



The Mixed Methods Approach to Evaluation

Michael Bamberger¹

NUMBER 1

JUNE 2013



Executive Summary

Over the past decade there has been an increased demand for mixed-methods evaluations to better understand the complexity of international development interventions and in recognition of the fact that no single evaluation methodology can fully capture and measure the multiple processes and outcomes that every development program involves. At the same time, no consensus has been reached by policy makers and evaluation practitioners as to what exactly constitutes a mixed-methods approach. This SI Concept Note aims at helping that discussion by defining mixed-methods as evaluation approaches that *systematically integrate* quantitative and qualitative research methodologies at all stages of an evaluation. The paper further discusses the most important strengths and weaknesses of mixed-methods approaches compared to quantitative and qualitative only evaluations and lists a number of implementation challenges and ways to address them that may be useful to both producers and consumers of performance and impact evaluations.

¹ Senior Adviser for Evaluations at Social Impact. The author would like to thank Kyle Block, Chiara Cruciano, Luca Etter, Leslie Greene and Danae Roumis for their valuable inputs and feedback.

Cover Picture: SI mixed-methods performance evaluation of the USAID Growth with Equity in Mindanao III (GEM-3) project in the Philippines. Photo Credit: Dr. Mary Judd, Team Leader for GEM-3 Evaluation, SI.

Section I –Defining Mixed Methods

Mixed methods is an evaluation approach that involves the systematic integration of different methodologies and methods² at all stages of an evaluation. The mixed methods approach normally refers to evaluation designs that combine Quantitative (QUANT) and Qualitative (QUAL) methods, but some evaluators also use the term to include evaluations that combine a number of different QUANT methods.³ Unless otherwise stated we will be discussing the most common usage, namely the combination of QUANT and QUAL methods and define a mixed methods evaluation as:

An approach to evaluation that *systematically integrates* QUANT and QUAL methodologies and methods at all stages of an evaluation.

It is important to distinguish the systematic integration of QUANT and QUAL methodologies and methods from the many evaluations that combine QUANT and QUAL methods in an ad hoc manner. For example, an evaluation that complements a QUANT household survey with a few case studies or focus groups that are selected in an unsystematic manner making it difficult to know if the selected cases or focus groups are representative of the population covered by the household survey would not be considered an appropriate application of the mixed methods approach.

Mixed methods can be used to strengthen monitoring and evaluation approaches used to assess any stage of

No single evaluation approach can fully understand and respond to the complexities of most development interventions. Understanding how development *really works* requires the integration of a wide range of evaluation tools and techniques drawn from multiple disciplines.

a development program – including diagnostic and planning studies, constructing baseline data, assessment of program implementation, and evaluating outcomes and impacts.

Section II—Rationale for Mixed Methods Evaluation

There has been a rapidly growing interest in mixed methods approaches to evaluation over the past decade. This is evidenced by the launch of several journals, books, handbooks, and websites. It is also an increasingly common practice in Calls for Proposals (sometimes referred to as Requests for Proposals, RFPs) and development agency evaluation guidelines to require the use of mixed methods. Perhaps the main reason for this growing interest is the recognition that no single evaluation methodology can respond to the many different kinds of evaluation questions of interest to clients and stakeholders, nor can any single design address the increasing complexity of development interventions.⁴ For example:

- Programs operate in complex and changing social, economic, ecological and political contexts and no single evaluation methodology can adequately describe the interactions among all of these different factors.⁵
- Program implementation and outcomes are affected by a wide range of historical, economic, political, cultural, organizational, demographic and natural environmental factors, all of which require different methodologies for their assessment.
- Programs also produce a range of different outcomes and impacts, many of which require different methodologies for measurement and assessment.
- Many important outcomes such as poverty, vulnerability, security and empowerment combine a number of different dimensions, which can be difficult to observe and measure (such as who

² The term “methodology” refers to the overall approach of the evaluation design (for example, using triangulation to compare estimates obtained from three different methods of data collection), while “method” refers to a particular technique such as the use of focus groups or participant observation or a household survey.

³ Theoretically one could also use the term mixed methods to refer to an evaluation that uses a number of different QUAL methods, but the term is rarely used in this way.

⁴ We will use “programs” as shorthand for projects, programs and more complex development interventions except where we wish to distinguish between different levels of complexity.

⁵ Many evaluations refer to all of the complex interactions among these contextual factors as part of the “system” in which the program is implemented, and there is an extensive literature on how systems analysis addresses the interactions among all of these factors. See for example B. Williams and R. Hummelbrunner’s *Systems Concepts in Action: A practitioner’s toolkit*. Stanford University Press, 2010.

controls the use of household productive resources, domestic violence, sexual harassment and leadership styles). An important application of mixed methods is to combine a number of different methods for collecting and interpreting data on a key outcome. Triangulation (discussed in Section VII) is then used to compare the estimates obtained from different methods and to assess whether they are consistent with each other.⁶

- Even seemingly simple projects often involve complex processes of organizational and behavioral change. While this may seem obvious, many evaluations of “simple” projects do not take these complex processes into account and simply conduct a pretest-posttest comparison of the project and control group population – ignoring the influence of contextual factors. This is an example where a rapid and economical mixed methods design can greatly enhance the evaluation.
- Programs change in response to how they are perceived and received by different sectors of the target (and non-target) population, and observing these processes of behavioral change often requires the application of different methodologies.⁷
- Even when the focus of the evaluation is on the assessment of outcomes and impacts, it is almost always necessary to assess the processes and quality of implementation – requiring different methodologies to those used to assess outcomes (and impacts).

Section III—Using mixed methods to assess impacts and outcomes – Impact Evaluation and Performance Evaluation

Mixed methods approaches can be used to strengthen both performance and impact evaluations, however, it is important to note the distinction between the two. Many organizations make a distinction between rigorous *impact evaluations* that require the use of experimental or quasi-experimental designs and other

kinds of *performance evaluations* of the achievement of goals, outcomes or objectives. Performance evaluations rely less on statistically rigorous methodologies, and are often also called results-based evaluations, rapid evaluations, ex-post or retrospective evaluations.

When rigorous statistical designs can be used, the methodology for assessing the contribution of the intervention to the observed outcomes is well understood. However, the process of assessing the contribution of a program intervention to desired goals and outcomes for other kinds of performance assessments is much more challenging and there is no standard and widely accepted set of methodologies.

USAID evaluation policies and guidelines, for example, distinguish between *impact evaluations* and *performance evaluations*. An impact evaluation is defined as:

“Impact evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a “treatment” or a “control” group provide the strongest evidence of a relationship between the intervention under study and the outcome measured.”

Ideally, impact evaluations are planned and designed before an intervention occurs, and analysis is produced at or after project completion and is used to plan or scale-up future projects. In contrast performance evaluation is defined:

“...as a range of evaluation options that can be produced at any phase of the program cycle to improve performance.”

⁶When estimates obtained from different methods appear not to be consistent, this often leads to a deeper understanding of the concepts being studied. For example, using different methods to explore the concept of “poverty” will often reveal that many women are equally concerned about vulnerability and lack of access to a social support network as are men.

