



PERFORMANCE MONITORING & EVALUATION

TIPS

CONDUCTING MIXED-METHOD EVALUATIONS

ABOUT TIPS

These TIPS provide practical advice and suggestions to USAID managers on issues related to performance monitoring and evaluation. This publication is a supplemental reference to the Automated Directive System (ADS) Chapter 203.

INTRODUCTION

This TIPS provides guidance on using a mixed-methods approach for evaluation research. Frequently, evaluation statements of work specify that a mix of methods be used to answer evaluation questions. This TIPS includes the rationale for using a mixed-method evaluation design, guidance for selecting among methods (with an example from an evaluation of a training program) and examples of techniques for analyzing data collected with several different methods (including “parallel analysis”).

MIXED-METHOD EVALUATIONS DEFINED

A mixed-method evaluation is one that uses two or more techniques or methods to collect the data needed to answer one or more evaluation questions. Some of the different data collection methods that might be combined in an evaluation include structured observations, key informant interviews, pre- and post-test surveys, and reviews of government statistics. This could involve the collection and use of both quantitative and qualitative data to analyze and identify findings and to develop conclusions in response to the evaluation questions.

RATIONALE FOR USING A MIXED-METHOD EVALUATION DESIGN

There are several possible cases in which it would be highly beneficial to employ mixed-methods in an evaluation design:

- When a mix of different methods is used to collect data from different sources to provide independent estimates of key indicators—and those estimates complement one another—it increases the validity of conclusions related to an evaluation question. This is referred to as triangulation. (See TIPS 5: Rapid Appraisal, and Bamberger, Rugh and

Key Steps in Developing a Mixed-Method Evaluation Design and Analysis Strategy

1. In order to determine the methods that will be employed, carefully review the purpose of the evaluation and the primary evaluation questions. Then select the methods that will be the most useful and cost-effective to answer each question in the time period allotted for the evaluation. Sometimes it is apparent that there is one method that can be used to answer most, but not all, aspects of the evaluation question.
2. Select complementary methods to cover different aspects of the evaluation question (for example, the how and why issues) that the first method selected cannot alone answer, and/or to enrich and strengthen data analysis and interpretation of findings.
3. In situations when the strength of findings and conclusions for a key question is absolutely essential, employ a triangulation strategy. What additional data sources and methods can be used to obtain information to answer the same question in order to increase the validity of findings from the first method selected?
4. Re-examine the purpose of the evaluation and the methods initially selected to ensure that all aspects of the primary evaluation questions are covered thoroughly. This is the basis of the evaluation design. Develop data collection instruments accordingly.
5. Design a data analysis strategy to analyze the data that will be generated from the selection of methods chosen for the evaluation.
6. Ensure that the evaluation team composition includes members that are well-versed and experienced in applying each type of data collection method and subsequent analysis.

Mabry [2006] for further explanation and descriptions of triangulation strategies used in evaluations.)

- When reliance on one method alone may not be sufficient to answer all aspects of each evaluation question.
- When the data collected from one method can help interpret findings from the analysis of data collected from another method. For example, qualitative data from in-depth interviews or focus groups can help interpret statistical patterns from quantitative data collected through a random-sample survey. This yields a

richer analysis and can also provide a better understanding of the context in which a program operates.

There are a number of additional benefits derived from using a mix of methods in any given evaluation.

- Using mixed-methods can more readily yield examples of unanticipated changes or responses.
- Mixed-method evaluations have the potential of surfacing other key issues and providing a deeper understanding of program context that should be considered when analyzing data and developing findings and conclusions.

- Mixed-method evaluations often yield a wider range of points of view that might otherwise be missed.

DETERMINING WHICH METHODS TO USE

In a mixed-method evaluation, the evaluator may use a combination of methods, such as a survey using comparison groups in a quasi-experimental or experimental design, a review of key documents, a reanalysis of government statistics, in-depth interviews with key informants, focus groups, and structured observations. The selection of methods, or mix, depends on the

nature of the evaluation purpose and the key questions to be addressed.

