2019
- ○ 2020
- ○ 2021
- ○ 2022

3 Region of Study

- ○ Global
- ○ Africa
- ○ Asia
- ○ Europe and Eurasia
- ○ Latin America and the Caribbean
- ○ Middle East and North Africa

4 USAID Education Policy Priority of the Study

Select all that are applicable

- ☐ Basic education and foundational skills
- ☐ Access for marginalized groups
- ☐ Higher education
- ☐ Youth and workforce development
- ☐ Other:

5 Type of Study

Select all that are applicable

- ☐ Experimental/Quasi Experimental/Impact Evaluation
- ☐ Non-Experimental/Quasi Experimental Quantitative or Performance Evaluation
- ☐ Qualitative

Quality Assessment Tool Survey

3. Introduction

The questions about the quality of the study found in this section address items that are typically found in the executive summary, introduction, or background section of a report. Since not every report follows the same structure, you may need to search for the relevant items elsewhere in the report of the study you are reviewing

6 Are clear study questions that are appropriate to the stated purpose of the study included in the report?

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.

"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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 1 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

7 If applicable, are study hypotheses included in the report?

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. Quantitative and qualitative studies may require hypotheses, depending on the study design and purpose.

"Partial" score could be given if some but not all of the hypotheses are explicitly described. Please explain the reason for a partial score in the comment box below (optional).

“NA” score should be given for quantitative and qualitative study designs which do not require hypotheses.

This question corresponds to item 2 in the “Study Quality Assessment Tool”.

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

8 Are the study questions appropriate to the conceptual/theoretical framework or theory of change?

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.

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.

For evaluation studies or studies addressing an intervention, this means that questions should be based on the intervention's theory of change or results framework.

"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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 3 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

9 Does the report acknowledge/draw upon existing relevant research?

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.

"Partial" score could be given when only some of the questions are informed by existing knowledge. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 4 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

10 Does the report explain the local context in sufficient detail as it relates to the study purpose and questions?

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.

"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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 5 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

Quality Assessment Tool Survey

4. Methods

The questions about the quality of the study found in this section address items that are typically found in the methods section of a report. Since not every report follows the same structure, you may need to search for the relevant items elsewhere in the report of the study you are reviewing

11 | Is the methodology appropriate for answering posed study questions?

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 method (or methods) for a particular study also balances cost, 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; for such studies, the assessment of the methodology must include the review of the study design vis-a-vis each stated study questions.

"Partial" score could be given if the methodology proposed is appropriate for some, but not all posed questions. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 15 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

12 Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report?

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 project the human subjects involved. USAID requires that evaluations are conducted to the highest ethical standards. As such, evaluations must be ethical, fair, and take into consideration cultural and contextual factors that may influence findings or how the findings are used. Informed consent/assent must be received from all study participants. There are many resources to provide in-depth ethical guidelines for research and evaluation in education and studies involving children and vulnerable populations. 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. 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.

"Partial" score could be given if the report mentions that ethical principles were integrated in the study without including the informed consent/assent protocols in an annex. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 34 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

13 Was/were research clearance(s) appropriate to the study obtained prior to starting data collection, as documented in the report?

USAID-funded studies which involves human subjects must consult an IRB and receive IRB approval. Studies funded by other donors must follow IRB/ERC 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 report provides documentation of research clearance but it is unclear if clearance was obtained prior to data collection. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 35 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

14 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?

The study questions and methodology should be culturally relevant and informed by relevant local stakeholders. This could be done during in-country design workshops as well as through meeting with the ministry or other relevant stakeholders. Where appropriate, the study should use inclusive, gender-sensitive, and participatory research methods, such as a Gender and Power (GAP) Analysis.

Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 11 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

15 Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made?

USAID requires that an evaluation report identifies 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; what types of interviews were conducted; with whom they were conducted (e.g., key informant interviews, individual interviews with beneficiaries, group interviews) and; detailed information on the kinds of analyses that were conducted (e.g., correlations, regressions, content analysis, pattern analysis).

"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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 9 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

16 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?

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 (back-translation) 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 if the results of the pilot were used 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 an item 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 who or how the results were used. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 12 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

17 Does the counterfactual meet standards of rigor?

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. Considerations about its rigor may include a review of information in the report about baseline equivalence, differential attrition, etc.

Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given if the study is not an Impact Evaluation or a study using an experimental/quasi experimental design.

This question corresponds to item 16 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

18 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?

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.

"Partial" score should be assigned when the study is purposeful with considering gender in data collection or considering variable impacts on gender but not any other stratifiers. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 14 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

19 | 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?

For **quantitative studies**, 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 may 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) or 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), (3) and the sample size. 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.

For **qualitative studies**, 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 to include, 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.

"Partial" score should be given if only some of these details were discussed or presented. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 19 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

20 Does the report document the steps taken to ensure that data were collected with a high degree of reliability?

USAID recommends that data collection methods 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 qualitative studies may include audio recording, videotaping and transcribing interviews.

"Partial" score could be given if steps to ensure the reliability of some, but not all data collected, are described. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 25 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

21 Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated?

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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 20 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

22 If applicable, was internal consistency of the instrument(s) established and documented?

Instruments which measure a scale comprised of a set of items or indicators must ensure that all of 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 internal consistency was established and documented for some but not all instruments, if multiple instruments were used in the study. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given for studies which do not use multi-item instruments where multiple items are intended to measure the same variable.

This question corresponds to item 26 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A

Comment

23 For studies where data is collected by a team, was inter-rater reliability established and documented?

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. The study should describe how inter-rater disagreement was measured and addressed. Inter-rater reliability statistics (like raw agreement and kappa) are measurements of the consistency between assessors. USAID recommends that in addition to an assessor evaluation process during training, that researchers/evaluators have two or more assessors in a sample-base collect data from the same respondent at the same time to compute the inter-rater reliability statistics for the field data collection.

"Partial" score could be given for studies that describe how inter-rater disagreement was measured but does not describe how it was addressed. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given for qualitative studies or for survey/assessment studies in which data was not collected by multiple enumerators.

This question corresponds to item 27a in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

24 If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team?

In qualitative studies 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. Inter-coder reliability statistics (like Krippendorff's alpha and Cohen's kappa) are some measurements of the consistency between coders that may be included in the report.

"Partial" score could be given for studies that describe how inter-coder disagreement was measured but does not describe how it was addressed. Please explain the reason for a partial score in the comment box below (optional).

"NA" 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.

This question corresponds to item 27b in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

Quality Assessment Tool Survey

5. Limitations

The questions about the quality of the study found in this section address items that are typically found in the limitations section of a report. Since not every report follows the same structure, you may need to search for the relevant items elsewhere in the report of the study you are reviewing

25 Is the report open and clear about limitations inherent to the study design and with its implementation?

It is common for researchers or evaluators to encounter expected or unexpected interferences with study design or the implementation of the study. Researchers/evaluators are obligated to include these "study limitations" and a description of the impact they may have had on the study. Limitations to the implementation of the study should be clearly presented. Clarity around study limitations is particularly important if they directly impact the evaluator's/researcher's ability to credibly and effectively answer a study question or impact 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.

"Partial" score could be given if the report mentions limitations without discussing them in detail. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 6 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

26 For evaluations, is the report open and clear about study limitations due to issues with the implementation of the intervention being evaluated?

Interventions frequently evolve in a way that may compromise the integrity of the study design. For instance, a new component of the intervention may be introduced midway through the implementation. Another example might be poor records of the implementation itself making it impossible for the evaluators to establish to what the observed effects might be attributed. Any such limitations of the intervention itself (not the evaluation) should be reported and their implications for the evaluator's ability to credibly answer the evaluation question discussed. 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. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given to studies that do not evaluate a specific intervention.

This question corresponds to item 7 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

27 | Is the report open about potential biases due to the study team composition?

USAID encourages study teams to include at least one evaluation specialist, host country team members, and a team leader who is external to USAID. USAID also requires that evaluation team members certify their independence by signing statements disclosing any conflict of interest or fiduciary involvement with the project or program 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. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 8 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

28 Does the report mention steps to mitigate common biases or threats to the integrity of the study?

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.

Some common threats to the integrity of **quantitative studies** may include non-equivalence at baseline, non-compliance, spillover, systematic attrition. Some common biases for quantitative studies may include confounding bias, selection bias, experimenter bias.

Some common threats to the integrity of **qualitative studies** may include threats to trustworthiness such as participant non-availability. Some common biases for qualitative studies may include selection bias, researcher bias.

Other threats to the integrity/trustworthiness and other common biases may be discussed in the report as well.

"Partial" score could be given if some, but not all threats or biases identified are discussed. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 18 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

29 | Is the report open and clear about how the act of doing the study may have biased the findings?

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). 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. 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.

Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 21 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

30 | Does the report adequately address missing data/non-response?

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 qualitative studies. For quantitative evaluations, the report may also mention using post-stratification to adjust weights for non-response.

