New Challenges and Opportunities for Multi and Mixed Methods Research (MMMR) in the Evaluation of International Development Programs

Michael Bamberger
Independent Evaluation Consultant, OR, USA

ABSTRACT
The purpose of this article is to review potential opportunities for the application of Multi and Mixed Methods Research (MMMR) in the evaluation of international development programs. MMMR is used to cover the mixing, combining, and integration of techniques, approaches, concepts, language, and philosophies within a particular social science discipline; and the mixing, combining, and integration across different social science disciplines. In this article, multi-method and mixed methods research approaches are not distinguished. Rather than focusing in depth on 1 or 2 potential applications, the article discusses 6 areas of development evaluation (i.e., integrating new information technologies, strengthening experimental and quasi-experimental designs, evaluating complexity and equity, evaluating gender equality and women’s empowerment, and identifying unintended outcomes of development programs). These areas are used to illustrate the potential contributions that MMMR can make in major areas of development evaluation. The article concludes by identifying some of the challenges inherent in the application of MMMR approaches.

KEYWORDS
International evaluation; mixed methods research; multi-methods research

Although the term multi-method research has been around at least since the publication of Brewer and Hunter’s (1989) Multimethod Research: A Synthesis of Styles, as yet, there is no consensus on how multi-method research should be defined and what is the distinction between multi-method and mixed methods (MM) approaches (see, however, the International Journal of Multiple Research Approaches website; it offers related operational definitions of these terms). Although multi-method approaches are widely applied by researchers in many disciplines, more attention has been focused on MM research, which has become increasingly popular over the past 20 years (sometimes referred to as the Third Research Paradigm) as evidenced by the publication of the Journal of Mixed Methods Research, three handbooks (i.e., Sage [with two editions] and Oxford University Press), and many books. One proposed way to distinguish between multi-method research and MM is that, whereas multi-method research usually involves the mixing, combining, or integration of techniques, approaches, concepts, language, and philosophies within a research tradition, in contrast, MM research usually involves the integration of techniques, approaches, and so forth across research traditions. Furthermore, MM usually involves a focus on integrating quantitative (QUANT) and qualitative (QUAL) methods. However, these distinctions are probably too limiting. Many evaluations involve combining both QUANT and QUAL approaches within a particular discipline (e.g., using sample surveys to develop a typology for selecting follow-up case studies or focus groups), whereas, on the other hand, multiple methods are frequently used in cross-disciplinary studies (Bevan, 2009).

An important dimension that has received less attention concerns the need to understand the different ways that QUANT methods are used in different social science disciplines. For example, the econometric methods that are widely used in both macro- and microeconomics involve quite different assumptions than those underlying QUANT methods used in sociology, demography, ethnography, or public-sector research. For example,
when econometric analysis is used in pretest-posttest comparison group designs, the problem of missing variables is often addressed by making assumptions about time-invariance that tend not to be used in the same way in many other disciplines (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2011; Khandker, Koolwal, & Samad, 2010).

A recent development that opens up new horizons for multi-method research concerns the rapid growth of big data and data analytics that permits the integration of many different data sets into an integrated data platform (see the following section). This adds a new dimension because the use of many kinds of data analytics entails researchers making the assumption that different kinds of data that have been collected in different ways and via different assumptions can be integrated into a single data platform (Mayer-Schönberger & Cukier, 2014; Meier, 2015; Siegel, 2013). Frequently, these large, multidimensional data sets are analyzed using data mining methods combined with correlation analysis, an approach that is frowned upon by many mainstream evaluators (Bamberger, 2016b; Siegel, 2013). There are also differences in how the nature of data is assessed, and differences in the role of theory.

Data analytics also involve the use of predictive modelling based on Bayesian statistics that many evaluators view as challenging both conventional experimental designs, and the kinds of analysis that can be conducted (see following section). Bayesian predictive modeling analysts criticize experimental designs for being backward-looking and assessing causality based on baseline conditions that will be five or more years out of date by the time the post-test analysis is completed. Consequently, it is argued that, in a rapidly changing world, the findings are of little practical relevance by the time they are published. Also, the fact that the cost of data collection for randomized controlled trials (RCTs) is normally very high means that most statistical significance tests can only be applied for the total sample and it is not possible to disaggregate to examine operational differences between sub-populations (Rose, 2016). Some of these considerations relate directly to design issues, whereas others relate to both design and analysis.

So, big data, with its ability to lead to results that can be generalized to future project scenarios; and the analysis of large and diverse data, previously far beyond the scope of conventional evaluations, not only offer tremendous opportunities, but also provide challenges for development evaluation. Multi-method researchers can make a major contribution to building bridges between big data analysts and conventional evaluators.

Although most evaluators continue to favor either a predominantly QUANT or a predominantly QUAL approach, there has been a growing movement to recognize mixed methods and multi-method evaluations as a distinct approach with its own methodologies and epistemological positions. Mixed methods research and, by extension, multi-method research have been called the third path (Gorard & Taylor, 2004), the third research paradigm (Johnson & Onwuegbuzie, 2004), and the third research community (Teddlie & Tashakkori, 2009).

Although the term mixed methods (MM) is commonly used to refer to designs that combine elements of both QUANT and QUAL approaches (Bamberger, Rugh, & Mabry, 2012), recently, the term has been broadened to include designs that combine two or more social science disciplines, without requiring that both QUANT and QUAL approaches must always be included (Bevan, 2009). In the international development field, much of this discussion concerns the combination of development economics (which plays a dominant role in development evaluation) with another discipline such as demography, sociology, or ethnography. In many of these instances, researchers representing both disciplines will use QUANT and QUAL methods, but each set of researchers operates within a different paradigm, often using different ways to formulate and test hypotheses and perhaps operating with different epistemological and ontological assumptions.