SELECTION OF DATA COLLECTION METHODS – AN EXAMPLE

The selection of which methods to use in an evaluation is driven by the key evaluation questions to be addressed. Frequently, one primary evaluation method is apparent.

For example, suppose an organization wants to know about the effectiveness of a pilot training program conducted for 100 individuals to set up their own small businesses after the completion of the training.

The evaluator should ask what methods are most useful and cost-effective to assess the question of the effectiveness of that training program within the given time frame allotted for the evaluation. The answer to this question must be based on the ***stated outcome*** expected from the training program. In this example, let us say that the organization's expectations were that, within one year, 70 percent of the 100 individuals that were trained will have used their new skills and knowledge to start a small business.

What is the best method to determine whether this outcome has been achieved? The most cost-effective means of answering this question is to

survey 100 percent of the individuals who graduated from the training program using a close-ended questionnaire. It follows that a survey instrument should be designed to determine if these individuals have actually succeeded in starting up a new business.

While this sounds relatively straightforward, organizations are often interested in related issues. If less than 70 percent of the individuals started a new business one year after completion of the training, the organization generally wants to know why some graduates from the program were successful while others were not. Did the training these individuals received actually help them start up a small business? Were there topics that should have been covered to more thoroughly prepare them for the realities of setting up a business? Were there other topics that should have been addressed? In summary, this organization wants to learn not only whether at least 70 percent of the individuals trained have started up a business, but also how effectively the training equipped them to do so. It also wants to know both the strengths and the shortcomings of the training so that it can improve future training programs.

The organization may also want to know if there were factors outside the actual intervention that had a bearing on the training's success or failure. For example, did some individuals find employment instead? Was

access to finance a problem? Did they conduct an adequate market analysis? Did some individuals start with prior business skills? Are there factors in the local economy, such as local business regulations, that either promote or discourage small business start-ups? There are numerous factors which could have influenced this outcome.

The selection of additional methods to be employed is, again, based on the nature of each aspect of the issue or set of related questions that the organization wants to probe.

To continue with this example, the evaluator might expand the number of survey questions to address issues related to the effectiveness of the training and external factors such as access to finance. These additional questions can be designed to yield additional quantitative data and to probe for information such as the level of satisfaction with the training program, the usefulness of the training program in establishing a business, whether the training graduate received a small business start-up loan, if the size of the loan the graduate received was sufficient, and whether graduates are still in the process of starting up their businesses or instead have found employment. Intake data from the training program on characteristics of each trainee can also be examined to see if there are any particular characteristics, such as sex or ethnic background, that can be correlated with the survey findings.

It is important to draw on additional methods to help explain the statistical findings from the survey, probe the strengths and shortcomings of the training program, further understand issues related to access to finance, and identify external factors affecting success in starting a business. In this case, the evaluation design could focus on a sub-set of the 100 individuals to obtain additional qualitative information. A selected group of 25 people could be asked to answer an additional series of open-ended questions during the same interview session, expanding it from 30 minutes to 60 minutes. Whereas asking 100 people open-ended questions would be better than just 25 people, costs prohibit interviewing the entire group.

Using the same example, suppose the organization has learned through informal feedback that access to finance is likely a key factor in determining success in business start-up in addition to the training program itself. Depending on the evaluation findings, the organization may want to design a finance program that increases access to loans for small business start-ups. To determine the validity of this assumption, the evaluation design relies on a triangulation approach to assess whether and how access to finance for business start-ups provides further explanations regarding success or failure outcomes. The design includes a plan to collect data from two

other sources using a separate data collection method for each source. The first data source includes the quantitative data from the survey of the 100 training graduates. The evaluation designers determine that the second data source will be the managers of local banks and credit unions that survey respondents reported having approached for start-up loans. In-depth interviews will be conducted to record and understand policies for lending to entrepreneurs trying to establish small businesses, the application of those policies, and other business practices with respect to prospective clients. The third data source is comprised of bank loan statistics for entrepreneurs who have applied to start up small businesses. Now there are three independent data sources using different data collection methods to assess whether access to finance is an additional key factor in determining small business start-up success.