"Partial" score could be given if information about valid responses is provided to some, but not all data used in the findings. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 28 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

Quality Assessment Tool Survey

6. Findings, Conclusions, Recommendations or Discussion Sections

The questions about the quality of the study found in this section address items that are typically found in the findings or discussion section of a report. Since not every report follows the same structure, you may need to search for the relevant items elsewhere in the report of the study you are reviewing

31 If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates?

USAID recommends that the margin of error be reported along with the findings from statistical samples.

"Partial" score could be given if some but not all of the statistical data are presented to include standard errors and confidence intervals, or if the statistical data is presented with either standard errors or confidence intervals. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given if the study does not use inferential statistical methods.

This question corresponds to item 24 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

32 Does the analysis include triangulation of data from different sources?

Typically, stronger bodies of evidence are likely to emerge if similar findings are obtained from different types of data (e.g., tests, interviews, observations) and respondent types (e.g., students, parents, teachers). It is important that contradictory data be taken into account when discussing the findings.

"Partial" score could be given if data from different sources are presented but the findings don't connect them into a coherent narrative. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given if the study does not use multiple data sources.

This question corresponds to item 17 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

33 Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings?

The purpose of a report is to provide the evaluators' or researchers' findings and recommendations on each and every study question. Accordingly, 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.

"Partial" score could be given if the answers are provided in the report but not the Executive Summary, or if some but not all questions are answered. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 29 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

34 Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections?

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.

For qualitative studies, 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)

"Partial" score could be given if the report discusses alternative explanations or inconsistencies in the findings. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 22 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

35 Does the report address the external validity (for quantitative studies) or transferability (for qualitative studies) of findings?

Quantitative findings 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. **Qualitative findings** 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.

"Partial" score could be given if the external validity or transferability of some, but not all key findings, are discussed in the report. Please explain the reason for a partial score in the comment box below (optional).

"NA" score could be given in case this study did not intend to have data from a sample extrapolated to a population.

This question corresponds to item 23 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

36 For impact evaluations, is a cost analysis of the intervention being evaluated included in the report?

Reporting on the findings from a cost analysis should be clear on all elements that may be useful for making decisions.

USAID requires all impact evaluations to include a cost analysis of the 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:
Details about the intervention, such as the ToC, 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
Details about the beneficiaries: who and where they are, including geography, age and sex, marginalization status, and other relevant details
Cost estimates: what is included/not included and why, whether recurrent and non-recurrent costs are separate, whether contributions were costed out or listed alongside final estimates
Perspectives: whose perspective(s) are reflected in the cost estimates
Cost modeling: prospective or respective, assumptions, and data limitations
Major cost drivers: what they are, what factors have greatest influence on cost estimates (i.e., contextual factors, beneficiary characteristics, intervention features, etc.)
Computation: how cost estimates were computed, including assumptions and the computations in an annex

"Partial" score could be given if some but not all of the required cost analysis details are included in the report. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given for studies that do not include an impact evaluation.

This question corresponds to item 10 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

37 Does the report list steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders and incorporate stakeholder feedback in the report?

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 help 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).

Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 13 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

38 Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings?

Visuals must be used to facilitate understanding of the findings by general 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.

"Partial score" could be given if the report uses visuals to an insufficient extent. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 33 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

39 If recommendations are made, are they specific, relevant, actionable, and based on the findings?

It is important that recommendations be practical, action-oriented, and specific as well as relevant to the study.

"Partial" score could be given if some but not all recommendations meet the criteria or if some but not all of the criteria is met for all recommendations. Please explain the reason for a partial score in the comment box below (optional).

"NA" score should be given to studies which are not intended to produce recommendations.

This question corresponds to item 31 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)
- ○ N/A (**survey-design:placeholder-points-na**)

Comment

Quality Assessment Tool Survey

7. Overall

Please answer questions about the clarity and presentation of the entire report

40 Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations?

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.

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.

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.

"Partial" score could be given if there is a clear, logical connection between some but not all of the study elements. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 32 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

41 Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)?

Reports should be written in an accessible way to non-experts. Excessive use of research terminology is also undesirable; the report should favor terminology that its intended audience is expected to be familiar with.

"Partial" score could be given if part but not all of the report is written in an accessible way. Please explain the reason for a partial score in the comment box below (optional).

This question corresponds to item 30 in the "Study Quality Assessment Tool".

- ○ Yes (**2 pts**)
- ○ No
- ○ Partial (**1 pts**)

Comment

Quality Assessment Tool Survey

8. Feedback about the tool

Please share your feedback about this tool, taking into consideration how easy or intuitive it was for you to use. Your feedback will help us improve the tool for use by a wider audience

42 Prior to completing the review, did you do any of the following orientation activities?

	Yes	No
Use the E-Orientation Module	○	○
Watch the Video Walk Through of Online Survey	○	○
Practice using the Online Survey	○	○
Attend a Listening Session	○	○

43 On a scale of 1-5, how easy was it to use the tool?

Please explain your response in the comment box below.

(with 1 being easiest, 5 being hardest)

Comment

44 On a scale of 1-5, how relevant do you find the tool to research and evaluation?

Please explain your response in the comment box below.

(with 1 being the least relevant, 5 being completely relevant)

Comment

45 Please provide specific feedback about the tool

You can provide feedback on the content, the wording, the structure, or anything that you believe is relevant

Enter answer

46 On a scale of 1-5, how likely are you to use this tool in your work?

Please explain response in the comment box below.

(with 1 being highly unlikely, 5 being extremely likely)

Comment

47 On a scale of 1-5, how likely are you to recommend the tool to your colleagues?

Please explain your response in the comment box below.

(With 1 being highly unlikely, 5 being extremely likely)

Comment

48 What should we name the tool?

While "Study Quality Assessment Tool" is an accurate title, it is a little awkward. We would love to know what you would call this tool! If we choose one of your ideas we will give you credit for the name when the tool is published. Please share your name and email in the comment box to allow us to assign accurate credit.

Enter answer

Title idea 1

Enter answer

Title idea 2

Comment

ANNEX C: LIST OF DOCUMENTS

The documents included in the review were comprised of evaluation reports, research reports, special evaluations, other USAID supported studies and documents, evaluation reports, and assessment studies about education. Briefs, infographics, planning documents, scopes of work, and evaluation design reports were not included in the study. Reports about studies that did not collect primary data, such as desk reviews, were also removed, to match the scope of the tool.

