

Tip Sheet 3: Assessing the Soundness of Evaluations

Step 3 of this guide states that analysts assess the soundness (reliability and validity) of the evaluations they deem relevant to their fragmentation, overlap, and duplication reviews. Specifically, analysts should assess all selected evaluations against basic standards for research design, conduct, analysis, and reporting. This tip sheet outlines questions that analysts can consider when addressing the soundness of evaluations.

All program evaluations included in a fragmentation, overlap, and duplication review should be assessed against basic standards for research design, conduct, analysis, and reporting. Major weaknesses in these areas that affect the reliability or validity (soundness) of each evaluation's findings must be identified and considered in using the evaluation and placing confidence in the findings. Whether experiment, case study, survey, or content analysis, each evaluation should be questioned as to its reliability and validity. Questions such as the following will determine the overall usefulness of an individual evaluation:

- Are the evaluation's objectives stated? Were the objectives appropriate with respect to the developmental stage of the program (e.g., new program or mature program)?
- Is the evaluation design clear? Was the design appropriate given the evaluation's objectives? Was the indicated design in fact executed?
- Did the variables measured relate to and adequately translate to the evaluation's objectives?
- Are sampling procedures and the evaluation sample sufficiently described? Were they adequate?
- Are sampling procedures such that policymakers can generalize to other persons, settings, and times of interest to them?
- Is an analysis plan presented and is it appropriate?
- Were the selection and training of data collectors adequate?
- Were there procedures to ensure reliability across data collectors?
- Were there any inadequacies in data collection procedures?
- Were problems encountered during data collection that affect data quality?
- Are the statistical procedures well specified and appropriate to the task?
- Are the conclusions supported by the data and the analysis?
- Are evaluation limitations identified? What possibly confounds the interpretation of the evaluation's findings?

This list shows some of the issues that should be raised in reviewing program evaluations. The information derived by answering these questions should lead to an overall judgment of the usefulness of each evaluation. It does not mean, however, that evaluations with design or other weaknesses are automatically excluded from consideration. Instead, if such evaluations are included, a judgment should be made about the confidence that can be placed in their findings in relation to the findings of other evaluations.

Key GAO Reports

Designing Evaluations: 2012 Revision. [GAO-12-208G](#). Washington, D.C.: January 2012.

Assessing the Reliability of Computer-Processed Data. [GAO-09-680G](#). Washington, D.C.: July 2009.

The Evaluation Synthesis. [GAO/PEMD-10.1.2](#). Washington, D.C.: March 1992.