In this example, the total mix of methods the evaluator would use includes the following: the survey of all 100 training graduates, data from open-ended questions from a subset of graduates selected for longer interviews, analysis of training intake data on trainee characteristics, in-depth interviews with managers of lending institutions, and an examination of loan data. The use of mixed-methods was necessary because the client organization in this case not only wanted to know how effective the pilot training course was based

on its own measure of program success, but also whether access to finance contributed to either success or failure in starting up a new business. The analysis of the data will be used to strengthen the training design and content employed in the pilot training course, and as previously stated, perhaps to design a microfinance program.

The last step in the process of designing a mixed-method evaluation is to determine how the data derived from using mixed-methods will be analyzed to produce findings and to determine the key conclusions.

ANALYZING DATA FROM A MIXED-METHOD EVALUATION – DESIGNING A DATA ANALYSIS STRATEGY

It is important to design the data analysis strategy before the actual data collection begins. Having done so, the evaluator can begin thinking about trends in findings from different sets of data to see if findings converge or diverge. Analyzing data collected from a mixture of methods is admittedly more complicated than analyzing the data derived from one method. This entails a process in which quantitative and qualitative data analysis strategies are eventually connected to determine and understand key findings. Several different techniques can be used

to analyze data from mixed-methods approaches, including parallel analysis, conversion analysis, sequential analysis, multilevel analysis, and data synthesis. The choice of analytical techniques should be matched with the purpose of the evaluation using mixed-methods. Table 1 briefly describes the different analysis techniques and the situations in which each method is best applied. In complex evaluations with multiple issues to address, skilled evaluators may use more than one of these techniques to analyze the data.

EXAMPLE OF APPLICATION

Here we present an example of parallel mixed-data analysis, because it is the most widely used analytical technique in mixed-method evaluations. This is followed by examples of how to resolve situations where divergent findings arise from the analysis of data collected through a triangulation process.

PARALLEL MIXED-DATA ANALYSIS

Parallel mixed-data analysis is comprised of two major steps:

Step 1: This involves two or more analytical processes. The data collected from *each* method employed must be analyzed separately. For example, a statistical analysis of quantitative data derived from a survey, a set of height/weight measures, or a set of government statistics is

conducted. Then, a separate and independent analysis is conducted of qualitative data derived from, for example, in-depth interviews, case studies, focus groups, or structured observations to determine emergent themes, broad patterns, and contextual factors. ***The main point is that the analysis of data collected from each method must be conducted independently.***

Step 2: Once the analysis of the data generated by each data collection method is completed, the evaluator focuses on how the analysis and findings from each data set can inform, explain, and/or strengthen findings from the other data set. There are two possible primary analytical methods for doing this – and sometimes both methods are used in the same evaluation. Again, the method used depends on the ***purpose*** of the evaluation.

- In cases where more than one method is used specifically to strengthen and validate findings for the same question through a triangulation design, the evaluator compares the findings from the independent analysis on each data set to determine if there is a convergence of findings. This method is used when it is critical to produce defensible conclusions that can be used to inform major program decisions (e.g., end or extend a program).
- To interpret or explain findings from quantitative analysis,

evaluators use findings from the analysis of qualitative data. This method can provide a richer analysis and set of explanations affecting program outcomes that enhance the utility of the evaluation for program managers. Conversely, patterns and associations arising from the analysis of quantitative data can inform additional patterns to look for in analyzing qualitative data. The analysis of qualitative data can also enhance the understanding of important program context data. This method is often used when program managers want to know not only whether or not a program is achieving its intended results, but also, why or why not.