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Accelerated Quality Education for Children in Liberia (AQE) Midterm Performance Evaluation	2019	Africa	Yes	79	Performance Evaluation	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00W5CV.pdf	Y
Preschool Education in Morocco: Challenges and Key Potential Inputs	2022	MENA	No	34	Other study	Qualitative Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZHCD.pdf	Y
Rapid Education and Risk Analysis Colombia	2020	LAC	Yes	150	RERA	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00X363.pdf	Y
End-line Evaluation in Nepal	2022	Asia	Yes	349	Impact and performance evaluation	Mixed Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZJPW.pdf	Y
Baseline Evaluation of the Second Phase (2021–2026) in Cote d'Ivoire	2022	Africa	No	201	Other study	Mixed Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZJZ3.pdf	Y
Final Evaluation of the First Phase (2015–2021) in Côte d'Ivoire	2022	Africa	No	224	Performance Evaluation	Mixed Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZJZ2.pdf	Y
Enhancing Quality in Pre-Primary Education in Lebanon in Times of Crisis: Final Report	2022	MENA	Yes	91	Other study	Qualitative Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZHDN.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Early Childhood Education Research Study 2020 Baseline Report	2020	Africa	No	199	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X8HZ.pdf	Y
Itegure Gusoma: Get Ready to Read. Early Childhood Development Programme Rwanda Baseline Report	2018	Africa	No	88	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X3Z7.pdf	Y
Zambia Early Childhood Education Research Study: 2022 Endline Report	2022	Africa	No	163	Assessment	Quantitative Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZN5T.pdf	N
LCP Cities' System Capacity Development (CI-CAP) Project: Challenges and Opportunities in Financing the Education Requirements of Early Grade Learners	2020	Asia	Yes	70	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZDKP.pdf	Y
USAID Education Data Activity: Language Mapping Exercise Report	2019	Africa	No	28	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00THW7.pdf	Y
A Study on Remote Radio Lessons to Support Early Grade Kinyarwanda Learning in Rwanda	2020	Africa	No	78	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X4G5.pdf	N
Educating Children Together Phase 2 (ECT2) Final Evaluation Report	2022	Africa	No	137	Impact Evaluation	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00Z94F.pdf	Y
Evaluation of the FY 2018 Promoting Autonomy for Literacy and Attentiveness Through Market Alliances (PALAM/A) Project - Baseline Report	2022	Asia	No	183	Other study	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZJD2.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Food for education project phase 2: midterm evaluation report	2019	Africa	No	344	Performance Evaluation	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00XF2T.pdf	Y
La Paz, Honduras Baseline Evaluation – La Paz Expansion	2022	LAC	Yes	112	Other study	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZHGV.pdf	Y
Mozambique Program Impact Evaluation	2022	Africa	No	180	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZKH1.pdf	N
Philippines Innovation Ecosystem Assessment, 2019 Update	2020	Asia	Yes	66	Assessment	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00ZGD1.pdf	N
Bar ama Baro - "Teach or Learn": Somalia's Accelerated Quality Learning Program Baseline Evaluation	2022	Africa	Yes	589	Other study	Quantitative Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZJF2.pdf	Y
Research Report on Publishing Collaboratives	2018	Africa	NA	79	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XSKC.pdf	N
Report: Cost of Teaching and Learning Materials/Data and Evidence for Education Programs	2021	Global	No	33	Other study	Quantitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00ZGZ3.pdf	Y
Pedagogical Management Model Based on Proyecto Educaccion's Experience	2018	LAC	Yes	61	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00T2W6.pdf	N
Rapid Education and Risk Analysis- Dominican Republic	2019	LAC	No	88	RERA	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00WFWZ.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Preliminary Report on COVID-19 Research: Data Collection and Analysis for the Early Grade Reading Study (EGRS), the Reading Support Project (RSP) and the Language Benchmarking Study	2021	Africa	No	97	Impact Evaluation	Quantitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XGST.pdf	N
Morocco e-Takwine Learning Management System (LMS) Study	2022	MENA	No	96	Other study	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZG4Q.pdf	Y
Public Financing of Education in Haiti, 2010 - 2018: Independent Report	2018	LAC	Yes	82	Other study	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TQSG.pdf	Y
Syria Essential Services II Northeast Syria Education Sector Assessment	2019	MENA	Yes	137	Assessment	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00W4VZ.pdf	N
Brief Assessment of Basic Education in Bosnia and Herzegovina: The Follow-On. Final Report	2018	E&E	No	86	Assessment	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TP9S.pdf	N
Education Performance, Improvement, Communication, and Knowledge (EPIC): Evaluation Report	2021	Global	No	129	Performance Evaluation	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XBNF.pdf	N
All Children Reading Asia: USAID/Burma Education and Youth Sector Assessment	2021	Asia	Yes	98	Assessment	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00X7QP.pdf	N
Northern Education Initiative Plus (NEI plus) Midline Institutional Capacity Assessment (ICA) of Bauchi State Education Agencies and	2018	Africa	Yes	104	Assessment	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TH16.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Local Government Education Authorities (LGEAs)									
Nicaragua Rapid Education and Risk Assessment	2018	LAC	Yes	65	RERA	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TQSF.pdf	N
South Sudan Rapid Education and Risk Analysis Report	2018	Africa	Yes	91	RERA	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PBAAJ830.pdf	Y
Final Assessment for Environment Education Program in Birds Head Seascape	2021	Asia	No	28	Assessment	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XXGV.pdf	N
Gender Analysis of the Government to Government Component of the Sindh Basic Education Programme	2018	Asia	Yes	161	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TVPG.pdf	N
Education Data Language Mapping Exercise Update	2022	Africa	No	30	Other Study	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00Z7TH.pdf	Y
Nonstate Schooling in the Middle East & North Africa	2021	MENA	NA	131	Other study	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XC5S.pdf	N
ABC+ Advancing Basic Education in the Philippines: Political Economy of Basic Education Provisioning in Region 6 (Western Visayas). Final Narrative Report	2020	Asia	Yes	59	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XJ6G.pdf	N
Leveraging Low-Cost Private Schools in Northern Ghana: Exploring Private Sector Partnerships to Support Education for All	2021	Africa	No	114	Assessment	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00Z32N.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Rapid Education and Risk Analysis Cox's Bazar - Final Report	2018	Asia	Yes	77	RERA	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TJTG.pdf	Y
Enhancing School Management and Planning Project General Assessment	2019	MENA	Yes	127	Assessment	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00WJP8.pdf	Y
Baseline Study Nepal FY20, Nepal (2020-2024)	2022	Asia	Yes	296	Other study	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZF7D.pdf	Y
Baseline Study in Bangladesh	2022	Asia	Yes	190	Other study	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZGK8.pdf	Y
Research and Learning of School Meals Program in Africa	2022	Africa	Yes	210	Other study	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZG4Z.pdf	Y
Kyrgyz Republic 2017 – 2021 Project - Final Evaluation	2022	Asia	No	167	Performance Evaluation	Mixed Methods	Other	https://pdf.usaid.gov/pdf_docs/PA00ZB67.pdf	N
Integration of Social and Emotional Learning into Basic Education Programming: Findings from Eight Case Studies	2021	Global	NA	69	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XXD9.pdf	Y
Education Systems Strengthening Research in sub-Saharan Africa: Final Report	2018	Africa	NA	52	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TCW5.pdf	N
Teacher Mobility Study	2018	Africa	Yes	53	Other study	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00X14S.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Policy Review of Minimum Professional Standards for Learning Facilitators of Nonformal Accelerated Education Programs in Nigeria: Implications for Nonformal Education Programs in Northeast Nigeria and USAID-AENN Project	2021	Africa	Yes	33	Other study	Qualitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XM65.pdf	Y
Teachers Training on Gender Mainstreaming Within Learning and School Environments: Assessment of Developing a Professional Development Course for Teachers on Gender with Jordan's Ministry of Education	2021	MENA	Yes	78	Assessment	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00XM33.pdf	Y
ASPIRE ELP Final Report	2018	Africa	No	35	Impact evaluation	Quantitative Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TFG5.pdf	Y
Teacher Rationalization, Retention, and Language Study: National Situation Analysis	2018	Africa	No	97	Other study	Mixed Methods	Other	http://pdf.usaid.gov/pdf_docs/PA00TQB8.pdf	N
Midterm Performance Evaluation of the Workforce Improvement and Skills Enhancement Activity (WISE); Technical Assistance, Training Activities and Capacity Building	2018	MENA	No	331	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00T6M7.pdf	Y
LAC Support Contract Jamaica Advance Performance Evaluation	2022	LAC	No	94	Performance Evaluation	Mixed Methods	YWFD	https://pdf.usaid.gov/pdf_docs/PA00ZHRR.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Technical Vocational Education and Training Strengthening for At-Risk Youth (TVET SAY): Public Opinion Survey on Technical Education in the Southern Caribbean Municipalities of Bluefields, Laguna de Perlas, Nueva Guinea, and La Desembocadura del Rio Grande Southern Caribbean Coast Autonomous Region (RACCS)	2020	LAC	Yes	106	Survey	Quantitative Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00X614.pdf	N
Engaging Young Agripreneurs: Options to Include Youth in Private Sector Extension and Advisory Services in Rwanda and Uganda	2020	Africa	NA	66	Other study	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00XR1Z.pdf	N
Mindanao Youth for Development (MYDev) Program FY17 Impact Evaluation Report & FY18/19 Extension Performance Evaluation Report: Measuring Youth's Employment, Perceptions and Engagements, and Skills	2019	Asia	Yes	33	Impact Evaluation	Experimental/Quasi-Experimental	YWFD	http://pdf.usaid.gov/pdf_docs/PA00W5Q2.pdf	N
Monitoring and Evaluation Support for Collaborative Learning and Adapting Activity: Performance Evaluation of Generating Entrepreneurs and Sustainable Synergies (GENESIS)	2020	LAC	Yes	227	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WK94.pdf	N
Midterm Performance Evaluation of USAID Career	2018	MENA	No	164	Performance Evaluation	Qualitative Methods	YWFD	http://pdf.usaid.gov/pdf_	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Center Activity: Final Evaluation Report								docs/PA00T2SQ.pdf	
Evaluation of the Ethiopia Youth Potential Activity: Youth Cohort Study	2019	Africa	Yes	171	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00TJMR.pdf	Y
Building the Potential of Youth Activity Youth Cohort Study Midline Report	2018	Africa	Yes	143	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00TDXB.pdf	Y
Final Performance Evaluation of USAID/Ethiopia's Building the Potential of Youth Activity	2020	Africa	Yes	106	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WJD2.pdf	Y
Generation Kenya Proof of Concept Study	2019	Africa	No	169	Other study	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WQ3N.pdf	Y
USAID/DRC Integrated Youth Development Activity (IYDA) Rapid Education Risk Assessment & Do No Harm Conflict Sensitivity Analysis (RERA / DNH)	2019	Africa	Yes	107	RERA	Qualitative Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WBMN.pdf	Y
Livelihoods for Resilience Activity: Labor Market Assessment	2018	Africa	Yes	59	Assessment	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00TCZ9.pdf	Y
Final Evaluation Report: Evaluation of the USAID Kunci Workforce Development Initiative	2020	Asia	No	94	Performance Evaluation	Qualitative Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00X529.pdf	Y
Huguka Dukore Akazi Kanoze Performance Evaluation	2019	Africa	No	86	Performance evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WGVG.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
DRC Benchmarking report	2022	Africa	Yes	87	Other study	Experimental/Quasi-Experimental	YWFD	https://pdf.usaid.gov/pdf_docs/PA00ZQ3R.pdf	Y
Sri Lanka YouLead Performance Evaluation	2022	Asia	No	83	Performance Evaluation	Mixed Methods	YWFD	https://pdf.usaid.gov/pdf_docs/PA00ZJKF.pdf	Y
Punjab Youth Workforce Development (PYWD) Project: Tracer Study Report	2019	Asia	Yes	52	Other study	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WCKV.pdf	N
Youth Leadership for Agriculture (YLA): End-of-Activity Evaluation	2020	Africa	Yes	130	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00WQWR.pdf	N
USAID Monitoring, Evaluation, and Learning Activity Training for Employment Activity (TEA): End of Project Evaluation Report	2021	MENA	Yes	81	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00Z3GP.pdf	N
Zimbabwe: Works Impact Evaluation Report	2018	Africa	No	60	Impact Evaluation	Experimental/Quasi-Experimental	YWFD	http://pdf.usaid.gov/pdf_docs/PA00SXBH.pdf	N
Final Performance Evaluation: Bridges to Employment in El Salvador	2020	LAC	Yes	194	Performance Evaluation	Mixed Methods	YWFD	http://pdf.usaid.gov/pdf_docs/PA00X2P5.pdf	N
Morocco Higher Education Situational Analysis Report	2019	MENA	No	135	Other study	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00WDRR.pdf	Y
Vietnam Tertiary Education Assessment	2019	Asia	Yes	83	Assessment	Qualitative Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00TSMD.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
USAID/Kosovo Transformational Leadership Impact and Performance Evaluation Project TLP University Partnerships Program: Performance Evaluation	2020	E&E	No	92	Performance evaluation	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XQPJ.pdf	Y
Impact-Med Activity Review: Competency-Based Medical Education (CBME) Advancement and Sustainability	2022	Asia	Yes	87	Other study	Qualitative Methods	Higher Education	https://pdf.usaid.gov/pdf_docs/PA00ZKDZ.pdf	Y
Mid-term Performance Evaluation: University Scholarship Program 7 (USP 7)/ Higher Education Scholarship (HES) Program. Final Evaluation Report	2021	MENA	Yes	61	Performance Evaluation	Qualitative Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XS8B.pdf	Y
