Combining Quantitative and Qualitative Methods for Program Monitoring and Evaluation: Why Are Mixed-Method Designs Best?

Michelle Adato

Despite significant methodological advances, much program evaluation and monitoring data are of limited utility because of an over-reliance on quantitative methods alone. While surveys provide generalizable findings on what outcomes or impacts have or have not occurred, qualitative methods are better able to identify the underlying explanations for these outcomes and impacts, and therefore enable more effective responses. Qualitative methods also inform survey design, identify social and institutional drivers and impacts that are hard to quantify, uncover unanticipated issues, and trace impact pathways. When used together, quantitative and qualitative approaches provide more coherent, reliable, and useful conclusions than do each on their own. This note identifies key elements of good mixed-method design and provides examples of these principles applied in several countries.

Over the last decade, development programs in Latin America, Asia, and Africa have increasingly undertaken rigorous impact evaluation. Despite advances, much evaluation and program monitoring data have limited utility because of an over-reliance on quantitative methods alone. While surveys provide essential data on whether or not changes have occurred as a result of a program, qualitative methods identify the underlying explanations for why we do or do not observe these changes. Survey methods will tell us, for example, the rate of change in attended hospital births, while qualitative methods will explain why some women now go to hospitals to give birth while others will not, despite a program designed to encourage their attendance. Qualitative methods also improve survey design, identify social and institutional impacts that are hard to quantify, and uncover unanticipated processes or outcomes. Mixed-method approaches are necessary, because whether development programs work as intended depends not only on how efficiently resources and knowledge are transferred, but also on complex economic and social dynamics in households, communities, and institutions. These dynamics cannot be disentangled through surveys alone. This note provides guidance on how to combine quantitative and qualitative methods for monitoring and evaluation (M&E) to maximize the ability to assess program performance and interpret and act on that information. This note also includes examples of different mixed-method designs used in Haiti, South Africa, Nicaragua, Turkey, and Zimbabwe.

What Do Mixed-Method Evaluation Designs Offer?

Quantitative methods provide uniform measures of project outputs and impacts, for example, the number of farmers trained or vaccines adminis-
tered, or changes in income, crop yields, school enrollment, or child stunting. Representative sample sizes ensure that findings are generalizable among a wider population. Econometric analysis further enables inferences of causality and relationships between impacts and explanatory variables.

Quantitative methods perform less well in explaining these results, particularly when explanations involve issues that are hard to quantify, but are often fundamental to understanding program results—such as beliefs and perceptions, social relationships, administrative bottlenecks, or institutional dynamics. Qualitative methods do better at capturing these issues because they use more flexible questions, ask for open-ended responses, thoroughly explore the topic, and promote rapport between researchers and research subjects, which results in more candid responses. Observation methods independently confirm or contradict what people say. There is, however, a trade-off between depth and breadth, and smaller sample sizes in qualitative studies mean that findings are rarely statistically representative of a broad population. Quantitative and qualitative evaluation methods compensate for each other’s weaknesses, and each approach provides more value when used in a mixed-method design, providing information and conclusions that are more coherent, reliable, and useful than those from single-method studies.

Box 1 provides examples of issues that tend to be best addressed by either a quantitative or qualitative approach. Note, however, that categories represent relative strengths for emphasis, but not a dichotomy. Each topic can potentially be addressed in different ways by quantitative and qualitative methods, yielding different types of information. Furthermore, surveys can include questions with open-ended responses, and qualitative data can be quantified.

Most importantly, the approaches work in complementary ways to address a given issue. In the evaluation of the Child Support Grant in South Africa, for example, qualitative methods identified the full range of adolescent high-risk behavior and its economic and social drivers. These were turned into survey questions and responses carefully tailored to this program context. Furthermore, the focus groups tested a key assumption used to construct the survey’s control group, tested respondent stratification and recruitment strategies, and provided data that would later help with interpretation of survey data.

