

Integrating Qualitative Methods into Investment Climate Impact Evaluations Subtitle

Alejandra Mendoza Alcántara

Michael Woolcock



WORLD BANK GROUP

Trade and Competitiveness Global Practice Group

&

Development Research Group

Poverty and Inequality Team

December 2014

Abstract

Incorporating qualitative methods into the evaluation of development programs has become increasingly popular in recent years, both for the distinctive insights such approaches can bring in their own right and because of their capacity to complement the strengths—and where necessary correct some of the weaknesses—of quantitative approaches. Some initial work deploying mixed methods has been undertaken in the assessment of investment

climate reforms, but considerable room for expansion exists. This paper summarizes some of the key principles and practices underpinning mixed methods evaluations in development, highlight some notable examples of how such work has been conducted (and the particular contributions it has made), and offers some guidelines for those seeking to increase the sophistication and utility of qualitative methods in the evaluation of investment climate reforms.

This paper is a product of the Trade and Competitiveness Global Practice Group and the Poverty and Inequality Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at amendoza1@worldbank.org and mwoolcock@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Integrating Qualitative Methods into Investment Climate Impact Evaluations¹

Alejandra Mendoza Alcántara
Trade and Competitiveness Global Practice
World Bank

Michael Woolcock
Development Research Group
World Bank

Key words: investment climate, regulation, impact evaluation, mixed methods, triangulation
JEL codes: B5, K2, M13

¹ This paper was written as part of the Investment Climate Impact Program. For detailed and helpful comments on an earlier draft we are grateful to David McKenzie, Audrey Sachs and Paul Shaffer (though of course we remain solely responsible for enduring errors of fact and/or interpretation). Email addresses for correspondence: amendoza1@worldbank.org and mwoolcock@worldbank.org.

1. Introduction

In recent years a growing number of impact evaluations have been conducted in the area of private sector development. Of these, the most commonly explored topics relate to small and medium size enterprises (SME) such as the effectiveness of training programs to improve performance and management, and the impact of microfinance and insurance on the expansion of agribusiness. The focus of this paper, however, is more recent work exploring the role of investment climate reforms in promoting the growth of the private sector, and the specific ways in which qualitative research—in combination with quantitative methods (i.e., mixed methods approaches) —can contribute to evaluations of such reforms. This paper provides a (partial) literature review on the use of mixed methods in international development evaluations and a framework that can be used to identify those types of investment climate interventions wherein the integration of qualitative methods can add significant (and distinctive) value.

Examples of investment climate reforms², which may be implemented at the national/subnational level and to all or specific sectors, include:

Business regulation: Reforms designed either to remove unnecessary or antiquated regulations, or to introduce new ones aiming to make it easier for businesses to get started and to operate securely. These reforms include activities such as business registration, business licensing (including sector licenses), construction permits, business inspections, regulatory governance, information and communication technology solutions for regulatory reform, and environmental regulation.

Competition policy: Reforms aimed at removing sector-specific constraints that affect market competition and support effective antitrust and competition rules in key sectors.

Investment policy: Reforms aiming to attract and retain investment, such as investment promotion activities, investment incentives, reforms to remove barriers to entry, and regulations to protect investment.

² Source: World Bank Group Investment Climate Department; see www.investmentclimate.org

Investment climate reforms may not always be amenable to (or even appropriate for) impact evaluation methods that have risen to prominence in recent years, such as randomized controlled trials (RCTs) and quasi-experimental methods.³ The nature of the interventions can involve challenges in identifying or distinguishing among different types of firms; reforms may be implemented at the national level or to specific sectors, making comparisons or construction of counterfactuals difficult (if not impossible); capturing changes in firm behavior as a response to reforms might involve a long causal chain of observable and unobservable processes. In the face of such challenges, the integration of qualitative and quantitative methods (i.e., using mixed-method approaches) can be a constructive, pragmatic response to the challenges of conducting impact evaluations: helping to identify appropriate counterfactuals, to assess the efficacy of causal links and processes, and to guide decisions about the wisdom of cancelling, continuing or scaling up an intervention, or replicating it elsewhere.

Although there are increasing calls to recognize the shared and complementary roles of qualitative and quantitative research methods⁴, for present purposes it is helpful to begin with their respective core strengths and characteristics. Where quantitative research methods include econometrics, experiments and mathematical modeling based on numerical data drawn from surveys and official accounts; qualitative research methods include case studies, process tracing, and ethnography, which draw on data derived from in-depth interviews, focus groups, participation, observation, historical documents, media images, and texts. However, qualitative data are increasingly being quantified. For example, due to recent advances in computing power, it is now possible to scan huge volumes of text to discern patterns with regards to (say) the changing incidence and meaning of key words⁵, and to incorporate questions capturing subjective perceptions into large-N household surveys. Similarly, some anthropologists also engage in experimental studies (Ensminger, 2000), while comparative political scientists often

³ An RCT (randomized controlled trial) randomly assigns a pre-determined fraction of eligible project beneficiaries to receive the project; such beneficiaries are called a treatment group. The remaining eligible beneficiaries do not receive the project, and thus comprise the control group. The difference in outcomes between the treatment and control group is the impact of the project. A quasi-experimental method constructs a comparison group when it is not possible (for political, ethical, financial or logistical reasons) to randomly assign groups.