WHEN FINDINGS DO NOT CONVERGE

In cases where mixed-method evaluations employ triangulation, it is not unusual that findings from the separate analysis of each data set do not automatically converge. If this occurs, the evaluator must try to resolve the conflict among divergent findings. This is not a disaster. Often this kind of situation can present an opportunity to generate more nuanced explanations and important additional findings that are of great value.

One method evaluators use when findings from different methods diverge is to carefully re-examine the raw qualitative data through a second and more in-depth content analysis. This is done to

determine if there were any factors or issues that were missed when these data were first being organized for analysis. The results of this third layer of analysis can produce a deeper understanding of the data, and can then be used to generate new interpretations. In some cases, other factors external to the program might be discovered through contextual analysis of economic, social or political conditions or an analysis of operations and interventions across program sites.

Another approach is to reanalyze all the disaggregated data in each data set separately, by characteristics of the respondents as appropriate to the study, such as age, gender, educational background, economic strata, etc., and/or by geography/locale of respondents.

The results of this analysis may yield other information that can help to resolve the divergence of findings. In this case, the evaluator should attempt to rank order these factors in terms of frequency of occurrence. This further analysis will provide additional explanations for the variances in findings. While most professionals build this type of disaggregation into the analysis of the data during the design phase of the evaluation, it is worth reexamining patterns from disaggregated data.

Evaluators should also check for data quality issues, such as the validity of secondary data sources or possible errors in survey data

from incomplete recording or incorrect coding of responses. (See TIPS 12: Data Quality Standards.) If the evaluators are still at the program site, it is possible to resolve data quality issues with limited follow-up data collection by, for example, conducting in-depth interviews with key informants (if time and budget permit).

In cases where an overall summative program conclusion is required, another analytical tool that is used to resolve divergent findings is the data synthesis method. (See Table 2.) This method rates the strength of findings generated from the analysis of each data set based on the intensity of the impact (e.g., on a scale from very high positive to very high negative) and the quality and validity of the data. An overall rating is assigned for each data set, but different weights can then be assigned to different data sets if the evaluator knows that certain data sources or methods for collecting data are stronger than others.

Ultimately, an index is created based on the average of those ratings to synthesize an overall program effect on the outcome. See McConney, Rudd and Ayres (2002) to learn more about this method.

REPORTING ON MIXED-METHOD EVALUATIONS

Mixed-method evaluations generate a great deal of data,

and, to profit from the use of those methods, evaluators must use and analyze all of the data sets. Through the use of mixed-method evaluations, findings and conclusions can be enriched and strengthened. Yet there is a tendency to underuse, or even not to use, all the data collected for the evaluation. Evaluators can rely too heavily on one particular data source if it generates easily digestible and understandable information for a program manager. For example, in many cases data generated from qualitative methods are insufficiently analyzed. In some cases only findings from one source are reported.

One way to prevent underutilization of findings is to write a statement of work that provides the evaluator sufficient time to analyze the data sets from each method employed, and hence to develop valid findings, explanations, and strong conclusions that a program manager can use with confidence. Additionally, statements of work for evaluation should require evidence of, and reporting on, the analysis of data sets from **each** method that was used to collect data, or methodological justification for having discarded any data sets.

REFERENCES

Bamberger, Michael, Jim Rugh and Linda Mabry. *Real World Evaluation: Working Under Budget, Time, Data and Political Constraints*, Chapter 13, "Mixed-Method Evaluation," pp. 303-322, Sage Publications Inc., Thousand Oaks, CA, 2006.

Greene, Jennifer C. and Valerie J. Caracelli. "Defining and Describing the Paradigm Issue in Mixed-methods Evaluation," in *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms*, Green and Caracelli eds. *New Directions for Evaluation*. Josey-Bass Publishers, No. 74, Summer 1997, pp 5-17.

Mark, Melvin M., Irwin Feller and Scott B. Button. "Integrating Qualitative Methods in a Predominantly Quantitative Evaluation: A Case Study and Some Reflections," in *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms*, Green and Caracelli eds. *New Directions for Evaluation*. Josey-Bass Publishers, No. 74, Summer 1997, pp 47-59.