Assessment of World Learning's Malawi Scholarship Program (MSP) Report	2019	Africa	No	99	Performance Evaluation	Mixed Methods	Higher Education	ASSESSMENT OF WORLD LEARNING'S MALAWI SCHOLARSHIP PROGRAM (MSP) REPORT (edu-links.org)	Y
Research, Evidence, and the Global Innovation Ecosystem: A Performance Evaluation of the Use and Utility of the Higher Education Solutions Network to Solve Development Challenges	2021	Global	No	410	Performance Evaluation	Mixed Methods	Higher Education	https://pdf.usaid.gov/pdf_docs/PA00XG65.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Sustainable Higher Education Research Alliances (SHREA): Mid-term Sustainability Assessment Report	2019	Asia	No	144	Assessment	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00WC2B.pdf	Y
Sustainable Higher Education Research Alliances (SHERA): Baseline Study Report	2018	Asia	No	49	Other study	Qualitative Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00T19P.pdf	Y
Science, Technology, Research and Innovation for Development (STRIDE): Performance Evaluation	2021	Asia	Yes	713	Performance Evaluation	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XWRX.pdf	Y
BUILD-IT Partnerships Sustainability Review	2021	Asia	Yes	71	Assessment	Qualitative Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XVRR.pdf	Y
Final Performance Evaluation of the Higher Education for Economic Growth Activity	2018	LAC	Yes	109	Performance evaluation	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00T881.pdf	Y
U.S.-Pakistan Centers For Advanced Studies: Midterm Evaluation	2018	Asia	Yes	226	Other study	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00TN5C.pdf	Y
Feasibility Study of the US-Pakistan Knowledge Corridor Scholarship Program	2018	Asia	Yes	91	Other study	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PBAAH928.pdf	N
Sustainable Higher Education Research Alliances (SHERA) Final Evaluation	2020	Asia	No	67	Performance Evaluation	Quantitative Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00X19J.pdf	N
University Led Innovation in Uganda: ResilientAfrica Network (RAN)	2020	Africa	Yes	82	Other study	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XC4B.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Situational Analysis of Higher Secondary Education in Bangladesh	2021	Asia	Yes	269	Other study	Mixed Methods	Higher Education	http://pdf.usaid.gov/pdf_docs/PA00XX8G.pdf	N
Scholarships and Training for Egyptian Professionals Activity (STEP): End-of-Project Performance Evaluation	2019	MENA	No	216	Performance evaluation	Mixed Methods	Higher Education	https://pdf.usaid.gov/pdf_docs/PA00TZBC.pdf	N
USAID/Mali Education Emergency Support Activity (EESA): Final Evaluation	2020	Africa	Yes	98	Performance Evaluation	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00X2KJ.pdf	Y
Teacher Motivation and Incentives Study Phase III	2019	Africa	Yes	64	Other study	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00WHTW.pdf	Y
Study of Incidence of Disability among Early Grade Learners in Senegal: Qualitative Research and Review of Existing Data	2020	Africa	Yes	54	Other study	Qualitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00X4X5.pdf	Y
Northern Education Initiative Plus (NEI+): End Line Performance Evaluation Report	2021	Africa	Yes	149	Performance Evaluation	Qualitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00XB17.pdf	Y
USAID India Final Gender and Social Inclusion in Education Report	2019	Asia	No	40	Other study	Qualitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00ZBBP.pdf	Y
The Malawi Girls' Empowerment through Education and Health Activity (ASPIRE): 2017 Performance Evaluation Report	2018	Africa	No	181	Performance evaluation	Mixed Methods	Access for Marginalized Groups	USAID ASPIRE 2017 Performance Evaluation	Y
Literacy Landscape Assessment in the Democratic Republic of Congo: Assessment Report	2020	Africa	Yes	109	Assessment	Qualitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00X79F.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
STOP Girl Trafficking Program: Final Evaluation	2019	Asia	Yes	97	Performance Evaluation	Qualitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00W45B.pdf	Y
Mid-term Evaluation of USAID/Mali Girls Leadership and Empowerment through Education (GLEE)	2021	Africa	Yes	123	Performance Evaluation	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00XR8J.pdf	Y
STOP Girl Trafficking Program: Endline Assessment	2019	Asia	Yes	98	Assessment	Quantitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00W459.pdf	Y
Mali Girls Leadership and Empowerment through Education: Baseline Report	2019	Africa	Yes	193	Other study	Quantitative Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00WB7B.pdf	Y
MEER Disability Inclusive Education Study Final Report	2022	MENA	Yes	89	Other Study	Qualitative Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZH9W.pdf	Y
Impact Evaluation of Nepal's Business Literacy Program	2020	Asia	Yes	171	Impact Evaluation	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00WS3N.pdf	Y
Assessment of Low-Cost Private Schools in FtF/RING II Districts in Northern Ghana	2019	Africa	No	99	Assessment	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00X417.pdf	Y
Apatseni Mwayi Atsikana Aphunzire AMAA Avaluation and Research: Baseline Report for School Construction Activity in Machinga and Balaka Districts Part I	2018	Africa	No	155	Other study	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_docs/PA00TK77.pdf	N
Performance Evaluation: Inclusive Education and Sports Program	2018	LAC	Yes	201	Performance Evaluation	Mixed Methods	Access for Marginalized Groups	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
								docs/PA00W6ZC.pdf	
USAID/Malawi Apatseni Mwayi Atsikana Aphunzire (AMAA) Evaluation and Research: Performance Evaluation	2022	Africa	No	108	Performance Evaluation	Qualitative Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZDBQ.pdf	N
Evaluation and Research Final Report: Apatseni Mwayi Atsikana Aphunzire Amaa	2022	Africa	No	84	Impact and performance evaluation	Mixed Methods	Access for Marginalized Groups	https://pdf.usaid.gov/pdf_docs/PA00ZJQ2.pdf	N
Niger Education Community Strengthening (NECS) Program: Final Performance Evaluation	2018	Africa	Yes	218	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TFQN.pdf	Y
Uganda Performance and Impact Evaluation for Literacy Achievement and Retention Activity (LARA): Midterm Impact and Final Performance Evaluation Report	2020	Africa	Yes	89	Impact and performance evaluation	Mixed Methods/Experimental/ Quasi-Experimental	Foundational Skills	https://dec.usaid.gov/dec/GetDoc.axd?ctID=ODVhZjk4NWQtM2YyMi00YjRmLTkxNjktZTcxMjM2NDBmY2Uy&rID=NTgwMzA3&pID=NTYw&attchmnt=VHJIZQ==&uSesDM=False&rIdx=MzA0Nzkx	Y
CARE India-Endline Report-Start Early Read in Time	2018	Asia	No	100	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W7ND.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Ghana Transition-to-English Plus (T2E+) Impact Evaluation: Baseline Report	2022	Africa	No	80	Other study	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZDMD.pdf	N
Monitoring, Evaluation and Coordination Contract (MECC): ACCELERE! Activity 1 Reading Impact Evaluation Report	2020	Africa	Yes	223	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XCRN.pdf	N
USAID/Uganda Final Performance and Impact Evaluation for Literacy Achievement and Retention Activity (LARA)	2021	Africa	Yes	88	Impact Evaluation	Mixed Methods/Experimental/Quasi-Experimental	Foundational Skills	https://dec.usaid.gov/dec/GetDoc.axd?ctID=ODVhZjk4NWQtM2YyMi00YjRmLTkxNjktZTcxMjM2NDBmY2Uy&rID=NTg0MDQw&pID=NTYw&attchmnt=VHJ1ZQ==&uSesDM=False&rIdx=MzA4NDk3	N
USAID Tusome Pamoja Pre-primary Endline Assessment Report	2019	Africa	No	67	Assessment	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X8DB.pdf	Y
Sindh Reading Program Early Grade Reading Assessment Endline Report	2019	Asia	Yes	219	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W5GG.pdf	Y
The Cost-Effectiveness of Classroom Based Libraries on Students Reading Skills	2020	Asia	Yes	44	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
								docs/PA00WRHT.pdf	
Midterm Performance Evaluation of the USAID/Laos Learn to Read Activity	2021	Asia	No	134	Performance evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZF4F.pdf	Y
Alternative Basic Education in Somalia External Performance Endline Evaluation	2021	Africa	Yes	175	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X9T5.pdf	Y
EGRA Midline Report	2022	Africa	No	117	Assessment	Quantitative Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZJPQ.pdf	Y
USAID/Liberia Read Liberia Impact Evaluation Classroom Practices Report 2019	2019	Africa	Yes	52	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z3FP.pdf	Y
Ghana Early Grade Reading Program Impact Evaluation - 2018 Midline Report	2019	Africa	No	284	Impact Evaluation	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00XHTV.pdf	Y
Improving Reading in Djibouti: Midterm Performance Evaluation of the Djibouti Early Grade Reading Activity (DEGRA)	2021	Africa	No	154	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z8H1.pdf	Y
Evaluation of Complementary Reading Project (CRP) Grants Initiatives	2018	Asia	Yes	90	Other study	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQFV.pdf	Y
Lecture Pour Tous Study of the Knowledge, Attitudes, and Practices (KAP) of the Ministry of National Education - Midline Report	2020	Africa	Yes	60	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X33B.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Nepal Early Grade Reading Program Performance Evaluation 2019: Final Report	2020	Asia	Yes	181	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WGR4.pdf	Y
Pashto Pilot Midline Report	2019	Asia	Yes	25	Other study	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQKX.pdf	Y
2019 Regional Early Grade Reading Assessment (USAID ACR Asia Philippines)	2020	Asia	Yes	84	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WKGV.pdf	Y
Vamos Ler!: Deep Learning Adaptive Study 2 (DLAS 2)	2019	Africa	No	108	Other study	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WGV8.pdf	Y
USAID Jordan RAMP: Early Grade Reading and Mathematics Initiative Midline Survey Summary Report	2018	MENA	Yes	29	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W7WZ.pdf	Y
Naogaon: USAID's Reading Enhancement for Advancing Development (READ) Activity	2018	Asia	Yes	60	EGRA Report	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TBKF.pdf	Y
Ghana Numeracy Pilot Impact Evaluation: 2017 Baseline Report	2018	Africa	No	200	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00SWTG.pdf	Y
Read Liberia Activity: 2021 District Education Monitoring Approach (DEMA) Group Administered Literacy Assessment (GALA)	2021	Africa	Yes	52	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZBT8.pdf	Y
Early Grade Reading and Math Project (RAMP): Impact Evaluation Final Report	2019	MENA	Yes	602	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WKQ9.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Rapid Assessment: Effects of the Covid-19 Pandemic on Student Retention, Instruction and Learning, and Irregular Migration of Students and their Families in El Salvador. Final Report	2022	LAC	Yes	88	Assessment	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z98Q.pdf	Y
Final Evaluation Report: Haiti Early Reading Program (ERP) Baseline Evaluation	2018	LAC	Yes	158	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNP6.pdf	Y
Northern Education Initiative Plus: Early Grade Reading Assessment Midline Report	2018	Africa	Yes	197	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WGT7.pdf	Y
ABC+: Advancing Basic Education in the Philippines Baseline Report	2020	Asia	Yes	120	EGRA Report	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X45G.pdf	Y
Variation Study Endline Report	2020	Asia	Yes	159	Impact Evaluation	Mixed Methods/Experimental/ Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WRHX.pdf	Y
USAID's Early Grade Reading Program II (EGRP II) in Nepal-Baseline Report Vol. 2, COVID-19 Response: The Home- and Community-Based Schooling Intervention	2021	Asia	Yes	62	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZBHB.pdf	Y
2017 Early Grade Reading Assessment: Sindh - Sindh Reading Program	2018	Asia	Yes	160	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNJZ.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
USAID Honduras Reading Activity 2018 Associated Factors Study	2018	LAC	Yes	51	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XM8H.pdf	Y
Impact Evaluation of the Western Cape Emergent Literacy Intervention in South Africa	2018	Africa	No	238	Impact Evaluation	Mixed Methods/Experimental/ Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T61D.pdf	Y
Assessing the Functionality and Sustainability of Community Engagement Structures for Early Grade Reading: Final Report	2020	Asia	Yes	56	Assessment	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X2X1.pdf	Y
USAID Tusome Pamoja: Field Study on Gender and Learning	2020	Africa	No	51	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X252.pdf	Y
USAID-funded Pakistan Reading Project: PRP Baseline Variation Study	2019	Asia	Yes	35	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQM7.pdf	Y
FATA Baseline Report: Student and Teacher Assessment	2018	Asia	Yes	29	Assessment	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQWQ.pdf	Y
Midterm Performance Evaluation: the Latin America and Caribbean Reads Capacity Program	2019	LAC	NA	98	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W5R8.pdf	Y
Data-Driven Instruction in Honduras: An Impact Evaluation of the Educaccion-PRI Promising Reading Intervention	2019	LAC	Yes	199	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WDW1.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
USAID Impact Evaluation of the Makhalidwe Athu Project (Zambia)	2018	Africa	No	139	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00SZJS.pdf	Y
Pakistan Reading Project (PRP) 2020 Early Grade Reading Endline Supplementary Research	2022	Asia	Yes	48	Other study	Qualitative Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZJJG.pdf	Y
Impact Evaluation of the Early Grade Reading Activity (EGRA) in Malawi	2018	Africa	No	165	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T3Q6.pdf	Y
Lecture Pour Tous MEN KAP Baseline Study Report	2018	Africa	Yes	49	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X339.pdf	Y
Sindhi End line Student and Teachers Assessment Report	2018	Asia	Yes	29	Assessment	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQKT.pdf	Y
USAID/Niger Education & Community Strengthening (NECS+) Early Grade Reading Assessment Monitoring Report	2018	Africa	Yes	74	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W573.pdf	Y
Reading Support Project: Final Design Evaluation Report	2019	Africa	NA	179	Other study	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WJHJ.pdf	Y
Rwanda Early Grade Reading Assessment: Baseline Report 2018 USAID Soma Umenye Project	2020	Africa	No	222	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X3C5.pdf	Y
Midline Evaluation in Timor-Leste	2022	Asia	No	295	Impact Evaluation	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00Z63K.pdf	Y