Qualitative methods explain process. Quantitative methods can determine, for example, whether dissemination through farmers’ organizations leads to increased adoption of an agricultural

<table>
<thead>
<tr>
<th>Box 1: Examples of Issues Normally Studied through Quantitative and Qualitative Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantitative</strong></td>
</tr>
<tr>
<td>- Household demographics</td>
</tr>
<tr>
<td>- Targeting accuracy</td>
</tr>
<tr>
<td>- Participation rates (for example, in training or services)</td>
</tr>
<tr>
<td>- Impacts (for example, on production, income, expenditure, employment, education, health, or nutrition)</td>
</tr>
<tr>
<td>- Intrahousehold decision making</td>
</tr>
<tr>
<td>- Service quality (for example, waiting times, availability of supplies, accuracy of rations, staff absenteeism)</td>
</tr>
<tr>
<td>- Test scores</td>
</tr>
</tbody>
</table>

Source: Author’s compilation.
technology. Qualitative methods will tell us about the social and political relationships that explain why different types of farmers join, and about formal and informal practices—factors necessary for understanding whether such organizations are likely to generate the intended outcomes under different circumstances. Such process issues “can be crucial to understanding impact, as opposed to simply measuring it” (Rao and Woolcock 2003). The study of process is also an important component of program monitoring. Studying what actually occurs during program implementation can determine whether failure to achieve intended outcomes or impact results from design failure or implementation failure (Bamberger, Rao, and Woolcock 2010).

Box 2 presents the wide variety of ways in which quantitative and qualitative methods support each other. Triangulation is a central function of mixed-method M&E designs: comparing quantitative and qualitative datasets to see how each one confirms, challenges, or explains the other. In the evaluation of the conditional cash transfer (CCT) program in Turkey, a survey was used to measure program impacts on attendance rates at school and health check-ups, while the qualitative research collected the full range of economic, political, and sociocultural explanations for attendance and lack of attendance at both. The survey found, for example, that the CCT program raised secondary school enrollment for girls by 10.7 percent, but enrollment rates were still very low at the secondary level: 38.2 percent for secondary school girls nationally, and lower in some regions. The qualitative study found that sociocultural beliefs and practices, and especially gender issues, frequently overpowered the financial incentive of the transfer. The issues hindering the success of the CCT program included the belief that women’s primary roles are as wives and mothers, the perceived lack of benefits of education, fear of girls’ sexuality and male advances bringing harm to family reputation and honor, exacerbated by inadequate transportation and location of schools that were perceived to put girls at further risk. The findings explained why the CCT was successful in some contexts and not others, and the need for complementary interventions (Adato 2008).

Single-method studies don’t have this ability to explain. In the evaluation of a CCT for nutrition in Brazil, the survey found a small negative effect of the program on children’s weight gain. The evaluators speculated, based on anecdotal information, that mothers might have deliberately kept their children underfed due to a mistaken belief that they would lose benefits if they gained weight (Morris and others 2004). Qualitative methods could have been used to test whether this explanation—or a different one—was likely to be correct. In another example from a study of domestic violence, a survey found a strong correlation between domestic violence and female sterilization. This would have been difficult to

### Box 2: Complementarities between Qualitative and Quantitative Methods

Qualitative methods support quantitative survey research through:

- Identifying hypotheses
- Suggesting or confirming the validity of identification strategies, constructs, proxies, assumptions, and instrumental variables
- Identifying and testing topics, questions, response options, and local meanings of language for surveys
- Identifying sampling and measurement error
- Determining the direction of causality
- Identifying unanticipated issues
- Making “unobservables” observable
- Triangulating, confirming, challenging, and explaining survey findings

Quantitative methods support qualitative methods through:

- Identifying topics and questions for investigation
- Identifying a stratification strategy and providing data for drawing a qualitative sample
- Providing background data on households, communities, and institutions
- Comparing profiles of qualitative sample communities or households with broader populations to determine degree of similarity or difference
- Testing the generalizability of qualitative findings
- Triangulating, confirming, and challenging qualitative research findings

Source: Author’s compilation.
explain without qualitative research, which found that husbands became more suspicious of their wives’ fidelity due to reduced risk of pregnancy, thereby increasing the risk of violence (Rao and Woolcock 2003). Surveys may be able to further test the generalizability of some of these findings, but qualitative work is necessary to identify these pathways.

Choosing among Methods

The most common qualitative methods used for M&E are focus groups, participatory appraisal, beneficiary/nonbeneficiary interviews, key informant interviews, and observation. These can be used in rapid appraisals, spending a day to several days in a locality or program delivery setting, or as part of extended case studies and ethnographic studies spending several weeks or months in one such location. Monitoring systems tend to use these methods through shorter data collection exercises at regular intervals. Which methods are selected depends in part on M&E budgets and time frames, but also on the purpose, stage of M&E at which they are used, and the types of issues to be investigated.