So, for the purposes of the present article, multi-method and mixed methods research (MMMR) is defined as an approach that either combines two or more research designs or methods within a particular social science discipline, or that combines two or more designs or approaches across different social science disciplines. This article does not distinguish between mixed methods and multi-method approaches, and leaves it to others to decide whether there is a useful difference between the two. Instead, this article proposes that MMMR should be based on the following principles:

1. First, most (but not necessarily all) MMMR evaluations require a combination of QUANT methods that answer questions about *how much, how many, and for how long*, with QUAL methods that dig deeper to understand lived experience, observe processes of, for example, behavioral change or program implementation, and assess the quality of services and programs.

2. Second, the choice of methods is critical and must be based on a full understanding of the strengths and weaknesses of each method and how they complement each other.

3. Third, the systematic use of triangulation to compare and confront information from different sources goes beyond the conventional use of triangulation as a consistency check to obtain the best estimate for a particular indicator. In MMMR, the application of triangulation is much broader, recognizing that different individuals and groups have different perspectives and values and there might not be a single best estimate.
• Fourth, MMMR requires a careful understanding of how the boundaries among different social science disciplines are defined and how they are applied in a particular evaluation (Hesse-Biber, 2015). One of the most important contributions of MMMR is to understand how conventional approaches from different disciplines can be reconciled and integrated. This also requires an understanding of the political dynamics of an evaluation and whether one discipline is dominant (as is frequently the case) and how this affects the way in which the different research tools and approaches are integrated.

• Finally, many proponents of MMMR believe that one of its purposes is to give voice to poor and vulnerable groups. From this perspective, MMMR are allied with empowerment, social justice, and equity-focused evaluation (Bamberger, Segone, & Tateossian, 2016; Mertens & Hesse-Biber, 2012).

All of these principles are directly relevant to the evaluation of international development programs. In fact, the impetus for the promotion of many of these principles came largely from the international development field. A good example is the emphasis on giving voice to poor and vulnerable groups wherein much of the literature originated in developing countries.

**MMMR in the Age of New Information Technologies**

The world we live in today is more connected and interdependent than at any time in human history. The changes are dramatic and many have occurred within the past few years, so that most people, including researchers and evaluators, are still attempting to grasp the implications. For example, it is estimated that 90% of the world’s knowledge has been generated during the past two years, and due to increases in digitalization and web data, many new actors have become producers, owners, and consumers of data. Between 2005 and 2015, the number of Internet users has more than tripled from 1 billion to 3.2 billion, and more households now own a mobile phone than have access to electricity or clean water (World Bank, 2016, p. 2). This explosion of information and its many new applications for both industrial and developing countries is often referred to as the data revolution (Bamberger, 2016).

Many new sources of data are becoming available, often in near real-time. Many of these data sources such as Twitter, satellite images, and electronic financial transfers (e.g., Automated Teller Machines [ATMs]) are so large that completely different systems are required for data analysis, often, but not always, requiring access to very powerful computers. Big data not only are used to collect and organize data, but also provide new methods of data analytics for data mining, data synthesis, analysis, and predictive modeling, as well as new methods to display and communicate the findings through data visualization. Furthermore, data analytics also make it possible to integrate many different kinds of data into an integrated data platform, which can then be used to identify new patterns and relationships among variables that had previously been stored separately and rarely linked. Another important tool, which is not yet widely used in evaluation, is machine learning (Domingos, 2015). Computers are taught to identify patterns in large and complex data. Computing power makes it possible to test, in some cases, thousands of different patterns and associations. Machine learning also offers a radically new way to develop a conceptual framework for an evaluation. There are several implications for MMMR. First, it is now possible to dramatically increase the number of data sources that can be used in many evaluations. For example, satellite images, analysis of Twitter and other social media, and analysis of ATM and phone records, as well as previously inaccessible secondary data sources, can be incorporated into an integrated data platform. Second, the speed and low cost of data collection and analysis make it possible to use more powerful data analysis techniques. Third, many big data analysts argue, that in cases where access to large and diverse data sets is possible, conventional quantitative evaluation models based on RCTs and quasi-experimental designs might become largely redundant (Mayer-Schönberger & Cukier, 2014). One of the limitations regarding the practical application of randomized designs is the high cost of data collection. Consequently, power analysis tools are used to estimate the smallest possible sample size to achieve a required level of statistical power for the total sample (Mayer-Schönberger & Cukier, 2014). Yet, even if there is adequate statistical power to detect a given effect size for the total sample, there remains a limitation of these designs in that it is frequently not possible to conduct disaggregated analysis to compare program effects on different groups, because sample sizes become too small. Due to the lower cost of data collection, big data often provide access to data on the total population, permitting much more detailed sub-group analysis. It could be argued that this is a practical, rather than a theoretical limitation of RCTs, but in the real-world, there are probably very few situations in which it would be economically feasible to permit the significant increases in sample size required for disaggregated analysis. There are, of course, many situations, such as pilot projects, where it may be possible to accept lower power levels for exploratory disaggregated analysis.
Despite the important issues of sample size and the limitations on the capacity for disaggregated analysis, a major strength of RCTs is that they address the issue of causality and explain the processes through which changes have taken place. An important limitation of predictive analytics is based on using data mining to generate large numbers of correlations. But correlations do not equal causality. This suggests that experimental designs and predictive analytics can be considered as complementary, rather than as competing approaches to development evaluation (Bamberger, 2016a). The debate on the limitations, and also the important strengths of RCTs is, of course, much broader than is the question of sample size (see, for e.g., Byrne, 2009) but space does not permit a fuller discussion.