⁷Realist evaluation is gaining popularity as a way to understand these complex processes of behavioral change (see R. Pawson and N. Tilly 1997 *Realist Evaluation*. Sage Publications)

Performance evaluations can address a range of questions relating to program relevance, program implementation mechanisms, participatory and consultative processes and lessons learned. Many of these questions relate to the extent to which the program is achieving its intended results and reaching the intended target populations. However, they are framed in a broader context as they also ask what the program is actually doing and delivering. For most programs within USAID, as in many other organizations, questions about impacts, effects, outcomes or results will be addressed through performance evaluation and without access to more rigorous impact evaluation methodologies. Normally performance evaluations will use non-experimental designs and weaker quasi-experimental designs to assess outcomes.⁸

Section IV—Strengths and weaknesses of QUANT and QUAL methods when used separately

Both QUANT and QUAL evaluation methods have significant strengths, but when the evaluation relies exclusively on one there are also significant weaknesses. Table 1 lists some of the strengths and weaknesses of QUANT evaluation designs. A major strength is the

ability to control for selection bias (the extent to which assignment to participate in the program (either by the agency or self assignment) influences project outcomes). On the other hand, the weaknesses of QUANT designs include difficulty capturing sensitive types of data or locating and interviewing difficult to reach groups. The process of quantitative data collection and analysis can often reduce rich information to a set of binary “Yes/No” or multiple choice questions. Strictly QUANT designs often do not consider the context in which programs are implemented even though it is possible to collect additional information on these factors. Many QUANT designs also do not take into consideration the project implementation process and how closely what happened on the ground conformed to the intended project design.

Table 2 shows that QUAL evaluation designs have the flexibility to adapt in response to changes in the program or the context in which it is being implemented as the nature of the program is better understood. It is also possible to work with much smaller, carefully targeted samples. The use of QUAL data can provide a richer and more holistic understanding of the program and its context. A narrative report with case studies or direct quotes is easier for many audiences to understand compared to QUANT reports with many tables. However, there

Table 1: Potential strengths and weaknesses of exclusive QUANT evaluation designs

Strengths	Weaknesses
<ul style="list-style-type: none"> • Statistical control of selection bias • The ability to generalize from the sample to the total population with a known level of statistical confidence • The magnitude and distribution of outcomes and impacts can be quantified • Potential to replicate data collection and analysis methods • Standardized methods for sample selection, data collection and analysis • Increases the credibility of the findings for many (but not all) stakeholders 	<ul style="list-style-type: none"> • Surveys cannot capture some types of sensitive information • Survey methods often find it difficult to identify and interview difficult to reach groups • No analysis of the context within which the program is implemented • Survey approach may alienate respondents • Delay in obtaining results • Data reduction loses information (rich responses are often reduced to “Yes/No” or multiple choice responses) • Difficulty in studying the program implementation process

⁸ Many textbooks classify impact evaluation designs into experimental designs; quasi-experimental designs; and non-experimental designs (where there is no matched comparison group). Most performance evaluations use non-experimental or weaker quasi-experimental designs. For a discussion of the most common types of quasi-experimental and non-experimental designs see Bamberger et al (2012) Chapter 11 and Appendix F, which gives an example to illustrate each design.

are also weaknesses. The flexibility of the design may be difficult for many readers who are used to seeing precise timelines, sample designs and draft survey instruments. The difficulty of generalizing from specific case studies and interviews and the presentation of multiple perspectives without a clear “bottom-line” is frustrating for readers who are used to having the findings summarized in bullet points. The interpretivist approach⁹ means that the reader must often rely on the judgment of the evaluator without any way to review the raw data or its interpretation.

Section V –The potential benefits of a mixed methods approach

Mixed methods approaches, *when used systematically*, offer the potential to combine the benefits of both QUANT and QUAL approaches while compensating for the limitations of both approaches when used in isolation. A well-designed mixed methods evaluation can offer a wide range of potential benefits:

- A well-designed mixed methods evaluation is able to draw on a much broader range of QUANT and QUAL tools, techniques and conceptual frameworks at all stages of the evaluation.
- Normally, the design will also incorporate professionals from different disciplines into the core evaluation team.

- Mixed methods assist evaluators in understanding how local contextual factors can explain variations in program implementation and outcomes in different locations.
- Mixed methods combine the representativeness of QUANT methods that allow for a generalization of findings from a sample to a larger population with the ability of QUAL methods to assess the effect of intervening variables (for example ethnicity, community leadership, etc.) on outcomes.

Section VI—Five key decisions when selecting the appropriate mixed methods design

The following section will discuss the application of mixed methods approaches for performance and impact evaluations. There are four key decisions that determine the appropriate mixed methods design:

Decision 1: What are the key questions that the evaluation must address?

Evaluations are normally designed to address one or more of the following questions¹⁰:

- To what extent can specific impacts or outcomes be attributed to the intervention?
- Did the intervention make a difference?

Table 2 Potential strengths and weaknesses of exclusive QUAL evaluation designs

Strengths	Weaknesses
<ul style="list-style-type: none"> • Flexibility to evolve • Sampling focuses on high value subjects • Holistic focus (“the big picture”) • Examine the broader context within which the program operates • Multiple sources provide complex understanding • Narrative reports more accessible to non-specialists • The use of participatory approaches makes it more likely that vulnerable and voiceless groups are heard 	<ul style="list-style-type: none"> • Lack of clear design may frustrate readers • Lack of generalizability • Multiple perspectives - hard to reach consensus • Individual characteristics not isolated • Interpretivist methods appear too subjective

⁹With the interpretivist approach, the presentation of findings and conclusions relies heavily on the interpretation of the research team, and primary data (transcripts, audio recordings) is not always included in the report that would allow the reader to judge the validity of the conclusions. However, QUAL evaluations have rigorous standards for ensuring the validity of the findings and for avoiding researcher bias. For further reading see Bamberger et al 2012, Chapter 7, and the definition of “interpretivist rigor” in the glossary

¹⁰The first four questions are taken from chapter 4 of DFID’s Broadening the Range of Designs and Methods for Impact Evaluations, 2012