Focus groups and participatory appraisal tend to be best suited for broad identification of issues and preferences. They are frequently used at early stages of evaluation design to inform the design of surveys, though they can also be used at later stages to provide data on how well a program is functioning and why. A main advantage of focus groups is that a large number of people can be included in the study in a relatively short period of time, maximizing the diversity of experiences and opinions identified while minimizing costs. Another advantage is that a group discussion can stimulate recollection and debate. Participatory appraisal methods, which combine visual exercises with discussions, provide more control and benefits to participants. The main limitations of focus groups are that there is relatively little time to establish rapport and trust, or to investigate issues in depth, and it is difficult to link the information to other datasets. People may also be less willing to discuss sensitive topics in groups (though not always), such as domestic violence or HIV (human immunodeficiency virus), particularly with respect to their own experience. Finally, minority opinions or those of the less powerful may not be revealed: careful disaggregation of focus groups by categories such as gender, wealth, age, or ethnicity is essential for reducing this risk. Still, focus groups are useful methods for rapid, low-cost identification of issues and for assessing beneficiary or service provider perceptions and experiences.

In-depth interviews and observations can be used at any stage of M&E to identify issues early on and to gather data once a program is underway. These methods allow the field-worker to pursue a topic until it is well understood. People may be more willing to respond candidly in individual interviews, and observations enable independent confirmation. These data can then be triangulated and analyzed in relation to other individual and contextual data. In evaluation, ethnographic case study methods are sometimes used, where field-workers live for a period of time (for example, three to six months) in program communities. These methods permit the most reliable picture of program processes and impacts, by providing the time to: establish strong rapport and trust with program stakeholders, conduct iterative sets of interviews, and observe household, community, and program interactions and key activities over time. These methods are, however, more time and resource intensive compared to focus group methods, and sample sizes are normally smaller.

Key informant interviews are essential for M&E, gathering the knowledge of program officials and staff, service delivery professionals, community leaders, business owners, contractors, and other stakeholders. Key informant interviews provide information and analysis based on day-to-day observations of the program. Another mixed-method approach, particularly valuable for operations or process M&E, are systematic observations of service delivery, combining quantitative instruments to record and rate observed conditions and practices with qualitative interviews and observations.

Key Issues in Mixed-Method Designs

Sequencing of methods. Although sequencing can be done in various ways, a best practice evaluation design might look like this: The evaluation starts with qualitative methods to identify key issues and gather information to inform survey design. This is followed by the baseline survey. The survey data are used to design and select the sample for a
new stage of qualitative research and to identify issues for investigation—such as findings that need explanation. Following this qualitative study, an evaluation survey mirrors the baselines, but adds some new questions identified in the qualitative study. A new phase of qualitative research then examines impacts and investigates survey findings. Depending on the evaluation design, resources, and needs, additional rounds may follow. For program monitoring, a subset of quantitative indicators and qualitative data can be collected at regular intervals, maintaining common indicators but also adapting indicators based on new findings. Many governments invest substantial resources in monitoring systems that collect large quantities of quantitative data that reflect expected outputs. They typically do not explain the reasons for good or poor performance, which limits the ability to respond. Complementary use of qualitative methods in monitoring systems can help provide these explanations and identify unanticipated issues and outcomes.

**Site and household selection.** While some qualitative studies involve a convenience sample of locations or households, a more rigorous approach uses survey data to stratify qualitative samples. The qualitative samples would then reflect characteristics of the quantitative sample. Within these stratified categories, households or individuals are often selected *purposively,* to ensure inclusion of households across the distribution; if a random sample is used, the sample should be large enough to capture this distribution. In an evaluation of an agroforestry intervention in Kenya, for example, survey data were used to select a sample of households for qualitative case studies that captured Luo and Luhya ethnic groups, male- and female-headed households, richer and poorer farmers, and agroforestry early adopters, late adopters, nonadopters, and *disadopters* (Place, Adato, and Hebinck 2007). In evaluations of CCT programs in Turkey, Nicaragua, and El Salvador, survey data were used to stratify the qualitative sample between households where children performed well and poorly on the key education and health indicators targeted by the program. In this way, the qualitative research could investigate the conditions and characteristics that explained this different performance, with and without the program (Adato 2008).

**Data analysis and integration.** Many M&E systems that collect quantitative and qualitative data fail to take advantage of their synergies. For example, data from alternating rounds of surveys and qualitative research are not always used to inform the questions for the alternating next round. Even more common is the failure to triangulate and integrate the findings at the analysis stage—thus losing much of the principle analytic power of mixed-method designs. Under the typical time pressure to complete evaluation or monitoring reports, data are analyzed and reported separately. It is critical that data integration becomes a priority, and that the time and resources needed for data integration at the analysis stage are included in the budget.

**What Do We Learn? Findings from Mixed-Method M&E**

Examples of mixed-method M&E from four countries are outlined in this section, illustrating the different purposes, designs, and methods discussed above.