Big data analytics is based on a radically different approach to experimental designs. This begins with data mining, which raises the hackles of many evaluators, which computes thousands or millions of correlations to find patterns in the data (Bamberger, 2016). Often, data analytics begins without any set of hypotheses or theoretical framework and the results simply identify relationships between, for example, variables A, B, and C and outcome D. Researchers do not use these analyses to explain causality but only to present correlations. Predictive analytics are then used to develop models that facilitate prediction of how different groups will respond to an intervention, for example, a job training and placement program for juvenile delinquents (Siegel, 2013). The major difference in the philosophies and approaches of conventional evaluators and data analysts is that, although the former spend considerable time, resources, and energy to ensure data quality and avoid bias, the mantra of many data analysts is most data are biased and often of poor quality, but it does not matter because tomorrow we will have new data. Data analysts often operate in a world where they continually have access to new real-time data so that their predictive models can be continually updated. However, much of the data is biased and is often of poor quality, but some data analysts argue that this is compensated for by the huge volumes of data and the frequency with which new data are generated and where the validity of the predictions is continually updated and improved.2

So, some of the questions, opportunities, and challenges for the incorporation of big data into a multi-method strategy for the evaluation of international development programs are the following:

a) **What are key paradigms, approaches, methods, and assumptions that underpin the application of big data?**

What are the main differences between big data analytics and conventional evaluation approaches, and how easily can these be reconciled and what seem to be the most difficult differences to address? Bamberger (2016b) identifies the following as some of the major differences between the approaches of data analysts and evaluators: (a) **Data quality**: evaluators spend time and resources to ensure the strongest, unbiased data, whereas data analysts accept a lower quality of data (the kinds that are actually available) because the data will be continually updated and improved; (b) **Integrating survey data which are generated at quite long time intervals (many surveys are only conducted annually or less frequently) with real-time data that are frequently updated but which are usually biased**. How to recognize the different uses for each types of data—rather than assessing all data by the usually evaluation quality standards; (c) **Data mining**: much of data analytics is based on data mining where multiple, large data sets analyzed using thousands or sometimes millions of correlations to identify patterns. Data mining is usually carried out without any theoretical framework or hypotheses. Evaluators are very suspicious of an analysis that is conducted without any theoretical guidance to direct the choice of variables or to check whether key information is not included in the analysis; (d) **Prediction versus attribution**: Most quantitative evaluations attempt to use an experimental or quasi-experimental design where a control/comparison group is used to control for other factors that might explain observed differences between project beneficiaries and non-beneficiaries. In contrast, data analytics normally involves the use of predictive analytic models that are based on correlation analysis and the use of Bayesian analysis to identify the probability of different outcomes. Many of these analyses predict likely outcomes for different sectors of the population (who will succeed and who will fail?) but without any explanation of the mechanisms that are driving these outcomes.

b) **What are the key conditions necessary for the successful application of data analytics?** Can these conditions be replicated in the evaluation of development programs? Examples of these conditions include the following: (i) What is the availability of the kinds of large data sets that are often available in the United States and Europe?; (ii) Are real-time data available to continually update the estimates and predictions?; (iii) Do development programs provide periodic progress indicators that can be used to test and update the predictive models?; and (iv) Is it possible to generate through big data the kinds of indicators required to evaluate development programs? Ensuring the appropriate questions are addressed becomes particularly critical when addressing issues of social justice, women’s empowerment, and vulnerability/exclusion: (v) Do development programs have management structures with the flexibility to respond to constantly updated real-time data and predictions? Note that this latter question does not imply that development programs have
weak and inflexible management but rather that the nature of most programs does not permit rapid changes of implementation strategies.

c) **What are the specific opportunities and challenges for multi-method approaches?** There are at least two questions to address here. The first question concerns how easily the paradigms, assumptions, methods, and tools of big data and data analytics can be reconciled with conventional evaluation approaches. Evaluators see many fundamental differences of approach that will be difficult to reconcile. The second question concerns the special issues involved in using multiple methods and data sources in an evaluation. A fundamental approach of data analytics is to integrate many data sets into a single, integrated data platform. This means that many of the key assumptions and data collection methods used to collect data on sensitive topics are largely ignored. So, for example, questions on domestic violence, information on household expenditures, and sentiment analysis from the analysis of a Twitter feed would all be reduced to a common metric. These questions, and particularly Question C, offer the opportunity for multi-method evaluators to develop strategies for integrating multiple data sets, based on different data collection methods, hypotheses, and analytical tools into a multi-method evaluation.

It is important to recognize that many of the predictive models were developed for on-line advertising (how does the click-rate for a particular site vary based on the placement of the text, font size, and color); or for the analysis of customer data for large retailers (Walmart was the pioneer where extensive demographic data are available on customers and daily purchase records are available, and the power of predictive models can be tested and refined on a daily basis (Siegel, 2013). However, data analytics are also starting to be used in the United States for the evaluation of social programs in fields such as crime prevention, education programs for disadvantaged youth, and recidivism (Siegel, 2013). There are some promising results, but these areas are still very new.