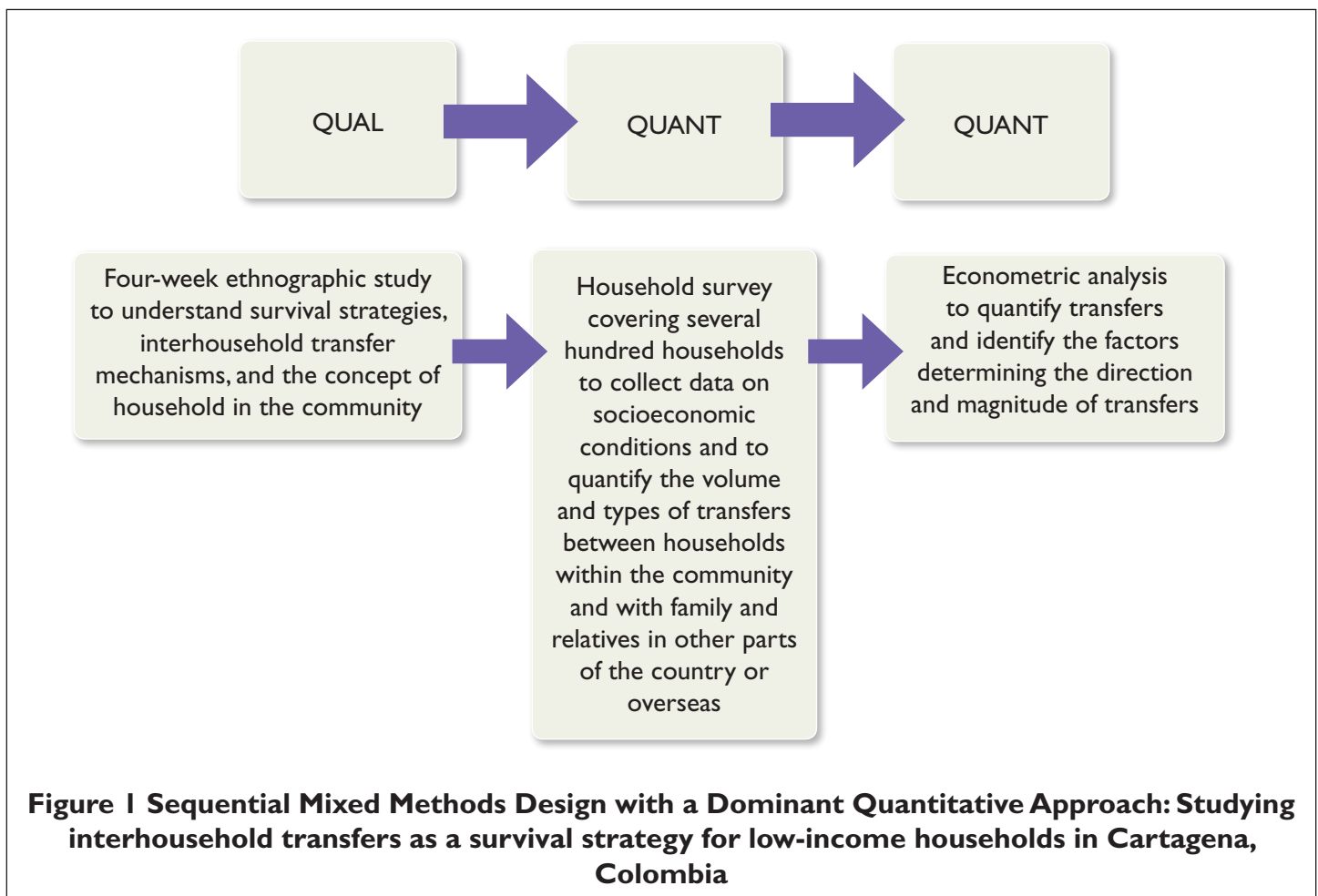
- C. How has the intervention made a difference?
- D. Will the intervention work elsewhere?
- E. Who benefited and who did not?
- F. Were there any unanticipated outcomes, and how important were they (positive or negative)?
- G. Is the program or other intervention sustainable?

The choice of questions will have an important influence on the choice of mixed methods design. For example, where possible Question A might best be addressed with a randomized control trial or pretest-posttest comparison group design complemented by QUAL and perhaps QUANT techniques to assess the process of project implementation. On the other hand, Question D might lend itself to a statistical design complemented by a QUAL contextual analysis to understand the unique characteristics of the project environment that were conducive to, or inhibited the

achievement of the intended outcomes that would be important to take into consideration when assessing potential replicability in other locations.

Decision 2: Will one approach (QUANT or QUAL) be dominant or will a balanced design be used giving equal weight to both approaches?

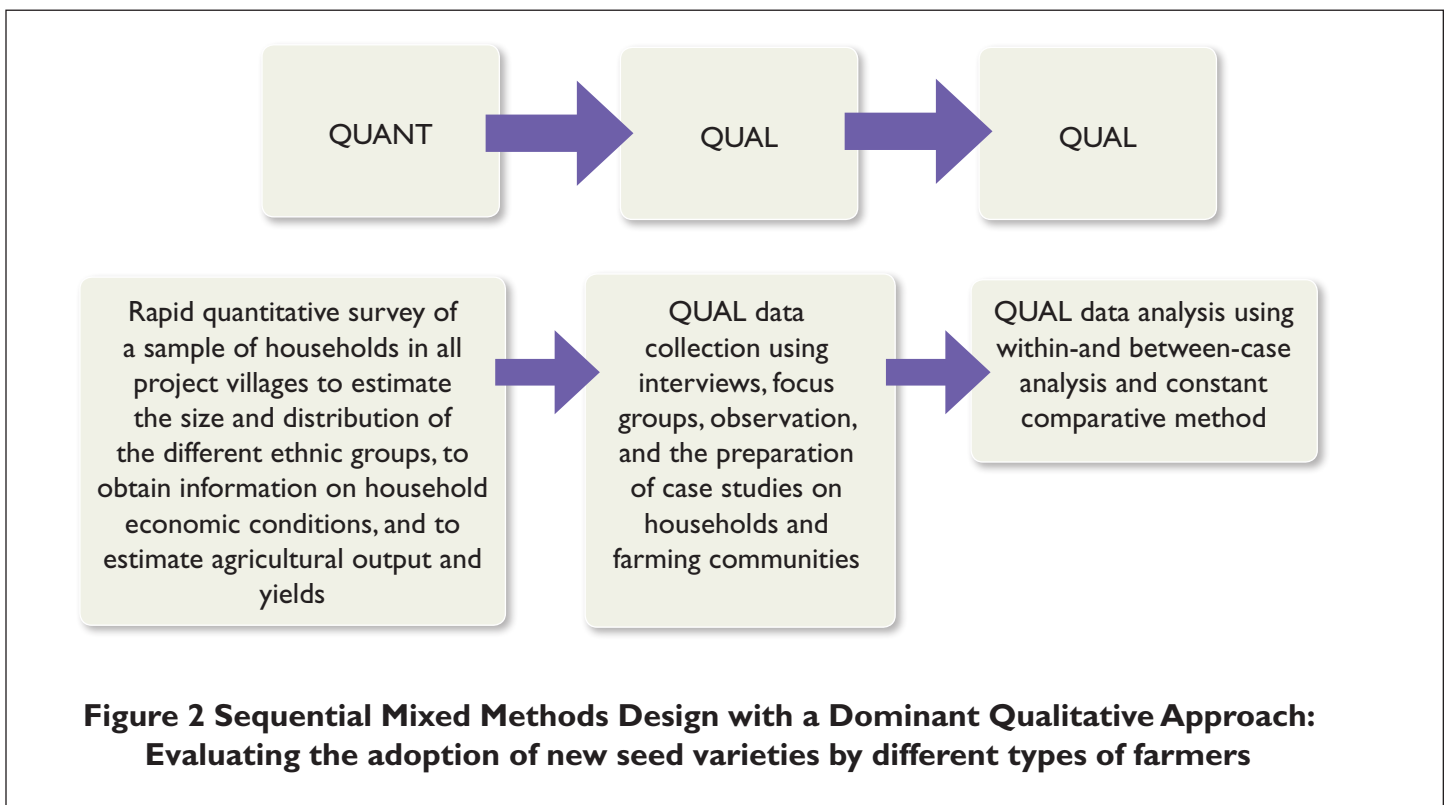
The choice of which orientation is dominant is important because QUANT and QUAL evaluations use mixed methods in very different ways and for very different purposes. Annex II provides examples of the different uses and purposes of mixed methods in QUANT and QUAL dominant evaluations. While the dominant orientation will be partly determined by the context and purpose of an evaluation, the relative importance given to QUANT and QUAL approaches will be significantly influenced by the evaluation preferences of the client and the research team. Different research designs can give different weights to



QUANT and QUAL approaches when evaluating the same project and addressing exactly the same evaluation questions.¹¹

Figure 1 describes a sequential design with a dominant QUANT orientation. This is a study of interhousehold transfers of money and goods as a survival strategy for poor urban households in Colombia.¹² The purpose of the study was to assess the patterns of transfers and to estimate whether they were sufficiently large to act as an informal social safety net providing help to the poorest sectors of the community in times of need. These interhousehold transfers are difficult to identify and measure, so an anthropologist lived in the community for a month to conduct a QUAL study to better understand the patterns of transfers. The data from the QUAL study was used to inform a questionnaire for a QUANT survey which was then administered to several hundred households. The data were analyzed using QUANT econometric analysis. This would be classified as a sequential QUAL-QUANT-QUANT mixed methods approach design.