**Monitoring a food-assisted maternal and child health and nutrition program in Haiti**

An operations research approach was used for the M&E system for World Vision’s food-assisted maternal and child health and nutrition program in Haiti (Loechl and others 2005). The objectives were to assess the implementation of service delivery, identify constraints to effective operation, and implement corrective actions. Quantitative methods used were structured observations and interviews at program delivery points. The qualitative methods were semistructured interviews with stakeholders and focus group discussions with the program staff. The service delivery points were: Rally Posts, where targeting, health education and services, and growth monitoring and promotion took place; Mothers’ Clubs, where smaller groups of participants gathered to discuss health and nutrition topics; and Food Distribution Points, where beneficiaries received monthly food rations. Selected findings from the service delivery points include:

*Operations at the Rally Posts:* These were found to be operating as planned; however, problems identified included crowding, a high participant/staff ratio, long waiting times, bottlenecks at registration, and the lack of supplies and transport for staff. Improvements were needed in the general education sessions and the com-
munication between health staff and caregivers. Measurement errors were also identified in weighing and plotting children’s weight on the growth chart; this was a critical area because the growth charts were used for targeting children for recuperative action.

Mothers’ Clubs: These were found to be highly popular among health staff and beneficiaries. A new behavior change and communication strategy and new materials and techniques had been recently developed to improve infant and young child feeding practices. The mixed-method approach enabled an objective assessment of the technical content of the sessions and health staff’s facilitation and teaching skills, which were found to have improved. However, ensuring the intended composition of the clubs was identified as an ongoing challenge, and continued supervision and retraining of the staff was recommended.

Food Distribution Points (FDPs). Observations of the FDPs identified excessive crowding and long waiting times, delays in arrival of the food and staff, and the reasons for these problems: bad road conditions, limited transport facilities, and fuel scarcity. Exit interviews revealed that a large proportion of beneficiaries did not receive the amounts of food commodities they were entitled to. The sharing of food commodities among other relatives, neighbors, and others was reported to be widespread. This was determined to be inevitable, and it was recommended that an additional indirect ration be provided to cover this, and that the program should continue to emphasize the use of fortified commodities with micronutrient content targeted to beneficiaries, especially young children.

After implementation of the recommendations, a new round of operations research was conducted to monitor the corrective measures and document improvements in the program.

Evaluation of the conditional cash transfer program in Nicaragua
The evaluation of the CCT in Nicaragua (Adato 2008) involved baseline and follow-up panel surveys with 1,359 households, conducted in 42 administrative units (comarcas) with and without the program. Survey data were later used to stratify households by high and low performance in health and education indicators, with a qualitative sample drawn from each category. In total, 120 households were included in the qualitative study in Nicaragua. Field-workers lived in the study communities for approximately four months, conducting interviews about program experiences and impacts and people’s attitudes and behaviour, and observing meal preparation; health and hygiene practices; shopping; beneficiary and community gatherings; health service delivery; health and nutrition education, and other program activities. Some of the benefits of this mixed-methods design are outlined below:

Targeting. The survey found that the program was well targeted, with undercoverage rates of 3 to 10 percent. The qualitative research found, however, that people saw themselves as “all poor” and did not understand why households were selected into or out of the program, resulting in several types of stress and tension in the communities. This led to recommendations to improve program communications and to provide some limited benefits to nonbeneficiary households.

Iron supplements. The survey found a large increase in the percentage of children receiving iron supplements: from under 25 percent to nearly 80 percent. However, it found no impact on the high anemia rates in this population. In initial interviews in the qualitative research, mothers said that they gave the supplements to their children. However, over time, the case study methods revealed a different picture: mothers were picking up the supplements but not giving them to their children because of the perception that iron negatively affected children’s stomachs and teeth.

“Stuffing” children before weighing. In the first phase of the program, if children twice fell below an established rate of weight gain, benefits could be suspended. Although this policy was dropped, the study found that many beneficiaries did not know this, and that to avoid what they believed would be a loss of their benefits, some mothers were stuffing their children with food and liquids on the day or days leading up to the weighing. This revealed important information about poorly conceived incentives, as well as the impact of inadequate communications.

Program impacts on gender relations. Concerns have been raised that giving cash transfers to women could cause tensions with their male partners, possibly contributing to domestic violence. The qualitative research was able to explore this delicate topic, but found that men largely supported women receiving the benefit, because they saw the CCT program as for children and believed that women would spend the cash more wisely.
Furthermore, the new resources in households helped to ease tensions. It also found that the program’s discourse on women’s empowerment and women’s receipt of the cash increased their self-confidence and gave them some new autonomy in certain spending decisions.