In international development, big data are starting to be used quite extensively in areas such as research on development (e.g., the relationship between varying unemployment rates and internal migration and tracking movements of refugees; Meier, 2015), and support to emergency programs (mapping location of earthquake victims, mapping drought hotspots design, and tracking trends in food prices). However, the evaluation community, in contrast to the review and program communities, has been slow to adopt big data approaches (Bamberger, 2016b). An important consequence for development planning and evaluation is that better data are becoming available on difficult-to-access populations. An example is the recent Afghan Census that was conducted in a high-security country, where access to the field was both difficult and dangerous (UNDP/UNOPS, UNICEF, UN-Women, & WFP, 2016). The census combined an ongoing demographic survey, satellite imagery, other remote sensing data, urban data, and Geographic Information Systems (GIS) statistical modelling. Data analytics was used to integrate the different data sources into a common platform, which was then used to obtain the best estimates from this incomplete population coverage and often weak data.²

Digital technologies, such as mobile phones, Global Positioning System (GPS) mapping, remote sensors, satellite imaging, digital payment programs, the Internet-of-Things, social networking, and on-line social planning systems (to name but a few), have been used very extensively by communities, non-governmental organizations (NGOs), governments, and international development agencies, and the impacts have been dramatic.³ These developments are closely linked to the equally dramatic growth of big data and smart data analytics, which can provide individuals, communities, NGOs, governments, and development agencies with sources of data—that would have been unimaginable even a few years ago—to identify development needs; to provide early-warning signals on potential emergencies or crises; and to plan, implement, and evaluate development programs. The generation and use of big data are closely linked to the broader applications of digital technology, and it is likely that, in the near future, many of the data sources used for program monitoring and evaluation (M&E) will be generated through broader technology delivery systems rather than through the kinds of stand-alone M&E studies that are commonly used today. So, future M&E systems are likely to be closely linked to broader systems encompassing program identification, design, and management. The evaluation of rural electrification programs is an example where much of the M&E data are already generated through digital systems used for payment and quality control of power delivery systems (World Bank, 2014).

However, despite the dramatic development of these technologies, there are a number of challenges that can limit the extent to which the promise of strengthening development programs in general, and evaluation in particular, is fulfilled and, in particular, the extent to which these technologies promote a more inclusive social development framework in which benefits reach the poorest and most vulnerable groups and where these groups have an active voice in decisions affecting their futures and where they are able to hold governments and development agencies to account. The concern is that, unless adequate political, regulatory, and social controls are put in place, most of the benefits could be enjoyed by wealthier and more powerful groups who may monopolize control of many of these technologies. There is also a concern that new information technologies
will be used extractively by governments, large development agencies, and corporations, resulting in poor and vulnerable groups having less, rather than more, information and control over decisions and policies affecting their lives. These are all challenges and opportunities whereby MMMR researchers and evaluators can potentially make important contributions.

The Potential Contributions of MMMR to Strengthening the Design of Development Evaluation

Using MMMR to Strengthen RCT and QED Evaluation Designs

Many evaluators argue that where they can be used, RCTs or, failing that, a strong quasi-experimental design (QED) are the most powerful designs for determining the extent to which programs have achieved their intended objectives (“Does the program work?”). Although still only used in a small proportion of evaluations, the popularity of RCTs has increased steadily over the past decade (White, 2009). However, there is an extensive literature on the limitations of RCTs, particularly with respect to the ways in which they are currently used and interpreted (Bamberger et al., 2012; Byrne, 2009). There are a number of ways in which MMMR can strengthen RCTs and QEDs. There are a number of ways in which MMMR can strengthen our understanding of the factual to complement the conventional focus on the counterfactual (White, 2011). Many RCTs and other QUANT [quantitative] evaluations fail to study how programs are actually implemented and often fail to assess implementation failure (in contrast to design failure that is usually studied) as an explanation of why intended outcomes were not achieved. MMMR are well-suited to bring together the multiple methods required to describe and assess the often complex processes of how programs are actually implemented on the ground. MMMR can also promote an understanding of the broader social, political, and economic context within which a program operates. For example, White’s (2002) work on poverty analysis illustrates how MMMR can provide a deeper understanding on issues such as labor exchange in Africa, and how community values affect access of different categories of women to loans in Zambia. White advocates incorporating ethnographic studies to inform QUANT analysis—for example, demonstrating the influence of the mother-in-law on whether the daughter-in-law adopts nutrition practices introduced under the Bangladesh Integrated Nutrition Program (White & Masset, 2007).

Despite the important benefits of RCT/QED designs, there are a number of important challenges when using an exclusively QUANT approach (Bamberger & White, 2007; Byrne, 2009). These challenges include failure to understand the context within which a program operates, ignoring how the program is actually implemented on the ground (the “black box” problem), limitations of exclusively QUANT data collection methods to capture sensitive information and interview difficult-to-reach groups, mono-method bias, the challenges of vanishing control groups, and weak external validity (Bamberger et al., 2012). RCTs also have limitations for equity-focused evaluation because many of the processes of social exclusion are difficult to observe and require greater use of QUAL and participatory methods (Bamberger & Segone, 2011).

There are a number of ways that MMMR can contribute to addressing these challenges (sometimes called “RCT+” designs). First, MMMR can strengthen QUANT designs. For example, sequential MMMR designs enable each stage of the evaluation to strengthen the rigor of the following stage and to enrich the interpretation of the previous stage. Combining deductive (QUANT) and inductive (QUAL) hypotheses can integrate econometric analytical models (formulated before data collection begins) with the flexibility to identify and test new hypotheses as the context of the program becomes better understood (QUAL emergent designs). Construct validity can also be strengthened by combining QUANT and QUAL indicators to develop multi-dimensional indices of complex constructs such as empowerment, vulnerability, resilience, and well-being. Second, MMMR can strengthen the sampling design. Combining random (QUANT) and purposive (QUAL) sampling combines statistical representation with the ability to explore small numbers of high-value participants in more depth. MMMR strategies can also assess the coverage of sampling frames and help identify excluded groups. This is critical because whereas great attention is paid to the ability of QUANT designs to address selection bias within the sampling frame, little attention is given to the fact that sampling frames often exclude important sectors of the target population—frequently the poorest and most vulnerable groups (Bamberger et al., 2012).