Figure 2 illustrates a sequential design with a dominant QUAL approach. This describes a hypothetical evaluation to assess the adoption of new varieties of seed by different types of rural families. The principal data collection methods are qualitative: interviews, focus groups, observation, and case studies of individual households and small farming communities. However, to obtain demographic information on the ethnic distribution of households, household economic conditions, and agricultural production, the evaluation begins with a rapid and inexpensive QUANT household survey covering a sample of households in all the villages covered by the agricultural extension project. The findings of this study were used to help identify the types of households to be studied in more depth through the QUAL data collection methods, and to ensure that the selected cases were broadly representative of the total survey population. Either of the previous evaluation methodologies could be modified to give equal weight to both QUANT and QUAL approaches. In the case of the interhousehold transfer study, the household survey could have been



¹¹ For a discussion of the QUANT-QUANT evaluation design continuum see Bamberger, Rugh and Mabry pp. 324-335, RealWorld Evaluation, 2012

¹² Wansbrough, Jones, and Kappaz 2000

complemented with QUAL case studies on families or informal transfer networks. These could then have been integrated into the analysis to compare the description and interpretation of the functions and operation of the transfer networks obtained from the QUAL studies with the findings of the econometric analysis. In the second example, a QUAL or QUANT study of marketing outlets could have been conducted to estimate the changes in sales of agricultural produce from the project areas and, possibly, the changes in the purchase of consumer goods by project area families.

Decision 3: Are QUANT and QUAL approaches used concurrently or sequentially?

QUANT and QUAL methods can either be used sequentially (one after the other) or concurrently (referred to as parallel). A QUAL-dominant sequential design used to evaluate the adoption of new seed varieties by different types of rural households may actually begin with a rapid QUANT study. This study will identify the different types of responses of rural households to the new seed varieties and to estimate the numerical frequency of each type of response (for example, early adopters, late adopters and non-

adopters). A sample of case studies is then selected to ensure that the cases are broadly representative of the different typologies. QUAL data collection methods such as case studies, in-depth interviews, key informants, participant observation are used, and the data is then analyzed qualitatively.

Figure 3 illustrates a concurrent (parallel) mixed methods design used to obtain independent estimates of household income and how it is affected by a project. QUANT and QUAL data collection methods were used consecutively to obtain three independent estimates of household income: a QUANT household survey, QUANT and QUAL observation techniques (such as noting the quality of house construction, ownership of consumer durables and equipment for running a small business), and QUAL focus groups. Estimates from the data collection sources were compared using the triangulation methods described in a later section.

In practice, sequential mixed methods designs are more widely used as they are easier to manage, particularly in situations where experienced supervisory staff is in limited supply. Consecutive designs have the potential advantage reducing the total time required for data collection and analysis, but they require tighter supervision and coordination.

SI in Action

SI conducted a mixed-methods performance evaluation of the “Growth with Equity in Mindanao III” (GEM-3) project in the Philippines. GEM-3 is a five-year program promoting development activities in the Autonomous Region in Muslim Mindanao (ARMM) and other conflict-affected areas of the region. SI applied a QUAL dominated mixed-methods approach that triangulated findings from a desk review of relevant secondary sources; key informant and group interviews with national, regional, and local government officials, USAID and other donor staff, representatives from civil society organizations, implementer staff, and project beneficiaries and non-beneficiaries with a quantitative household mini-survey of over 900 respondents covering 22 provinces, 54 municipalities, five cities, and 87 barangays.

Decision 4: Is a single-level or a multi-level design used?

Most mixed methods evaluation designs are conducted on a single level such as interviews with households or individuals. The examples given in Figures 1, 2, and 3 are all single-level evaluation designs. However, one of the potentially most useful applications of mixed methods is for the evaluation of multi-level programs such as the provision of health, education or other services at the state or district level. The evaluation of these programs typically requires an assessment of how programs operate at the district (or state) level, as well as at the level of the individual school or hospital/clinic, the classroom or individual health service, and at the level of the individual teacher or doctor as well as patient or student. They also require the analysis of the interactions among the different levels as the attitude, qualifications or other factors of the hospital director or the head teacher can influence the attitudes and behavior of individual teachers and health providers. Similarly the attitudes of the parents, community leaders and students or patients also have an important

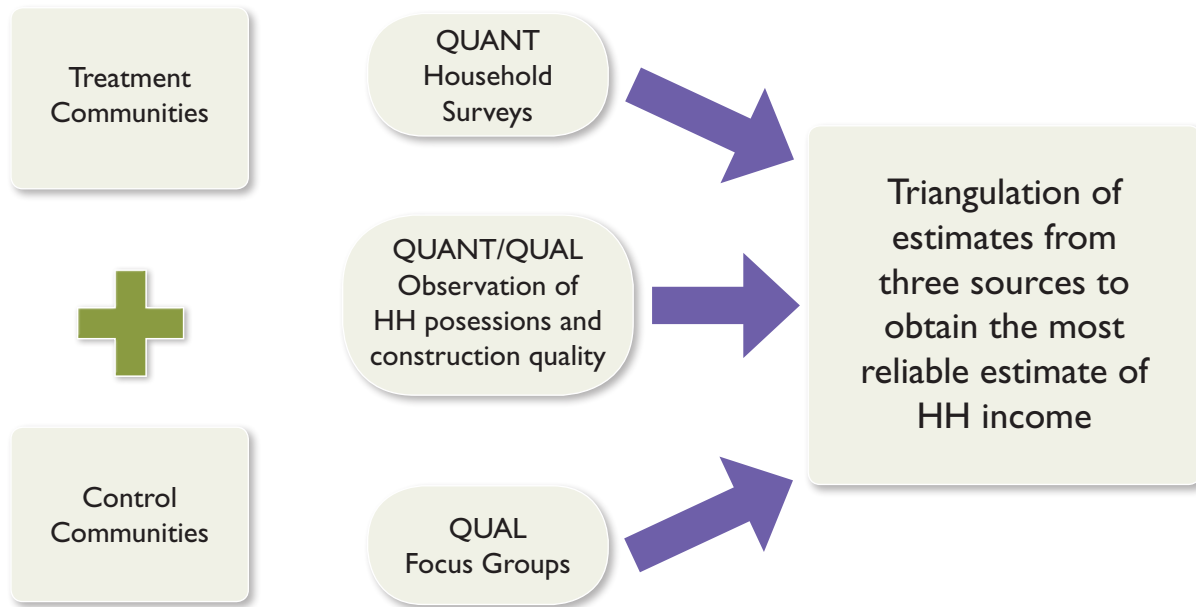


Figure 3 Consecutive Mixed Methods design: Triangulating QUANT and QUAL estimates of household income in project and comparison areas