McConney, Andrew, Andy Rudd, and Robert Ayres. "Getting to the Bottom Line: A Method for Synthesizing Findings Within Mixed-method Program Evaluations," in *American Journal of Evaluation*, Vol. 3, No. 2, 2002, pp. 121-140.

Teddlie, Charles and Abbas Tashakkori, *Foundations of Mixed-methods Research: Integrating Quantitative and Qualitative Approaches in the Behavioral Science*, Sage Publications, Inc., Los Angeles, 2009.

TABLE 1 – METHODS FOR ANALYZING MIXED-METHODS DATA¹

Analytical Method	Brief Description	Best for...
Parallel	Two or more data sets collected using a mix of methods (quantitative and qualitative) are analyzed independently . The findings are then combined or integrated.	Triangulation designs to look for convergence of findings when the strength of the findings and conclusions is critical, or to use analysis of qualitative data to yield deeper explanations of findings from quantitative data analysis.
Conversion	Two types of data are generated from one data source beginning with the form (quantitative or qualitative) of the original data source that was collected. Then the data are converted into either numerical or narrative data. A common example is the transformation of qualitative narrative data into numerical data for statistical analysis (e.g., on the simplest level, frequency counts of certain responses).	Extending the findings of one data set, say, quantitative, to generate additional findings and/or to compare and potentially strengthen the findings generated from a complimentary set of, say, qualitative data.
Sequential	A chronological analysis of two or more data sets (quantitative and qualitative) where the results of the analysis from the first data set are used to inform the analysis of the second data set. The type of analysis conducted on the second data set is dependent on the outcome of the first data set.	Testing hypotheses generated from the analysis of the first data set.
Multilevel	Qualitative and quantitative techniques are used at different levels of aggregation within a study from at least two data sources to answer interrelated evaluation questions. One type of analysis (qualitative) is used at one level (e.g., patient) and another type of analysis (quantitative) is used in at least one other level (e.g., nurse).	Evaluations where organizational units for study are nested (e.g., patient, nurse, doctor, hospital, hospital administrator in an evaluation to understand the quality of patient treatment).
Data Synthesis	A multi-step analytical process in which: 1) a rating of program effectiveness using the analysis of each data set is conducted (e.g., large positive effect, small positive effect, no discernable effect, small negative effect, large negative effect; 2) quality of evidence assessments are conducted for each data set using "criteria of worth" to rate the quality and validity of each data set gathered; 3) using the ratings collected under the first two steps, develop an aggregated equation for each outcome under consideration to assess the overall strength and validity of each finding; and 4) average outcome-wise effectiveness estimates to produce one overall program-wise effectiveness index.	Providing a bottom-line measure in cases where the evaluation purpose is to provide a summative program-wise conclusion when findings from mixed-method evaluations using a triangulation strategy do not converge and appear to be irresolvable , yet a defensible conclusion is needed to make a firm program decision. Note: there may still be some divergence in the evaluation findings from mixed data sets that the evaluator can still attempt to resolve and/or explore to further enrich the analysis and findings.

¹ See Teddlie and Tashakkori (2009) and Mark, Feller and Button (1997) for examples and further explanations of parallel data analysis. See Teddlie and Tashakkori (2009) on conversion, sequential, multilevel, and fully integrated mixed-methods data analysis; and McConney, Rudd, and Ayers (2002), for a further explanation of data synthesis analysis.

For more information:

TIPS publications are available online at [insert website].

Acknowledgements:

Our thanks to those whose experience and insights helped shape this publication including USAID's Office of Management Policy, Budget and Performance (MPBP). This publication was written by Dr. Patricia Vondal of Management Systems International.

Comments regarding this publication can be directed to:

Gerald Britan, Ph.D.

Tel: (202) 712-1158

gbritan@usaid.gov

Contracted under RAN-M-00-04-00049-A-FY0S-84

Integrated Managing for Results II