Third, MMMR can strengthen the quality and validity of data collection. Conventional QUANT designs often encounter problems collecting information on sensitive topics such as domestic violence, power relations in households, community organizations, and other groups. Using QUAL process analysis techniques such as observation, process tracing, panel studies, focus groups, key informant interviews, and participatory group consultation techniques such as Participatory Rural Appraisal (PRA) can also open up the program implementation “black box” (White, 2011). A useful technique, particularly for QEDs that do not have access to good baseline data, is the reconstruction of baseline data by creatively combining and triangulating a range of different QUANT
and QUAL sources of data (Bamberger, 2009). Similar reconstruction techniques can be used to estimate unobservables (missing variables). For example, this approach can be used to determine whether women who successfully started or expanded small businesses using micro-credit had greater prior business experience or came from families that were more supportive of women engaging in economic activities.

Fourth, MMMR can strengthen the quality of data analysis and interpretation. MMMR data analysis can combine QUANT and QUAL techniques to dig deeper, integrating QUAL richness with statistical generalization. For example, during recent years, there have been important advances in the QUANT analysis of QUAL narrative data (see Neuendorf, 2017, Chapters 4 and 5). Where possible, MMMR involves an attempt to budget time and resources to permit systematic analysis of differences between QUANT and QUAL findings, or to explore unexplained results from the QUANT analysis. In many cases, a rapid return to the field to check on unanticipated findings can have a high value because outliers often provide important insights into different patterns of behavior or social organization (Brown, 2000).

Fifth, triangulation can capture multiple perspectives on complex outcomes and impacts and develop strategies to understand the reasons for differences. These approaches are widely used in empowerment and equity-oriented evaluations to give voice to a wider range of voices, including those of the most vulnerable groups, particularly women (Mertens & Hesse-Biber, 2012). The World Bank is one of many organizations that now systematically incorporate MMMR into program evaluations (see Holland, 2007 for a number of case studies). Two other interesting examples are the assessment of the effectiveness of Community Driven Development in Indonesia in reducing community-level conflict (Barron, Diprose, & Woolcock, 2011) and the assessment of the social impact of social funds in Jamaica (Rao & Ibáñez, 2003).

The Contribution of MMMR to the Evaluation of Complex Development Programs

As interventions become larger and more complex, the difficulties of designing an evaluation that explains processes of causal change rapidly multiply (Bamberger, Vaessen, & Raimondo, 2016; Forss, Marra, & Schwartz, 2011). Complex programs involve many different processes and require the combination of multiple analytical frameworks and data collection methods. Also, programs operate in complex and changing micro- (i.e., community or local level), meso- (i.e., middle level such as the district level), and macro- (i.e., national or international level) environments and with complex interactions among the different levels (Bronfenbrenner, 1979). Furthermore, interventions are affected by cultural, historical, political, legal, and environmental factors, all of which require different evaluation methodologies. Programs also evolve and change in response to the context in which they operate and, consequently, it is necessary to employ an emergent evaluation design with the flexibility to respond to these changes. As a result of these factors, outcomes (intended and unintended) result from multiple causal pathways that are often non-linear and non-proportional and where a given outcome can result from different combinations of factors.

Further complications result from the fact that programs adapt to how they are perceived and used by different sectors of the target population. Feedback from the first groups of beneficiaries to their neighbors and friends can dramatically affect how a program evolves. The author observed a school feeding program in Central America where some teachers allowed younger siblings not in school to also receive breakfast with the students. Word quickly spread and mothers brought more young siblings each morning until finally some programs were transformed from a school feeding program to a child nutrition program with local farmers donating food and fathers helping to build school kitchens. In other schools, where teachers did not offer breakfast to siblings, the program continued as a normal school breakfast program. Theory-based evaluation (Leeuw, 2016), which is one of the underpinnings of MMMR, helps to understand the context, which is one of the main dimensions of complexity analysis. The context relates to complexity concepts such as embeddedness and connectivity. Theory-based evaluation recognizes and provides tools to analyze multiple, long, and interconnected causal pathways. The evaluation of infrastructure interventions illustrates the practical utility of theory-based approaches, including process tracking, for the analysis of these complex causal pathways (White, 2011). No single evaluation approach can address all of these complexity challenges and there is increasing interest in MMMR designs combining different frameworks, tools, and techniques—often from different social science disciplines (Bevan, 2009). MMMR has a number of advantages for addressing the challenges of complexity, which include the following:

- **Delineating the evaluand** (what is being evaluated). Complex interventions are frequently difficult to describe because different funding and implementing agencies provide different packages of services, often implemented in non-standard ways. These strategies are difficult to monitor, involving both QUANT and QUAL dimensions that require a MMMR approach. Approaches such as realist evaluation (Pawson, 2013), and theory-
based evaluation (Bamberger et al., 2012; Funnell & Rogers, 2011) describe MM strategies to capture these multi-dimensional approaches.

- **Describing complex causal pathways and outcomes/impacts.** Complex causal pathways are often non-linear, non-proportional, and involve feedback loops. MMMR tools such as theory of change and case studies can identify the main causal pathways and describe how these operate for different groups and in different contexts. MMMR also help define multi-dimensional outcomes.

- **Context-specific embeddedness of the intervention.** Program implementation and outcomes are also affected by a wide range of interacting contextual factors making the effects even more complex to trace and requiring a MM approach. A combination of methods is required to assess the effects of each factor.

- **Behavioral complexity.** Understanding the behavioral processes through which many complex programs achieve their outcomes requires the combination of QUANT methods such as socio-metric analysis, social network analysis, and attitude surveys with QUAL techniques such as participant observation, content analysis of audio and video recordings, focus groups, and in-depth interviews (Bamberger, Raftree, & Olazabal, 2016). The evaluation of complexity programs is another area where MMMR can make a major contribution. This is particularly the case because it is widely recognized by evaluators and program managers that current evaluation approaches are not well suited to assess programs that are constantly evolving, where outcomes are affected by the interaction among multiple local contextual factors and where difficult-to-capture processes of behavioral change play an important role in understanding what actually happens on the ground, and the multiple factors that contribute to outcomes.