SI in Action

SI is implementing a rigorous mixed-methods Impact Evaluation of the *A Ganar Alliance* in Honduras and Guatemala. *A Ganar* is a 7-9 month integrated job training program that combines sports-based field and classroom activities, vocational training, internships/apprenticeships and various follow-on activities to help at-risk youth (1) find jobs, (2) start or expand their business, or (3) re-enter the formal education system. SI is applying a QUANT dominated mixed methods approach that triangulates findings from a randomized control trial (RCT) IE with a robust qualitative component that includes in-depth case studies with 12 program participants and post-program interviews with 50 participants. Each multi-perspective case study involves interviews with 5 individuals (participant, their life skills and vocational training facilitators, mentor and employer) over multiple points in time.

influence. Mixed methods are well suited to examine these interactions among different levels, something which is difficult to do with exclusive QUANT or QUAL approaches.

Figure 4 illustrates how a multi-level mixed methods design could be used to evaluate the implementation and impacts of a school feeding program that operates in a number of different schools in a district. QUANT and QUAL data collection and analysis methods are used at each level, and QUAL methods (such as observation, key informants, review of reports) are used to understand interactions among the different levels. Furthermore, QUANT analysis is also able to explore variation at different levels through multilevel statistical analysis.

Section VII—Applying mixed methods at each stage of an evaluation

Stage 1: Hypothesis formulation and development of the theory of change

Generally, mixed methods designs permit the strengthening and broadening of research hypotheses by combining *deductive* (QUANT) and *inductive* (QUAL) hypotheses. Deductive hypotheses are normally defined

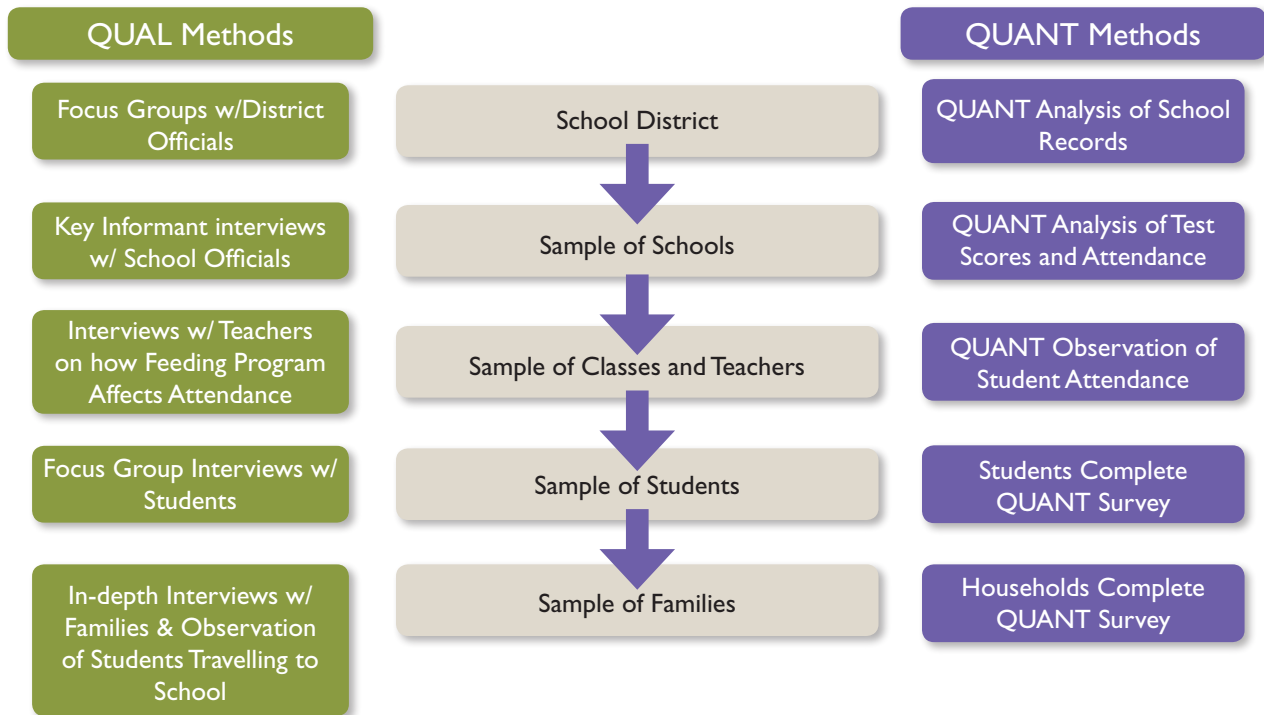


Figure 4 Multi-level Methods Design: The effects of a school feeding program on school attendance

at the start of the evaluation, based on evaluation theory or a review of the literature, and can be tested using experimental or quasi-experimental designs. Despite the statistical rigor with which they can be tested, a disadvantage of deductive hypotheses is that they do not have the flexibility to adapt to changes that arise in project design or implementation, or changes in the context within which the project is implemented. This is a major limitation as many programs undergo significant changes during implementation.

In contrast, many QUAL evaluations use *inductive* hypotheses that are refined and revised as more information is obtained on the project and its context. While it is usually not possible to test these hypotheses in such a rigorous manner, a big advantage is their flexibility to evolve and adapt. Mixed methods can combine the statistical rigor of deductive hypotheses with the flexibility and depth of understanding provided by inductive/emergent methods. This permits the exploration of a much broader range of hypotheses than is normally possible through conventional statistical designs.

Mixed methods designs are normally based on a theory of change which provides a framework for identifying the processes through which changes are expected to occur, for identifying and testing key assumptions, and for assessing the effects of contextual factors. A well-articulated theory of change can also provide a framework for identifying unanticipated outcomes – which often have important consequences for the program and which are difficult to assess through most conventional evaluation designs (Bamberger 2012).¹³

Stage 2: Stakeholder consultation

Mixed methods provide a range of tools for identifying a wider range of groups who are affected by the project and whose voices should be heard during program design, implementation and evaluation. At the program evaluation and design stage participatory group consultations provide mechanisms for involving community level stakeholders; while techniques such as snowball sampling,¹⁴ key informants and participant observation can be used to identify vulnerable and difficult to reach groups who are frequently not

¹³ Bamberger 2012 Unanticipated outcomes of development interventions: A blind spot for evaluation theory and practice. IPDET guest lecture. July 2012.