Evaluation of high yielding maize varieties in Zimbabwe

This evaluation (Bourdillon and others 2007) used data from a panel survey conducted in resettlement areas from 1983–84, 1987, and annually from 1992–2000. The surveys contained extensive information on agricultural and nonfarm activities, expenditure, assets, and other impacts. Qualitative household case studies, focus groups, and key informant interviews were conducted during a six-month period of fieldwork in 2001. Findings include:

Gender relations and intrahousehold resource control. Despite a modest reduction in household-level poverty, benefits to men from high yielding varieties (HYVs) of maize undermined women’s control of resources. Whereas men operated within the public commercial markets for HYV maize, women preferred the open-pollinated varieties (OPVs), which HYV maize had displaced, because OPV seeds and maize were marketed through informal networks where women operated. Women also did not have access to credit for the commercial fertilizer necessary for HYV maize, but not for OPVs. Although money from the sale of HYV maize was called “family money,” the qualitative research revealed that “family money” was really the household head’s money, kept in his bank account. This “family money” was often invested in cattle (traditionally male property), and in one case study a woman explained her fear that if her husband died, his relatives would take the cattle away and she would be left with nothing.

The significance of age for extension approaches. The qualitative research revealed generational differences in how farmers value knowledge. Young people trusted the knowledge of the national extension service officers, viewing them as trained and experienced. In contrast, older people trusted their own experiences and demonstration units. Cultural values and beliefs attributed wisdom to age, and older men especially found it hard to admit to limitations in their knowledge, preferring their own “practical” knowledge to what they saw as “theoretical” knowledge.

Culture and magic. In the case study communities, there was widespread belief in the effect of witchcraft on crop performance. People frequently attributed magical powers to those who achieved unusually high yields, and poor yields to theft of crops through witchcraft. In one resettlement area, people would not show interest in the crops of others, because observing how others grew their crops could arouse suspicions of witchcraft. In another area, there was a widespread belief that implements or animals lent to other farmers could be returned bewitched. This has important implications for farmer-to-farmer methods of dissemination and extension, and the expectation that farmers “learn from each other.”

Final Remarks

Although mixed methods are widely used by governments and international agencies, there are a number of reasons why it is still common to find single-method approaches. The high cost of survey research means that decisions are often made to allocate an entire evaluation budget to a single approach. Second, timelines are often perceived as too tight for iterative rounds of data collection. Third, researchers are usually trained in one approach—quantitative or qualitative—and do not sufficiently understand or appreciate the methods and value of the other. However, mixed-method research designs can be adapted to fit a given set of conditions, and the benefits are likely to far exceed the costs. Still, it is important to recognize that the open-ended nature of qualitative research methods requires a considerable degree of skill on the part of field researchers to obtain quality data, and that sufficient resources are needed to ensure a strong research design, a sample size large enough to capture heterogeneity, adequate time for fieldwork, and the systematic analysis and integration of data. If both quantitative and qualitative research are undertaken with rigor, then mixed-method M&E will result in a far better understanding of program results than either approach alone. This level of understanding is critical to provide effective feedback that will improve performance and enable programs to meet their goals.

References

This note series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREM notes are widely distributed to Bank staff and are also available on the PREM Website (http://www.worldbank.org/prem). If you are interested in writing a PREM note, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREM note please contact the PREM Advisory Service at x87736.


Acknowledgment
For their comments, the author thanks Laura Rawlings (Lead Social Protection Specialist, HDNSP), and the following members of the Poverty Reduction and Equity Group: Philipp Krause (Consultant), Gladys Lopez-Acevedo (Senior Economist), Keith Mackay (Consultant), and Jaime Saavedra (Acting Sector Director). The views expressed in this note are those of the author. To access other notes in this series, visit www.worldbank.org/poverty/nutsandbolts.

About the Author
Michelle Adato is Director of Social and Gender Assessment for the Millennium Challenge Corporation of the U.S. government, which she joined in late 2010. At the time that she wrote this note, she was a Senior Research Fellow at the International Food Policy Research Institute (IFPRI). For eight years, she was co-leader of IFPRI’s Global and Regional Program on Large-Scale Human Capital Interventions, which specialized in program evaluations using quantitative and qualitative methods, and she is the co-editor of two books based on multi-country impact assessments. She has a Ph.D. in Development Sociology from Cornell University and an M.P.A. from Harvard University’s John F. Kennedy School of Government.

For Further Reading