The Contribution of MMMR to the Evaluation of Equity and Gender Equality

There is an increasing awareness that international development cannot be adequately evaluated solely on the basis of aggregate indicators such as changes in gross domestic product (GDP), average household income, or country indices such as the Human Development Index (Kumar, 2013). Although these are important starting points, average progress on a set of aggregate indicators fails to identify sectors or groups that do not benefit from development or who might even be worse off. Studies by organizations such as United Nations International Children Emergency Fund (UNICEF; Bamberger & Segone, 2011) have revealed that there are many countries that have achieved significant increases in, for example, school enrolment or reductions in malnutrition or stunting, but where the gap between children from better off and poorer families has not decreased or might even have increased. A number of quantitative measures are now widely used to identify these discrepancies. One of the most common is quintile or decile analysis, which compares enrolment, access to services, malnutrition, and so forth, by income groups. Another method is public expenditure incidence analysis, which examines the distribution of public expenditure on public services by income group (Bamberger & Segone, 2011).

Although these approaches represent an important advance, there are still challenges to be addressed in understanding who benefits and who does not benefit from different aspects of development. One challenge concerns construct validity. Concepts such as poverty or vulnerability are multi-dimensional, but most of the quantitative indicators only measure a single dimension. A second challenge is that significant sectors of the poorest and most vulnerable groups are invisible and are not included in many surveys. In some cases, this is because the sampling frame misses certain sectors of the population such as families that do not own land, have a property title, or appear in the register of the internally displaced population or are not registered as refugees. In other cases, marginal groups might wish to remain invisible for fear of deportation or expulsion from the community. On a subtler level, many surveys only involve interviews of the household head or farm owner so that information might not be collected on, for example, the wife or female partner. This is important because there are many situations where the household consumes sufficient food and basic essentials to place them above the poverty line, but where intra-household allocation means that certain household members, for example, the elderly, women, or girls, are below the minimum acceptable level (Bamberger & Segone, 2011). Many quantitative methods fail to capture intrahousehold resource allocation. The fourth challenge is that many equity analyses are based on indicators of poverty, equity, or vulnerability defined by outside experts and do not involve consultations with the community on how they perceive poverty or equality.

There are a number of ways in which MMMR can contribute. Some of these ways are presented in the following sections.

- **Construct validity and the analysis of multi-dimensional concepts.** QUANT and QUAL methods can be combined to provide more nuanced measures of different dimensions. Take the example of poverty: surveys could be used to obtain quantitative measures of different dimensions of poverty (e.g., income, access to services,
education level). Qualitative scales or other indicators can be used for self-assessment on a poverty scale; participatory consultations, such as PRA or group consultations, can be used to understand how communities understand poverty; and participatory poverty maps can be constructed whereby community members rate other families on a poverty scale. Data analytics can then be used to create an integrated data platform that combines all of the indicators into a common metric. In cases where the researcher has access to big data, it is possible to obtain poverty indicators from a number of different sources, which are then combined and triangulated to obtain a multi-dimensional poverty indicator. Some of the indicators that are often used for this purpose include: records of ATM withdrawals, electronic financial purchases (e.g., seeds and agricultural inputs), analysis of Twitter and other social media to track changes in the number of words and phrases relating to poverty, and satellite images of the number of lorries transporting agricultural goods to and from the market.

Invisibility of the poorest and most vulnerable groups. There are a number of MMMR approaches. The problem of the incomplete coverage of the sampling frame can be addressed in several ways: intensive on-the-ground, house-to-house checks can be conducted in small areas to attempt to identify individuals, businesses, or households that were not included in the sampling frame; snowball sampling can be used where respondents in the original sample are asked to identify individuals or groups not included in the sample; observation and participant observation can be used to directly observe groups that might have been overlooked. Some of the same techniques can be used to track groups not wishing to be located, although the process is likely to be more difficult. In-depth interviews, group consultations, and participant observation can be used.

Community attitudes to poverty and vulnerability. Focus groups, social poverty maps, interviews with key community informants, and participant observation can all be useful techniques.

Intra-household resource allocation. This is often the most challenging dimension because detailed data collection might be required to document differences in consumption. A common QUANT approach is to obtain detailed consumption records, either to be recorded by the enumerator or to ask households to keep detailed records of consumption. Although these techniques work well for aggregated consumption, they are much more difficult to use to document how food and other resources are allocated among household members. Participant observation can be a useful technique if it is possible to gain the confidence and access to the household on a regular basis. Recently, smartphone apps have been developed, for example, asking a family member to take photos or videos of how food and other resources are allocated. In all of these examples, MMMR combines different indicators and methods within a particular discipline (e.g., ATM records and electronic financial transactions) with QUANT and QUAL indicators across disciplines (e.g., QUANT survey data and QUAL data from participatory group consultations).

Although space does not permit us to do justice to this important topic, it is important to mention that all of these issues are reflected, often in a more nuanced way, with respect to gender equality and women’s empowerment. Many agencies have traditionally relied on the collection of sex-disaggregated data on program outcomes to address gender inequities. However, gender research has clearly documented that the continued disadvantage of women on a broad range of development indicators can only be explained by the articulation of systems of social control that are supported by complex interactions among legal, economic, political, labor market, socio-cultural, historical, and psychological factors (Bowman & Sweetman, 2014). These mechanisms are subtle and cannot usually be captured by exclusively QUANT surveys or by a review of government and agency reports. This is an area where MMMR can contribute because gender analysis always requires a combination of QUANT and QUAL data that must often be collected at the micro, meso, and macro levels (Bamberger, 2013; Monitoring, evaluation & learning, 2014). MMMR triangulation is a valuable tool for confronting the value perspectives of different groups and for uncovering subjugated knowledge (Hesse-Biber, 2012). Interested readers are referred to the very extensive literature on the application of MMMR to the analysis of gender inequality (Bamberger 2013; Bamberger, Segone, et al., 2016; Bowman & Sweetman, 2014; Brisolara, Seig, & SenGupta, 2014; Brown, 2000; Monitoring, evaluation & learning, 2014; Hay, 2014; Hesse-Biber, 2012; Hesse-Biber & Griffin, 2015; Mertens & Hesse-Biber, 2012; Mulder & Amariles, 2014).