¹⁴ Snowball sampling is non-probability sampling method where interviewees refer the researcher to future interviewees from among their acquaintances.

consulted. There are also a range of participatory methodologies that can involve stakeholders during the implementation of the evaluation. Participant perspectives can also be included in the interpretation of evaluation findings through techniques such as Most Significant Change.¹⁵

Stage 3: The evaluation design

The following are some of the ways that mixed methods can be used to strengthen the evaluation design:

- Strengthening the breadth and depth of the key constructs that are being tested and that are used to explain the processes through which changes take place. This is important because many key constructs such as vulnerability, domestic violence, personal security and wellbeing are complex and multidimensional and there is a tendency in many QUANT evaluations to simplify these to one-dimensional indicators that are easier to analyze.
- Incorporating contextual variables into the analysis. Many evaluations either ignore contextual variables or only incorporate them in an *ad hoc* and anecdotal manner – often not including them in the statistical analysis. Mixed methods both study and incorporate these variables more systematically and also use techniques such as “quantizing” to incorporate these descriptive variables into the statistical analysis.¹⁶
- As discussed earlier, mixed methods also increase the flexibility of the evaluation to adapt to changes in project design or in the context in which programs operate.
- Mixed methods can also build analysis of the process of project implementation into a summative impact

evaluation design so as to resolve the so-called “black box” problem.¹⁷

- Mixed methods can also strengthen the matching of the treatment and comparison groups when this has to be done through judgmental matching or quasi-experimental methods. Techniques such as rapid diagnostic studies, focus groups, key informants, concept mapping¹⁸ and GPS mapping¹⁹ can be used to increase the number of criteria that can be used to match communities or other groups.

Stage 4: Sample selection

Mixed methods combine QUANT statistical precision and unbiased sample selection with the ability of QUAL samples to use small samples that focus on high-value subjects. There are a number of sampling methods that can be used including:

- Concurrent (parallel) mixed methods sampling (described above)
- Sequential sampling: a random sample may be used to identify subjects to be included in a purposive two-phase sample
- Multi-level mixed methods samples (the example of the school feeding program in Figure 4)
- More sophisticated variations of multi-level sampling include nested samples.²⁰

Stage 5: Data collection

Mixed methods draw on all of the conventional data collection methods to combine indicators that can be quantified (height, weight, age, income) with measures that can capture the qualitative and difficult to measure dimensions such as the reasoning that respondents

¹⁵ Rick Davies and Jess Dar (2005) The Most Significant Change (MSC) Technique: A guide to its use. CARE International. UK

¹⁶ One method is to transform descriptive/narrative material into a set of dummy variables that can easily be incorporated into the regression analysis

¹⁷ The “black box” refers to pretest-posttest evaluation designs that do not study what happens during project implementation and how closely what actually happens on the ground compares with the intended project design. This is an important issue when an evaluation does not find that any statistically significant differences were found in outcome variables between the project and comparison groups. With a “black box” design it is not possible to determine whether the failure to detect project impacts means that the project design and concept is not valid (design failure), at least not in this particular situation; or whether there were problems during implementation (implementation failure) so that the project was not able to test the project design and how well it might have worked if well implemented (Bamberger et al op cit pp. 182-3)

¹⁸ Mary Kane and William Trochim (2006) Concept Mapping for Planning and Evaluation (Applied Social Research Methods). Sage Publications.

¹⁹ Steven J. Steinberg and Sheila L. Steinberg (2006). GIS Geographic Information Systems for the Social Sciences: Investigating Space and Place. Sage Publications.

²⁰ Nested samples are sometimes used in multi-level mixed methods sample designs. The sample is selected in a series of stages where each successive stage is selected from within the previous one. An example is the selection of school systems, then the random selection of districts within each system, then the selection of high and low performing schools in each district, then the selection of third grade students in each selected school, and then the intensive study of a sample of students. Sample selection may be random at some stages and purposive at others (example adapted from Teddlie and Tashakkori (2009) Foundations of mixed methods research pp. 190-92. Sage Publications.

provide to questions about their life. QUAL data can also be used to assess the quality of services, attitudes and feelings, and to describe processes. A well-combined set of QUANT and QUAL data can provide a much fuller picture and understanding of the programs and phenomena that are being studied.

Stage 6: Data analysis and interpretation

There are a variety of mixed method data analysis approaches. Triangulation is a key element of the mixed methods approach with respect to ensuring the validity of data and enriching the analysis and interpretation of findings. At the stage of data collection this involves comparing estimates obtained from data collected through different methods to determine whether estimates are consistent, and in cases where they are not to try to understand the reasons for the differences. It is often assumed that triangulation only involves comparing estimates from data that have been collected in different ways: for example comparing what respondents say in response to questions in a structured survey with what is observed, or information obtained from in-depth interviews. While this comparison is

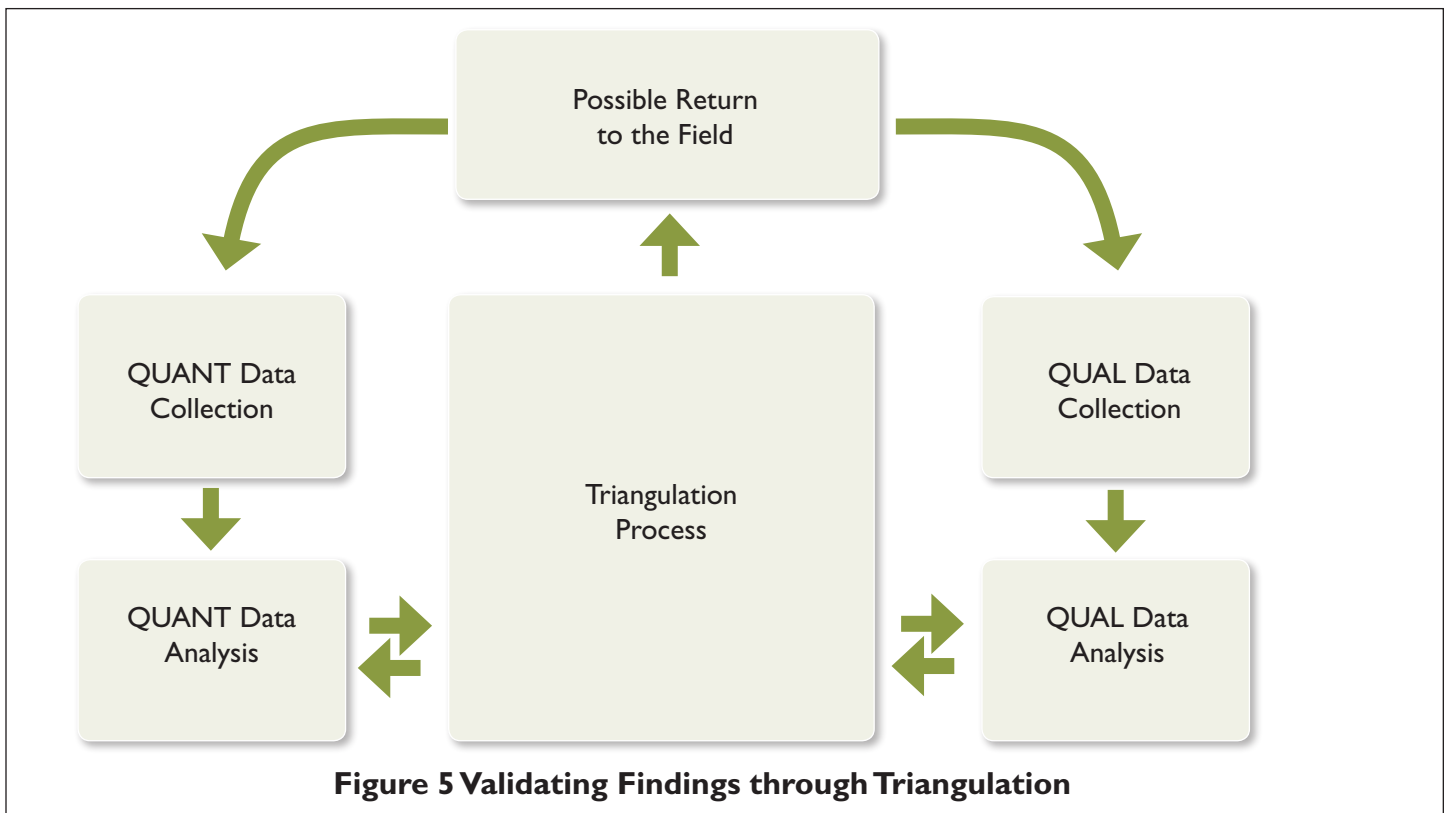
important, triangulation can also compare information collected by different researchers, collected at different times of day or at different times in the year, or collected in different locations (e.g. the home, the street, at work). Other mixed methods data analysis strategies include:

- *Parallel QUANT and QUAL analysis:* QUANT data is analyzed using conventional QUANT data analysis methods while QUAL data is analyzed using QUAL methods. The findings are then brought together and compared but there are no systematic methods for reconciling any inconsistencies.
- *Conversion methods:* QUAL data is converted into QUANT measures using scales or dummy variables (sometimes called “quantizing”) and the transformed variables are then included in the QUANT data analysis. Narrative data, such as transcripts from case studies or focus groups, can be transformed using QUAL data analysis software, such as Atlas.ti. Though less common, QUANT data can also be transformed into QUAL variables. For example, numerical information on changes in household income between the start of the project (say three years ago) and today could be converted into a QUAL scale with categories like: “large increase”, “small increase”, “no change”, “small decrease” and “large decrease”. This is not very common but it could be done if the results were to be presented to communities with a low level of numerical skills.
- *Sequential data analysis:* Each stage of the QUAL-QUANT (or QUANT-QUAL) data collection design is analyzed sequentially using appropriate QUAL and QUANT data analysis methods. The analysis of each stage will identify findings and issues to be incorporated into the subsequent phase.

Figure 5 illustrates how triangulation could be used to compare estimates of socioeconomic status obtained through a structured QUANT household survey measuring household income and from different QUAL methods measuring other wellbeing characteristics, using in-depth interviews, observation and key informants. Data collected through these different methods is discussed by the evaluation teams and if there are significant differences the teams discuss possible reasons for these differences and perhaps they will both review and re-analyze their data to better understand the differences. If the reasons for

SI in Action

SI conducted a Performance Evaluation of the MCC Albania Threshold Programs. These programs constituted of a number of interventions such as revision of the legal framework, introduction of electronic business registration and licensing processes, and awareness raising campaigns aimed at improving the rule of law and reducing corruption. SI applied a mixed-methods evaluation approach combining findings from a large quantitative enterprise survey with over 250 Albanian business owners with focus group discussions and key informant interviews with business owners, union representatives, judges, and politicians. The evaluation also used additional sources of QUANT data such as country level corruption perception indicators. By triangulating qualitative and quantitative methods, SI was able to measure corruption in a multifaceted way.



the differences are still not clear, the evaluation team might return to the field to explore further possible explanations, and the teams may then meet again to try to agree on the best way to explain the differences. In most cases it is usually possible to at least partially explain the differences and the team will then agree on which is the best estimate. Occasionally it will not be possible to explain the differences – and this will have to be explained in the report.

There are several important elements in how mixed methods evaluations use triangulation. First, it is a planned strategy as part of the process of data collection and analysis. It is assumed that there will often be differences and inconsistencies, and triangulation is designed to identify these differences. Second, there should be a defined strategy for trying to understand the reasons for the differences. Time and resources should always be budgeted for team discussions, additional analysis, or a rapid return to the field to seek further explanation. Refer to the sidebar example of a return to the field to check why one village reported that men were responsible for village water management, while in all other villages this water was managed by women. Was this a reporting error, or was there a more

important explanation? The latter proved to be the case. Third, it is important to recognize the effects that the characteristics of the interviewer as well as the time and location where the interview is conducted can have on responses. Large surveys will sometimes be able to compare responses from different interviewers, test for individual interviewer effects, or to compare responses from different categories of interviewers (gender, age group, ethnicity). Often this level of rigor will not be possible, but the evaluator team should always be aware of the influence of these kinds of factors on response rates, and they should try to control for these factors as much as possible.

Stage 7: Dissemination of findings

Some mixed methods evaluators would also argue that mixed methods facilitates a more creative approach to the dissemination of evaluation findings and the efforts made to ensure utilization by relevant stakeholders of the findings and the implementation of recommendations. In addition to the QUANT data that is collected and presented for most evaluations, mixed methods evaluations also collect narrative data and perhaps photographs, audio-recordings and video. Many of these evaluations also involve the affected

communities and other stakeholders as partners in the definition of what questions are to be asked and perhaps in the interpretation of findings. All of this information can be used to go beyond the traditional written report with text and tables, often accompanied by a presentation. Some may like or require conventional written reports, while other audiences may prefer some of the formats discussed below. Combined with a participatory philosophy about the purpose of an evaluation, the following are some of the creative ways that findings can be disseminated:

- Written reports may include photos and direct quotations from respondents
- People from the affected communities may be involved in presenting their perspective on the program and the evaluation to clients
- Clients may be invited to visit the affected communities where the presentations will be made partly by the residents and partly by the evaluation team
- Dissemination may be in the form of videos that highlight the entire evaluation process from design to analysis.

Conclusion

There continues to be a steady increase in the use of mixed methods evaluations. This is due in part to a recognition that most development interventions, even seemingly “simple” projects, involve complex processes of behavioral change, are influenced by a wide range of contextual factors, and produce a broad range of outcomes and impacts (both intended and unintended). No single evaluation design can fully capture all of these processes or measure all of the outcomes, and consequently there is a need for mixed methods designs that combine a broad range of QUANT and QUAL methods for the design, implementation, analysis and interpretation of the evaluations. Mixed methods can combine the in-depth insights provided by QUAL methods with the statistical analysis and ability to generalize provided by QUANT methods.

However, at this point in time, many so-called mixed methods evaluations combine QUANT and QUAL data collection and analysis in a somewhat ad hoc way, so there is a need to incorporate more systematic strategies for combining QUANT and QUAL approaches at all stages of the evaluation, including for hypothesis development and overall evaluation design. It is hoped that this SI Concept Note can contribute to the development of more systematic mixed methods strategies so as to fully benefit from this approach.