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Mid-term Performance Evaluation of the USAID READ Activity	2018	LAC	No	83	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00SZ86.pdf	Y
An Evaluation of Reading for Ethiopia's Achievement Developed Community Outreach (READ CO) Project	2019	Africa	Yes	96	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X599.pdf	Y
USAID/Philippines Basa Pilipinas Program. Reading is for Girls: A Study of the Role of Gender in Literacy Achievement in USAID Basa Pilipinas	2018	Asia	Yes	40	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T716.pdf	Y
Mozambique Educating Children Together Phase 3 (ECT3) - Baseline Evaluation	2022	Africa	No	57	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZB8R.pdf	N
USAID READ Community Outreach: Assessment on Sustainability	2019	Africa	Yes	44	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TRR8.pdf	N
Benin Keun Faaba III: Baseline Evaluation	2022	Africa	No	120	Assessment	Mixed Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZM68.pdf	N
Ethiopia READ Community Outreach: Internal Performance Assessment	2018	Africa	Yes	42	Assessment	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TRR5.pdf	N
Senegal All Children Reading Lecture Pour Tous: EGRA CI Analysis Report	2018	Africa	Yes	45	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X14P.pdf	N
All Children Reading Asia Analysis of Early Grade Reading Assessment in India	2018	Asia	No	119	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Initial Data Collection and Assessment Report								docs/PA00TG2F.pdf	
Mureke Dusome Performance Evaluation: Documenting Successful Approaches and Lessons Learned in Promoting Early Grade Reading through Sustainable School-Community Partnerships in Rwanda	2020	Africa	No	149	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WCGJ.pdf	N
2017 Early Grade Reading Assessment: Balochistan	2018	Asia	Yes	89	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK7.pdf	N
All Children Reading-Cambodia Student Performance in Early Literacy: Midterm Impact Report	2020	Asia	No	60	Impact Evaluation	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WKQM.pdf	N
Reading Support Project: Formative Implementation Evaluation of the Reading Support Project in South Africa	2020	Africa	No	578	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XH1V.pdf	N
Read Liberia Impact Evaluation Implementation Report 2019	2019	Africa	Yes	49	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z3FQ.pdf	N
Setting Reading Benchmarks in South Africa	2020	Africa	No	108	Other study	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X1NZ.pdf	N
Reading for Ethiopia's Achievement Developed Technical Assistance Project	2018	Africa	Yes	86	Performance evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00SZB8.pdf	N
USAID/Philippines Basa Pilipinas Program: Early Grade	2018	Asia	Yes	149	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Reading Assessment Final Evaluation Report 2018								docs/PA00T715.pdf	
Midterm Performance Evaluation: Soma Umenye Activity	2020	Africa	NA	127	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WNZX.pdf	N
2017 Early Grade Reading Assessment: Islamabad Capital Territory	2018	Asia	Yes	49	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK4.pdf	N
2017 Early Grade Reading Assessment: Sindh - Pakistan Reading Program	2018	Asia	Yes	153	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK1.pdf	N
2017 Early Grade Reading Assessment: Azad Jammu and Kashmir	2018	Asia	Yes	63	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK8.pdf	N
USAID Soma Umenye Local Early Grade Reading Assessment (LEGRA) Pilot Report	2020	Africa	No	48	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X74F.pdf	N
Variation Study Midline Quantitative & Qualitative Report	2020	Asia	Yes	86	Impact Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQM3.pdf	N
2017 Early Grade Reading Assessment: Gilgit-Baltistan	2018	Asia	Yes	94	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK3.pdf	N
Ghana Early Grade Reading Program Impact Evaluation: 2017 Baseline Report	2018	Africa	No	174	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00SWTF.pdf	N
Final Performance Evaluation of the USAID/Guatemala Lifelong Learning Project	2018	LAC	Yes	81	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
								docs/PA00TFPC.pdf	
Tusome Pamoja: Midline Findings Report	2018	Africa	No	37	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TZ35.pdf	N
Reading and Access Evaluation Report: Final Performance Evaluation of Amazonia Lee in Peru	2020	LAC	No	124	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WP1R.pdf	N
Performance Evaluation of Reading for Ethiopia's Achievement Developed Institutional Improvement (READ II)	2018	Africa	Yes	57	Performance Evaluation	Qualitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X5JR.pdf	N
Evaluation Report: Khagrachari	2018	Asia	Yes	52	Other study	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T7MS.pdf	N
Lecture Pour Tous Teacher Knowledge, Attitudes, and Practice Regarding Early Grade Reading: Baseline Study Report	2020	Africa	Yes	75	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X338.pdf	N
USAID Reading and Access Story Powered School Program Impact Evaluation: Endline Report	2019	Africa	No	79	Impact evaluation	Mixed Methods/Experimental/ Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z3KH.pdf	N
Whole-of-Project Performance Evaluation of the Reading for Success Project - Morocco	2020	MENA	No	166	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WK1N.pdf	N
Rwanda Endline Survey Tracking Literacy Knowledge,	2021	Africa	No	193	Survey	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Attitudes, and Practices at the School and Community Level								docs/PA00XXGX.pdf	
Leading Partnerships and Participation for Learning, Education, and Development (LPP-LED): Final Narrative Report	2020	Asia	Yes	54	Assessment	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00ZDKR.pdf	N
USAID/Liberia Read Liberia Impact Evaluation: Classroom Observation Report 2022	2022	Africa	Yes	37	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00ZQKF.pdf	N
2017 Early Grade Reading Assessment: Federally Administered Tribal Areas	2018	Asia	Yes	34	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK6.pdf	N
Results of Malawi National Reading Program Baseline Assessment of Standard 2 and 4 Learners	2018	Africa	No	137	Assessment	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T3Q5.pdf	N
A Study on Kinyarwanda Instructional Time in Lower Primary, Volume 1 of 2	2019	Africa	No	107	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XHVB.pdf	N
USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E) Early Grade Reading Assessment (EGRA): 2018 Endline Report	2019	Africa	Yes	121	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X5JW.pdf	N
READ Foundation End-line Assessment Study, Azad Jammu & Kashmir and Gilgit Baltistan	2019	Asia	Yes	35	Assessment	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQM2.pdf	N
Lecture Pour Tous Report on the Midline Survey of Knowledge, Attitudes, and Practices on Parental and	2020	Africa	Yes	68	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X33C.pdf	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
Community Engagement in Reading									
Reading Support Project: Summative Implementation Evaluation of the Reading Support Project in South Africa	2021	Africa	No	455	Performance Evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XGSV.pdf	N
Rwanda 2018 Early Grade Reading Baseline Assessment: Trends Observed, Lessons Learned and Recommendations for the Future	2019	Africa	No	43	EGRA Report	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X3BG.pdf	N
Equating Study: LARS IV, EGRA 2018, LEGRA 2021. USAID Soma Umenye	2021	Africa	No	44	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00Z2QZ.pdf	N
READ Foundation Baseline Assessment Study, Azad Jammu & Kashmir and Gilgit Baltistan	2018	Asia	Yes	36	Assessment	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00WQM1.pdf	Y
Mid-term Performance Evaluation of the Selective Integrated Reading Activity (SIRA) in Mali	2020	Africa	Yes	193	Performance evaluation	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00X2K9.pdf	N
Endline Evaluation of the Innovation for Improving Early Grade Reading Activity (IIEGRA) in Government Primary Schools (GPS)	2018	Asia	Yes	49	Assessment	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PBAAJ746.pdf	N
2017 Early Grade Reading Assessment: Khyber Pakhtunkhwa	2018	Asia	Yes	90	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00TNK2.pdf	N
USAID Pakistan Reading Project Batch 2 Impact Evaluation Study	2020	Asia	Yes	33	Impact Evaluation	Experimental/Quasi-	Foundational Skills	http://pdf.usaid.gov/pdf_	N