Some economists are beginning to incorporate QUAL approaches from ethnography and sociology to understand household and gender dynamics to strengthen the interpretations of their QUANT survey data. The previously cited examples of the mother-in-law effect in Bangladesh (White & Masset, 2007) and cultural factors excluding young women from access to loan programs in Zambia (Eriksen & Lensink, 2015), illustrate promising avenues for the application of MMMR.
The Contribution of MMMR to Identifying Unintended Outcomes of Development Programs

Experienced evaluators are aware that few development programs work out exactly as planned and, often, many have quite serious, unintended outcomes (UOs). Consequently, it is surprising that so many evaluations, including rigorous impact evaluations, fail to identify many of these UOs. Examples of UOs that are not captured by the evaluation include the following: (a) programs that promote women’s economic empowerment, which result in increased domestic violence from men partners; infrastructure investment programs intended to benefit low-income groups that are often co-opted by political insiders who buy-up property at below market prices before the start of the project; and road construction projects that result in disproportionate increases in traffic accidents to women and children. Theory-based evaluation is one of the key tools of MMMR and a well-articulated theory can identify causal chains and mechanisms through which UOs might occur. The framework should also identify and articulate a rival hypothesis (Carvalho & White, 2004) that explains why the program might not produce the intended outcomes (another way to identify UOs). On a broader level, all theory-based evaluations, and evaluations in general, should take into consideration contextual factors, which are also a major source of negative outcomes.

Sometimes, the failure to capture UOs results from underfunded evaluations that only allow researchers to spend limited time in the field, whereas in other cases, the reason is the lack of interest of evaluation clients to discover UOs. However, the important issue for the present discussion is that the failure to capture UOs often results from the methodological limitations of many evaluation designs, including supposedly rigorous RCTs and QEDs. RCTs are designed to determine whether there is a statistically significant difference in outcomes between the project and control group, but the design logic does not permit the detection of UOs because it is only intended to assess whether intended outcomes have been achieved. So, it is quite possible for an RCT to find that a project has been successful without being aware of UOs.

The situation is more complicated for theory-based evaluation designs such as theory of change (TOC). It is perfectly possible for a TOC to identify potential negative outcomes and some do this. However, experience shows that very many TOCs do not identify UOs. For example, a recent meta-analysis of 340 USAID evaluation reports revealed that only 15% reported on unplanned effects (Hageboeck, Frumkin, & Monschein, 2013).

There are a number of ways that MMMR can contribute to improving the ability of evaluations to identify UOs. First, diagnostic planning studies can help understand the social, political, and cultural context within which the program will operate (White, 2011). Most commonly, these studies only involve relatively short visits to project locations and meetings with key informants, but for larger evaluations, this might involve ethnographic evaluations wherein researchers spend weeks or months living in the communities (Barron et al., 2011). However, even a few well-planned days in the field before the evaluation design is finalized can improve the evaluation design. These studies help understand patterns of social organization and potential areas of conflict that might be exacerbated by the project. Building these into the results chains can identify specific areas and issues on which the evaluation should focus.

Second, process analysis can identify the mechanisms through which UOs occur. The observation process is greatly strengthened by a well-articulated TOC to help identify what to look for. It is also important to use flexible methods to identify processes and outcomes that had not been anticipated (Morell, 2010).

The two following studies illustrate different ways in which UOs were identified through MMMR designs. First, in their study of the effects of community-driven development on the levels of local conflict in Indonesia, Barron et al. (2011) spent almost a year reviewing the academic literature on the causes and types of conflict, conducting participant observation in a sample of communities, interviewing local experts and key informants, and reviewing local documentation on the levels and types of conflict. They were able to identify more than 60 forms of community-level conflict that were then built into the evaluation design. In the second example, the evaluation of a sexual health information campaign using Short Message Service (SMS) messages in Uganda included an RCT survey asking young men and women about their knowledge and attitudes towards the use of contraceptives and their risky sexual behaviors (Jamison, Karlan, & Raffler, 2013). It was found that risky sexual behavior decreased for both women and men. However, the follow-up, in-depth QUAL study revealed that as women became more aware of the dangers of risky sexual behavior, they refused to have sex with partners involved in high-risk behavior. An important UO was that men who were refused sex with their regular partners, sought sex outside of their regular relationship, thereby significantly increasing their high-risk sexual behaviors. This important finding was only discovered by using a MMMR design.

The main contributions of MMMR are to complement the conventional RCT and QED designs by strengthening the theoretical framework and by introducing new tools to collect contextual data, to better understand what happens during project implementation, and to study the behavioral processes through which intended changes
are expected to occur. These are important contributions that can only be achieved through the systematic use of MMMR designs.

Conclusion: Opportunities and Challenges for the Application of MMMR in Development Evaluation

We have used the term Multi and Mixed Methods Research (MMMR) to cover both research designs that combine different research methods within a particular social science discipline and designs that combine methods across different social science disciplines. In the latter case, conventional mixed methods almost always combine QUANT methods from one discipline with QUAL methods from a different discipline. However, it could be argued that this is not an essential requirement for MMMR, and it would be useful to explore further the different approaches to, for example, QUANT methods in different disciplines. For example, there are differences among the QUANT approaches used in econometrics, demography, and many branches of sociology.