Annex I Examples of how mixed methods are used in QUANT and QUAL dominant mixed methods evaluations

QUANT dominant mixed methods designs	
Reasons for using mixed methods	Mixed methods tools and techniques
1. Exploratory studies to help design the QUANT data collection instruments	<ul style="list-style-type: none"> • Rapid diagnostic studies • More in-depth ethnographic studies
2. To strengthen the validity of constructs, scale development, and indicators	<ul style="list-style-type: none"> • Combining several independent QUANT and QUAL indicators and using triangulation to assess validity and reliability
3. To collect data on sensitive topics	<ul style="list-style-type: none"> • Observation, participant observation, focus groups, key informants, in-depth interviews
4. To help identify and to interview difficult to reach groups	<ul style="list-style-type: none"> • Snowball sampling • Case studies • Participant observation
5. To explore outliers and findings that cannot be explained through the statistical analysis	<ul style="list-style-type: none"> • Rapid follow-up studies • Case studies • Focus groups
6. To explore in more depth different typologies and patterns identified in the statistical analysis	<ul style="list-style-type: none"> • Case studies • Participant observation
7. To illustrate different responses and outcomes	<ul style="list-style-type: none"> • Case studies • Participant observation
8. To explore patterns of behavioral change	<ul style="list-style-type: none"> • Realist evaluation • Observation • Key informants
9. To understand context	<ul style="list-style-type: none"> • Descriptive studies combined with compilation of available statistical data • Key informants
10. To examine the process of project implementation and the quality of services	<ul style="list-style-type: none"> • Panel studies of a small group of informants who are visited periodically to discuss what they have heard about the project and their own experiences • Observation, participant observation • Key informants
11. To enrich and broaden the analysis and interpretation of findings	<ul style="list-style-type: none"> • Triangulation
12. To strengthen generalizability of findings to other locations and contexts	<ul style="list-style-type: none"> • Analysis of contextual factors affecting outcomes
QUAL dominant mixed methods designs	
Reasons for using mixed methods	Mixed methods tools and techniques
1. Using selection procedures to ensure cases and subjects for in-depth analysis are broadly representative of the total population	<ul style="list-style-type: none"> • Using the sampling frame for the QUANT surveys to select the QUAL cases • Using preliminary analysis of the QUANT survey to create a typology of household/schools/farms responses to the program to select the QUAL cases
2. To permit quantification of the findings	<ul style="list-style-type: none"> • Using data from the QUANT surveys to estimate the proportion of the total population that fall into each of the categories covered by the case studies
3. To permit statistical analysis of narrative reports	<ul style="list-style-type: none"> • Content analysis and other statistical techniques for qualitative data analysis
4. To analyze the physical location of subjects, services and important resources and barriers affecting the program	<ul style="list-style-type: none"> • GIS mapping technology

Annex II Examples of evaluation designs at each point on the QUANT-QUAL design continuum

Design	Some applications
QUANT dominant designs	
A. Completely QUANT design	Evaluation based on the analysis of secondary survey data that covers both project and control areas. Where the samples are sufficiently large it may be possible to use techniques such as propensity score matching.
B. Dominant QUANT design	Evaluation is mainly based on the application of a structured questionnaire to a randomly selected sample of individuals or households. QUAL techniques are only used to develop, validate or illustrate the QUANT data collection and analysis. Exploratory QUAL interviews and observation used to test the survey instrument and a small number of case studies may be used to illustrate the main groups identified in the QUANT analysis. The case studies are only used for illustrative purposes.
Design strategies that give equal weight to QUANT and QUAL approaches	
C. Designs originating with a QUANT orientation that give equal weight to QUANT and QUAL components	Usually a large scale QUANT sample survey (e.g., households, communities, organizations) is used. QUAL techniques are used to conduct exploratory studies to identify issues and to formulate questions to help develop this instrument. The QUANT survey is complemented by process and/or contextual analysis to understand factors explaining differences in outcome in different project sites. Follow-up case studies or focus groups may be conducted to explore and illustrate in more depth some of the key issues arising from the survey analysis.
D. Designs that originate as MM without either a QUANT or QUAL orientation	QUANT surveys are combined with a range of different QUAL techniques. Sometimes the latter focus on process and contextual analysis, in other cases the focus is on the same unit of analysis as the surveys (e.g., individuals, households) but different data collection methods are used.
E. Designs originating with a QUAL orientation that give equal weight to QUANT and QUAL components.	QUANT surveys may be used to identify key issues or key groups to be explored in more depth. Selection procedures are used to ensure that units selected for in-depth study are at least broadly representative of the total population. Rapid follow-up QUANT surveys may be conducted to assess the generalizability of the QUAL findings.
QUAL dominant designs	
F. Predominantly QUAL design	A rapid QUANT survey is used either to identify the issues or groups to be covered in the in-depth QUAL studies or to show that the QUAL sample is reasonably representative of the total population
G. Completely QUAL design	The evaluation is based exclusively on QUAL techniques. In many cases data will be collected from a small number of individuals or groups complemented by general description of the setting. In other cases the study may have a broader focus using artifacts and other kinds of secondary data to study a community or broader culture.

Annex III Annotated Bibliography

Bamberger, M. (ed). 2000. *Integrating Quantitative and Qualitative Research in Development Projects*. Directions in Development. Washington, DC: World Bank.

Case studies on how integrated mixed-method evaluations have been conducted on education, health, poverty, and water supply projects in different parts of the world.

Bamberger, M; Rugh, J and Mabry, L 2012. *RealWorld Evaluation: working under budget, time, data and political constraints*. Sage Publications. Chapter 13 “Mixed methods evaluation”

Creswell, J. 2003. *Research Design: Qualitative, Quantitative and Mixed-Methods Approaches*. Thousand Oaks, CA: Sage.

Chapter 11 provides an overview of the main stages in the design and implementation of a mixed-method evaluation and of six mixed-method designs.

Greene, J. C. and V. J. Caracelli. 2003. “Making Paradigmatic Sense of Mixed Methods Practice.” Pp. 91–110 in *Handbook of Mixed Methods in Social & Behavioral Research*, edited by A. Tashakkori and C. Teddlie. Thousand Oaks, CA: Sage.

The authors argue that research decisions are rarely consciously rooted in philosophical assumptions or beliefs. They recommend that more attention should be paid to understanding the philosophical and methodological distinctions between QUANT and QUAL while at the same time balancing this with practical and political considerations relevant to the context of each study.

Greene, J.C. 2007. *Mixed Methods in Social Enquiry*. San Francisco. Wiley.

In addition to a discussion of the methodology of mixed methods, Greene examines the rationale, purposes and potentialities of mixed methods as a particular way of thinking. She argues that an important contribution of mixed methods is to generate new and more insightful understanding of complex social phenomena.

Mertens, D. M. 2003. “Mixed Methods and the Politics of Human Research: The Transformatory-Emancipatory Perspective.” Pp. 135–66 in *Handbook of Mixed Methods in Social & Behavioral Research*, edited by A. Tashakkori and C. Teddlie Thousand Oaks, CA: Sage.

The author argues that mixed methods are a powerful tool for contributing to the transformatory-emanicipatory perspective that recommends the adoption of an explicit goal for research to serve the ends of creating a more just and democratic society that permeates the entire research process, from the problem formulation to the drawing of conclusions and the use of results. Readers will find this provides a very different perspective than do some of the other more methodologically focused references.

Teddlie, C & Tashakkori, A. 2009. *Foundations of Mixed Methods Research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks. Sage.

This is probably the more complete and systematic presentation of all of the stages of the formulation, design, implementation, analysis and interpretation of mixed method designs.

Tashakkori, A. and C. Teddlie, eds. 2010 (Second Edition) . *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks, CA: Sage.

Provides a comprehensive but technically more difficult selection of readings on these questions. Chapter 24 discusses how mixed methods have been applied in the evaluation of international development projects and presents three detailed case studies two of which have a more quantitative focus while the third is more qualitative.