Title	Publication Year	USAID Region	EiCC (FY21)	# Pages	Study Type	Method	Education Policy Priority	File	Review Completed
						Experimental		docs/PA00WQKV.pdf	
Evaluation of Leer Juntos, Aprender Juntos Early Grade Reading Intervention in Guatemala: Final Report	2019	LAC	Yes	166	Impact Evaluation	Experimental/Quasi-Experimental	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00XJKG.pdf	N
Niger Education & Community Strengthening Early Grade Reading Assessment Report	2019	Africa	Yes	111	EGRA Report	Quantitative Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00W574.pdf	N
National Assessment of Reading Instruction, Standards 1-4, February 2019	2019	Africa	No	190	Other study	Quantitative Methods	Foundational Skills	https://pdf.usaid.gov/pdf_docs/PA00WCP4.pdf	N
Gender-Sensitive Education Management to Improve Reading Acquisition	2018	LAC	Yes	61	Other study	Mixed Methods	Foundational Skills	http://pdf.usaid.gov/pdf_docs/PA00T2W8.pdf	Y

ANNEX D: REVIEWER CHARACTERISTICS

A total of 163 potential reviewers signed up to participate, and many volunteered to review two or more reports. Ninety-two (56.4 percent) volunteers completed their reviews. While most (n = 64, 69.6 percent) volunteers reviewed one report, some volunteers completed two or three reviews.

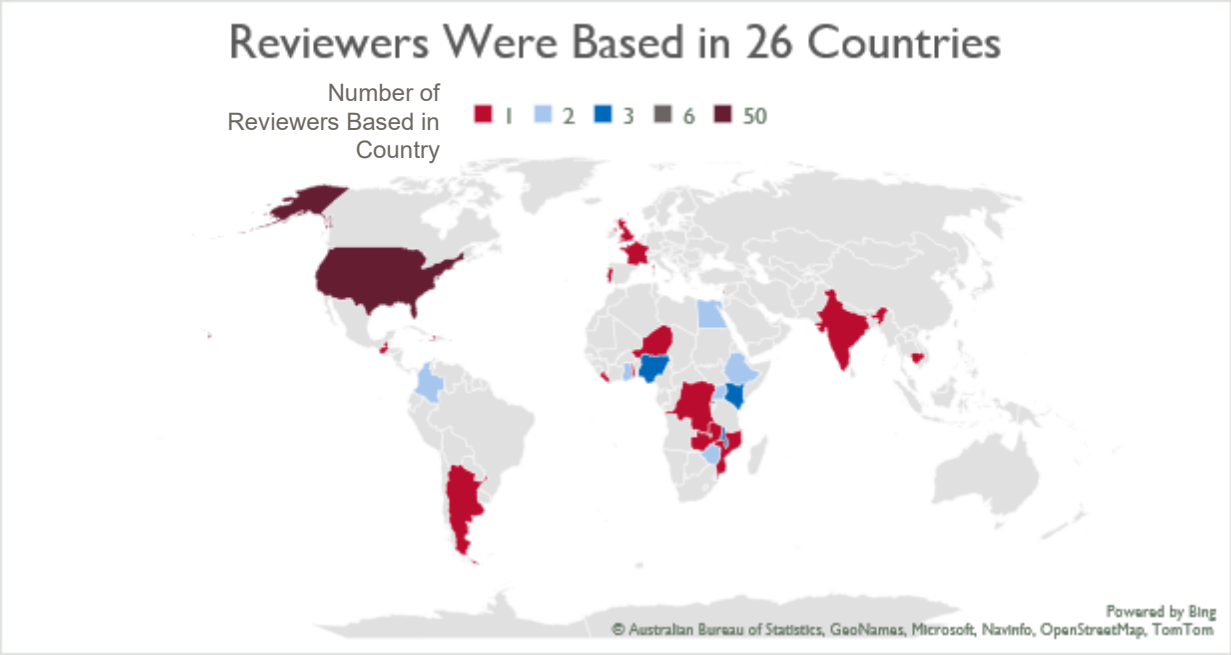
As Exhibit 30 shows, most volunteer reviewers were based in North America and Sub-Saharan Africa. Fewer reviewers based in Asia and Europe signed up or completed a review. This could indicate a gap in the solicitation approach, represent broader regional patterns of engagement with USAID education activities, or indicate other barriers that researchers and partners in those regions face.