This article does not distinguish between mixed methods and multi-method approaches. Rather, it illustrated how MMMR can enhance methodological rigor and the operational utility of international development evaluations by strengthening randomized and quasi-experimental designs, the evaluation of complex development interventions, the evaluation of equity and gender equality, and the identification of unintended outcomes. MMMR can also make important contributions to promoting social justice and giving voice to marginalized groups. A set of principles are proposed that could provide a framework and guidelines for the application of MMMR. All of these MMMR approaches are based on a set of principles presented earlier, where conventional evaluation designs do not work well. MMMR can also help identify unintended outcomes and address equity issues, both of which are frequently overlooked.

Before presenting these different examples of areas where MMMR can potentially contribute, the article begins by arguing that the rapid evolution of new information technologies, big data, smart data analytics, and Information and Communication Technology (ICT), will dramatically change how evaluations are designed and conducted. The fact that most of the evaluation community has not yet come to terms with these new technologies suggests that MMMR researchers might have the opportunity to take the lead in understanding the opportunities and challenges that these new technologies offer.

MMMR also offers important practical benefits for evaluators when working under budget, time, and data constraints. A judicious application of several complementary methods based on a well-articulated theory of change and using triangulation to increase validity might provide a more meaningful assessment of the evaluation questions than would be obtained by investing in a single method such as a sample survey. There are often possible trade-offs between maximizing sample size and accepting a smaller sample and lower power of the test so that some resources can be invested in complementary QUAL methods. One of the practical challenges when working with a relatively small sample is that it is usually not possible to compare statistical differences between sub-groups (heterogeneity). In these common situations, judicious use of case studies, focus groups, or purposive samples of high-value informants might help to understand causes and consequences of differences between groups at a relatively low cost.

Despite the many benefits of MMMR, there are a number of challenges that affect their wider application when evaluating international development programs. The use of multiple methods will often be more costly than the application of a single method. In addition to the direct financial costs, a MMMR approach will often require the use of additional or more experienced researchers, and the introduction of new research methods requiring additional training. Sometimes researchers might also feel threatened if asked to use data collection and analysis techniques with which they are not familiar, or which they might feel are less rigorous than the methods they normally use. Morgan (2014) also points out the practical challenge of collecting multiple sources of information and then finding that it is difficult to incorporate all of these sources into the analysis. These considerations and trade-offs can be particularly helpful for international development evaluation where access to data and other real-world constraints can be particularly challenging (Bamberger et al., 2012).

This article concludes by identifying some of the challenges facing the application of MMMR approaches. Although some of the greatest potential benefits come from inter-disciplinary MMMR strategies combining different disciplines, this involves a number of challenges. Different disciplines have different research approaches to, for example, hypothesis generation, evaluation design, data collection, and analysis. Perhaps most importantly, there are differences in terms of what constitutes credible evidence (Donaldson, Christie, & Mark, 2009). One of the challenges of introducing QUAL methods to QUANT-oriented agencies is that many QUAL research designs are not considered by many QUANT researchers to be sufficiently rigorous or professional. A final challenge
concerns the management of MM. Considerable time and effort is required for team-building and for the implementation and analysis of MM designs. Greater management flexibility will also be required to simultaneously manage, for example, a sample survey with precise sample selection and data collection strategies and an ethnographic study with an emergent design and changing strategies that adapt as more is learned about the area of study (Bamberger, 2016c).

MMMR has important potential benefits but also involves financial costs and organizational challenges. So, evaluators of international development efforts should be aware of both the costs and benefits when deciding whether and how to use MMMR. Evaluators and managers should also be aware of the hidden costs of relying on more familiar and perhaps more economical evaluation designs that might not capture important equity and gender issues and that might fail to identify important unintended outcomes.

Notes

1. These issues are discussed in two recent blogs authored by Bamberger (2016a, 2017).
3. See for example, the ICT-Works blogs, which document the extensive applications of these technologies in the development field (ict-works@inveneo.org)
4. See Bamberger, Raftree, et al. (2016), “The role of new information and communication technologies in equity-focused evaluation: opportunities and challenges,” Evaluation for a discussion of these challenges including why these technologies can lead to governments and donors adopting an extractive strategy whereby information on and about poor and vulnerable groups can be collected without their knowledge or involvement.
5. Digital Dividends identifies three sets of factors that have limited the achievement of the potential social and broad economic benefits that digital technologies offer: who controls digital technology, inequality of access and of resources to benefit from these technologies, and concentration and lack of competition. In this report, we will discuss a broader range of constraints.
6. The use of ethnographic techniques in Zambia found that although the program was intended to benefit younger women, many communities believed that if an older woman received a loan, it was not necessary to also give one to a young woman because it was assumed (wrongly) that funds from the first loan would be shared equally among all family members.
7. In sequential designs QUANT and QUAL data collection methods are introduced in sequence so that, for example, a QUAL ethnographic study can strengthen the design of a QUANT questionnaire, or a rapid QUANT survey can ensure that a sample of case studies is broadly representative of the total population.
8. For example, QUAL exploratory studies can strengthen the design of questionnaire, whereas a rapid QUANT follow-up study can often help interpret some unexpected findings in the QUANT data analysis.
9. For example, house-to-house visits can be made to identify schools, businesses, or certain population groups not captured in the sample; key informants and observational techniques, including participant observation can also be used to identify difficult-to-reach groups.
11. For example, the DFID theory of change for the design and evaluation of programs to combat violence against women specifically identifies factors that might negatively affect the success of a program (DFID, 2012). See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67336/how-to-note-vawg-1.pdf
12. This case was developed with Michele Tarsilla for a workshop at the European Evaluation Society Conference in Dublin in 2014.

References