Exhibit 30: Regional distribution of reviewers

Region	Signed Up (n = 163)	Completed Review (n = 92)
Sub-Saharan Africa	65 (40.0%)	29 (31.5%)
Asia	6 (3.7%)	2 (2.2%)
Europe	4 (2.5%)	3 (3.3%)
Latin America and the Caribbean	8 (5.0%)	5 (5.4%)
Middle East and North Africa	9 (5.5%)	3 (3.3%)
North America	71 (44.0%)	50 (54.0%)

Volunteers who signed up were based in 39 countries, while those who participated were based in 26 countries (Exhibit 31). While not everyone who expressed interest in the review was able to fully participate, the draft ASQ Tool was socialized across geographic boundaries and reached USAID partners outside the global north.

Exhibit 31: Countries represented by reviewers who completed the review



ANNEX E: STUDY QUALITY ASSESSMENT RESULTS

CONCEPTUAL FRAMING

Reports that met "Minimum Adequacy" for conceptual framing in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	39	57.4%
	Access to Education	20	16	80.0%
	Higher Education	14	11	78.6%
	YWFD	13	8	61.5%
	Other Priority	7	3	42.9%
Study Method	Experimental/Quasi	22	11	50.0%
	Quant Obs.	25	17	68.0%
	Qual Obs.	35	21	60.0%
	Mixed	40	28	70.0%
Region of Study	Global	4	3	75.0%
	Africa	57	37	64.9%
	Asia	37	17	45.9%
	EE	2	2	100.0%
	LAC	12	10	83.3%
	MENA	10	8	80.0%
Year Published	2018	34	21	61.8%
	2019	31	20	64.5%
	2020	18	10	55.6%
	2021	19	11	57.9%
	2022	20	15	75.0%
EiCC Status	EICC	74	47	63.5%
	Not EICC	48	30	62.5%
TOTAL		122	77	63.1%

Item-level data for conceptual framing

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Are clear study questions that are appropriate to the stated purpose of the study included in the report? (n=121)*	94	77.7%	13	10.7%	14	11.6%	0	1
If applicable, are study hypotheses included in the report? (n=58)	23	39.7%	24	41.4%	11	19.0%	64	0
Are the study questions appropriate to the conceptual/theoretical framework or theory of change? (n=120)	69	57.5%	22	18.3%	29	24.2%	0	2
Does the report acknowledge/draw upon existing relevant research? (n=120)	67	55.8%	19	15.8%	34	28.3%	0	2
Does the report explain the local context in sufficient detail as it	88	72.7%	17	14.0%	16	13.2%	0	1

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
relates to the study purpose and questions? (n=121)								

* *indicates the critically important question*

ROBUSTNESS OF METHODOLOGY

Reports that met "Minimum Adequacy" for robustness of methodology in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	27	39.7%
	Access to Education	20	6	30.0%
	Higher Education	14	9	64.3%
	YWFD	13	5	38.5%
	Other Priority	7	2	28.6%
Study Method	Experimental/Quasi	22	6	27.3%
	Quant Obs.	25	14	56.0%
	Qual Obs.	35	11	31.4%
	Mixed	40	18	45.0%
Region of Study	Global	4	2	50.0%
	Africa	57	24	42.1%
	Asia	37	11	29.7%
	EE	2	2	100.0%
	LAC	12	8	66.7%
	MENA	10	2	20.0%
Year Published	2018	34	15	44.1%
	2019	31	12	38.7%
	2020	18	5	27.8%
	2021	19	10	52.6%
	2022	20	7	35.0%
EiCC Status	EICC	74	29	39.2%
	Not EICC	48	20	41.7%
TOTAL		122	49	40.2%

Item-level data for robustness of methodology

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Is the methodology appropriate for answering posed study questions? (n=122)*	92	75.4%	3	2.5%	27	22.1%	0	0
Does the counterfactual meet standards of rigor? (n=63)*	44	69.8%	9	14.3%	10	15.9%	59	0
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	82	67.8%	9	7.4%	30	24.8%	0	1

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
target population(s), and presented in sufficient detail? (n=121)								
Does the report mention steps to mitigate common biases or threats to the integrity of the study? (n=122)	61	50.0%	48	39.3%	13	10.7%	0	0
Does the analysis include triangulation of data from different sources? (n=118)	80	67.8%	27	22.9%	11	9.3%	0	4

* *indicates the critically important questions*

CULTURAL APPROPRIATENESS

Reports that met "Minimum Adequacy" for cultural appropriateness in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	27	39.7%
	Access to Education	20	6	30.0%
	Higher Education	14	9	64.3%
	YWFD	13	5	38.5%
	Other Priority	7	2	28.6%
Study Method	Experimental/Quasi	22	6	27.3%
	Quant Obs.	25	14	56.0%
	Qual Obs.	35	11	31.4%
	Mixed	40	18	45.0%
Region of Study	Global	4	2	50.0%
	Africa	57	24	42.1%
	Asia	37	11	29.7%
	EE	2	2	100.0%
	LAC	12	8	66.7%
	MENA	10	2	20.0%
Year Published	2018	34	15	44.1%
	2019	31	12	38.7%
	2020	18	5	27.8%
	2021	19	10	52.6%
	2022	20	7	35.0%
EiCC Status	EICC	74	29	39.2%
	Not EICC	48	20	41.7%
TOTAL		122	49	40.2%

Item-level data for cultural appropriateness

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
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,	59	48.4%	36	29.5%	27	22.1%	0	0

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
and inclusive as appropriate? (n=122)*								
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? (n=122)	52	42.6%	31	25.4%	39	32.0%	0	0
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? (n=120)	62	51.7%	25	20.8%	33	27.5%	0	2
Does the report list steps taken to validate findings, conclusions, and recommendations (if applicable) with local stakeholders and incorporate stakeholder feedback in the report? (n=121)	47	38.8%	62	51.2%	12	9.9%	0	1

* *indicates the critically important question*

ETHICS

Reports that met "Minimum Adequacy" for ethics in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	18	26.5%
	Access to Education	20	9	45.0%
	Higher Education	14	8	57.1%
	YWFD	13	7	53.8%
	Other Priority	7	3	42.9%
Study Method	Experimental/Quasi	22	4	18.2%
	Quant Obs.	25	9	36.0%
	Qual Obs.	35	13	37.1%
	Mixed	40	19	47.5%
Region of Study	Global	4	1	25.0%
	Africa	57	22	38.6%
	Asia	37	11	29.7%
	EE	2	1	50.0%
	LAC	12	7	58.3%
	MENA	10	3	30.0%
Year Published	2018	34	12	35.3%

		N	# Adequate	% Adequate
	2019	31	9	29.0%
	2020	18	9	50.0%
	2021	19	8	42.1%
	2022	20	7	35.0%
EiCC Status	EICC	74	26	35.1%
	Not EICC	48	19	39.6%
TOTAL		122	45	36.9%

Item-level data for ethics

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Were ethical principles for the protection of human subjects integrated into the study approach and documented in the report? (n=122)*	56	45.9%	36	29.5%	30	24.6%	0	0
Was/were research clearance(s) appropriate to the study obtained prior to starting data collection, as documented in the report? (n=121)	39	32.2%	61	50.4%	21	17.4%	0	1

** indicates the critically important question*

VALIDITY

Reports that met "Minimum Adequacy" for validity in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	38	55.9%
	Access to Education	20	12	60.0%
	Higher Education	14	10	71.4%
	YWFD	13	7	53.8%
	Other Priority	7	1	14.3%
Study Method	Experimental/Quasi	22	14	63.6%
	Quant Obs.	25	14	56.0%
	Qual Obs.	35	17	48.6%
	Mixed	40	23	57.5%
Region of Study	Global	4	3	75.0%
	Africa	57	33	57.9%
	Asia	37	15	40.5%
	EE	2	2	100.0%
	LAC	12	9	75.0%
	MENA	10	6	60.0%
Year Published	2018	34	21	61.8%
	2019	31	18	58.1%
	2020	18	9	50.0%
	2021	19	12	63.2%
	2022	20	8	40.0%
EiCC Status	EICC	74	39	52.7%

		N	# Adequate	% Adequate
	Not EICC	48	29	60.4%
TOTAL		122	68	55.7%

Item-level data for validity

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Does the report explain in sufficient detail how the indicators or constructs used in the study capture the phenomenon being investigated? (n=120)	95	79.2%	15	12.5%	10	8.3%	0	2
Is the report open and clear about how the act of doing the study may have biased the findings? (n=121)	56	46.3%	54	44.6%	11	9.1%	0	1
If applicable to the study methods, are statistical data presented to include standard errors and confidence intervals around point estimates? (n=70)	30	42.9%	28	40.0%	12	17.1%	50	2
Does the report provide evidence that the findings are credible, such as through discussions of alternative interpretations in the findings and conclusions sections? (n=121)	73	60.3%	21	17.4%	27	22.3%	0	1
Does the report address the external validity (for quantitative studies) or transferability (for qualitative studies) of findings? (n=94)	62	66.0%	21	22.3%	11	11.7%	26	2

RELIABILITY

Reports that met "Minimum Adequacy" for reliability in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	26	38.2%
	Access to Education	20	11	55.0%
	Higher Education	14	7	50.0%
	YWFD	13	6	46.2%
	Other Priority	7	2	28.6%
Study Method	Experimental/Quasi	22	10	45.5%
	Quant Obs.	25	12	48.0%
	Qual Obs.	35	13	37.1%
	Mixed	40	17	42.5%
Region of Study	Global	4	1	25.0%
	Africa	57	25	43.9%
	Asia	37	14	37.8%
	EE	2	1	50.0%

		N	# Adequate	% Adequate
	LAC	12	6	50.0%
	MENA	10	5	50.0%
Year Published	2018	34	21	61.8%
	2019	31	12	38.7%
	2020	18	7	38.9%
	2021	19	7	36.8%
	2022	20	5	25.0%
EiCC Status	EICC	74	35	47.3%
	Not EICC	48	17	35.4%
TOTAL		122	52	42.6%

Item-level data for reliability

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Does the report document the steps taken to ensure that data were collected with a high degree of reliability? (n=122)	85	69.7%	23	18.9%	14	11.5%	0	0
If applicable, was internal consistency of the instrument(s) established and documented? (n=81)	44	54.3%	29	0.3580247	8	9.9%	40	1
For studies where data is collected by a team, was inter-rater reliability established and documented? (n=86)	34	39.5%	44	51.2%	8	9.3%	36	0
If applicable to the study methods, was inter-coder reliability established and documented for studies where data was coded by a team? (n=82)	31	37.8%	35	42.7%	16	19.5%	40	0
Does the report adequately address missing data/non-response? (n=120)	48	40.0%	48	40.0%	24	20.0%	0	2

OPENNESS AND TRANSPARENCY

Reports that met "Minimum Adequacy" for openness and transparency in 2023, by report factor

		N	# Adequate	% Adequate
Education Policy Priority	Basic Education	68	35	51.5%
	Access to Education	20	14	70.0%
	Higher Education	14	11	78.6%
	YWFD	13	9	69.2%
	Other Priority	7	5	71.4%
Study Method	Experimental/Quasi	22	9	40.9%
	Quant Obs.	25	16	64.0%

		N	# Adequate	% Adequate
	Qual Obs.	35	22	62.9%
	Mixed	40	27	67.5%
Region of Study	Global	4	1	25.0%
	Africa	57	36	63.2%
	Asia	37	20	54.1%
	EE	2	2	100.0%
	LAC	12	8	66.7%
	MENA	10	7	70.0%
Year Published	2018	34	20	58.8%
	2019	31	19	61.3%
	2020	18	11	61.1%
	2021	19	12	63.2%
	2022	20	12	60.0%
EiCC Status	EICC	74	41	55.4%
	Not EICC	48	33	68.8%
TOTAL		122	74	60.7%

Item-level data for openness and transparency

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Is the methodology explained in sufficient detail for a reader to understand the study design and the rationale for decisions made? (n=121)	94	77.7%	5	4.1%	22	18.2%	0	1
Is the report open and clear about limitations inherent to the study design and with its implementation? (n=122)*	97	79.5%	5	4.1%	20	16.4%	0	0
For evaluations, is the report open and clear about study limitations due to issues with the implementation of the intervention being evaluated? (n=99)	72	72.7%	15	15.2%	12	12.1%	22	1
Is the report open about potential biases due to the study team composition? (n=121)	55	45.5%	59	48.8%	7	5.8%	0	1
For impact evaluations, is a cost analysis of the intervention being evaluated included in the report? (n=55)	14	25.5%	31	56.4%	10	18.2%	67	0

* *indicates the critically important question*

COGENCY

Reports that met "Minimum Adequacy" for cogency in 2023, by report factor

		N	# Adequate	% Adequate
	Basic Education	68	56	82.4%

		N	# Adequate	% Adequate
Education Policy Priority	Access to Education	20	20	100.0%
	Higher Education	14	13	92.9%
	YWFD	13	11	84.6%
	Other Priority	7	5	71.4%
Study Method	Experimental/Quasi	22	17	77.3%
	Quant Obs.	25	24	96.0%
	Qual Obs.	35	28	80.0%
	Mixed	40	36	90.0%
Region of Study	Global	4	3	75.0%
	Africa	57	51	89.5%
	Asia	37	29	78.4%
	EE	2	2	100.0%
	LAC	12	11	91.7%
	MENA	10	9	90.0%
Year Published	2018	34	31	91.2%
	2019	31	22	71.0%
	2020	18	17	94.4%
	2021	19	17	89.5%
	2022	20	18	90.0%
EiCC Status	EICC	74	62	83.8%
	Not EICC	48	43	89.6%
TOTAL		122	105	86.1%

Item-level data for cogency

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Are all study questions and sub-questions answered in the report and in the Executive Summary with evidence from the findings? (n=120)	95	79.2%	8	6.7%	17	14.2%	0	2
Is the report supported by relevant visualizations (e.g., charts, maps, infographics) that help non-technical audiences easily understand the study findings? (n=119)	76	63.9%	16	13.4%	27	22.7%	0	3
If recommendations are made, are they specific, relevant, actionable, and based on the findings? (n=117)	89	76.1%	10	8.5%	18	15.4%	5	0
Is there a clear, logical connection between the study questions, conceptual framework, data, analysis, findings, conclusions, and recommendations? (n=122)*	91	74.6%	6	4.9%	25	20.5%	0	0

	Yes	% Yes	No	% No	Partial	% Partial	N/A	No Response
Is the report written in a style and language that the intended audience can understand (e.g., technical jargon is minimized and explained)? (n=122)	102	83.6%	4	3.3%	16	13.1%	0	0

** indicates the critically important question*

REVIEWER FEEDBACK

		1		2		3		4		5	
	Average	n	%	n	%	n	%	n	%	n	%
On a scale of 1-5, how easy was it to use the tool? (with 1 being easiest, 5 being hardest) (n=115)	2.504348	37	32.2%	28	24.3%	17	14.8%	21	18%	12	10%
On a scale of 1-5, how relevant do you find the tool to research and evaluation? (with 1 being the least relevant, 5 being completely relevant) (n=113)	4.345133	3	2.7%	3	2.7%	7	6.2%	39	35%	61	54%
On a scale of 1-5, how likely are you to use this tool in your work? (with 1 being highly unlikely, 5 being extremely likely) (n=109)	3.954128	5	4.6%	9	8.3%	22	20.2%	23	21%	50	46%
On a scale of 1-5, how likely are you to recommend the tool to your colleagues? (With 1 being highly unlikely, 5 being extremely likely) (n=109)	4.06422	4	3.7%	8	7.3%	17	15.6%	28	26%	52	48%

ENDNOTES

[i] Norman Blaikie, *Designing Social Research*, 2nd Edition (Malde, MA: Polity, 2009), 40, 70, 74.

[ii] Dahlia K. Remler and Gregg G. Van Ryzin, *Research Methods in Practice: Strategies for Description and Causation*, 2nd Edition (Los Angeles: Sage Publications, 2015), 5.

[iii] 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.

[iv] 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.
[v] 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
[vi] USAID. "Local Systems: A Framework for Supporting Sustained Development. A Mandatory Reference for ADS Chapter 220." July 28 2014. https://www.usaid.gov/sites/default/files/documents/1870/220maq.pdf
[vii] Kim, Daniel. "Introduction to Systems Thinking," The Systems Thinker. 1999, https://thesystemsthinker.com/introduction-to-systems-thinking/
[viii] Allen, Will. "Systems thinking," Learning for Sustainability. Accessed 28 October 2022, https://learningforsustainability.net/systems-thinking/
[ix] Coghlan, David, and Mary Brydon-Miller. "Systems Thinking." In *The SAGE Encyclopedia of Action Research*, edited by Coghlan, David, and Mary Brydon-Miller, 753-54. London: SAGE Publications Ltd, 2014. https://dx.doi.org/10.4135/9781446294406.n333.
[x] Kim, Daniel. "Introduction to Systems Thinking," The Systems Thinker. 1999, https://thesystemsthinker.com/introduction-to-systems-thinking/
[xi] 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/
[xii] Allen, Will. "Systems thinking," Learning for Sustainability. Accessed 28 October 2022, https://learningforsustainability.net/systems-thinking/