⁴ The word “methods” is used here to refer to the various techniques used for conducting evaluations, including the design, collection, analysis and interpretation of data. For a more nuanced view, see Hentschel’s (1999) useful distinction between methods and data in development research.

⁵ Ravallion (2012), for example, analyzes the changing frequency with which the word ‘poverty’ has appeared in all printed English-language documents (that are available on-line) published since 1800.

use tables derived from Boolean algebra and fuzzy sets analysis⁶ to assess interaction effects in small-N studies of (say) the conditions under which revolutions or institutional transitions occur (see Thelan and Mahoney, 2011; and Goertz and Mahoney, 2012).

The more modest objective of this note is to provide an overview of how a systematic integration of qualitative and quantitative methods—what is frequently called taking a ‘mixed methods’ approach—can usefully enrich an impact evaluation, especially as it pertains to assessing the effectiveness of investment climate interventions. It proceeds as follows. Section 2 outlines the distinctive characteristics (and strengths and weaknesses) of qualitative and quantitative approaches, and argues for the importance of assessing empirical findings regarding an intervention’s efficacy against a theory of change. Section 3 considers how the unique insights drawn from qualitative methods can be integrated throughout the different stages of an evaluation. Section 4 provides a framework to denote when qualitative methods are especially useful, focusing on the context of investment climate and private sector development interventions. Finally, section 5 provides an overview of the most common qualitative data collection techniques and provides some concrete examples of how they have been deployed in investment climate impact evaluations.

2. Why are mixed-method approaches useful for impact evaluation?

The main rationale for the systematic integration of qualitative and quantitative methods is that both approaches complement the others’ limitations; this is particularly so with regards to the ‘breadth’ and ‘depth’ of information that together is needed to optimally describe and explain complex phenomena. In this way, integrating qualitative methods in impact evaluation helps reveal the ways in which different causal mechanisms—singularly or in combination—generate observed outcomes and thereby enable evaluators to assess the intervention’s theory of change; i.e., both whether and *how* impact is achieved (Bamberger et al, 2008). Table 1 below summarizes the key ways in which both approaches are used in data collection, design, and analysis and interpretation.

⁶ Such tables typically array a series of binary variables (i.e., coded 1/0) to assess when particular combinations of otherwise complex factors (e.g., the presence or absence of a financial crisis, dependence on natural resource rents) are associated with specific outcomes (e.g., transitions to democracy).

Table 1: Characteristics of Quantitative and Qualitative Methods as Used in Impact Evaluations

	Quantitative Methods	Qualitative Methods
Research Questions	Usually derived deductively (e.g., from knowledge gaps in the literature); seek to demonstrate ‘precise’ causal effect (impact) of x on y for relatively large populations; can also draw on qualitative insights to refine questions for specific contexts	Usually derived inductively (e.g., by refining questions as they emerge in situ); focus on process concerns—how outcomes were attained, how different types and combinations of mechanisms generated different outcomes for different groups
Data Collection	Use data collection methods such as surveys with closed ended questions; this standardizes but limits the depth and variability of the information that is obtained	Use data collection methods such as focus groups to capture in-depth, context-specific information; are also used to ensure that questions in surveys are worded and sequenced in ways that all parties understand (‘construct validity’)
Evaluation Design	Use techniques to diminish selection bias (and other confounding factors); seek to ensure representativeness and comparability of project and non-project samples to enhance quality of statistical inference (‘internal validity’)	Can help to discern and discuss issues that are ‘unobservable’ statistically (including identifying good instruments); weaknesses in ‘breadth’ and representativeness are compensated for by strengths in ‘depth’ and understanding of causal mechanisms
Analysis and Interpretation	Quantifies the magnitude of impact to try to determine <i>whether</i> an observed aggregate outcome can be causally attributed (probabilistically) to the intervention; but even the most ‘rigorous’ (‘well-identified’) analysis rarely provides warrant for inferring that similar results will obtain elsewhere (or at larger scale) (‘external validity’)	Is best suited to informing discussions regarding <i>how, why</i> and <i>for whom</i> a given intervention worked (or not); thus can help explain (and foster learning from) variation in outcomes and/or implementation processes, and usefully contribute to discussions about the possible generalizability of given findings to novel contexts, populations and scales of operation

Using qualitative and mixed methods can enhance the robustness of the underlying model of causal inference (i.e., improve internal validity) and thereby diminish the influence of various sources of bias (e.g., selection bias, by observing ‘unobservable’ factors shaping program

placement and participation) and measurement error (e.g., discrepancies in terms of how survey questions are understood by respondents and researchers).⁷ Results obtained from qualitative analysis may support the conclusions obtained from the quantitative research but enable researchers to go beyond the measurement of impacts and provide specific evidence of *how* impact was achieved and for whom—i.e., it can facilitate the exploration of variation across time, space and groups, by showing how local context characteristics and implementation dynamics interact. In a recent study from Indonesia, for example, even neighboring villages performed quite differently; a key factor shaping this variation was whether local leaders supported or resisted the project, even though these villages were participating in the same project being implemented by the same people (Barron et al, 2011).

For ethical and political reasons, for example, it may be important to identify not just average treatment effects associated with a given regulatory reform, but why some social groups (such as women, ethnic minorities, the rural poor) fare worse than others. Similarly, from an operational perspective, an intervention may be failing because one seemingly small but crucial component of an otherwise tightly integrated implementation chain is weak (e.g., users' perceptions of the implementing agency's trustworthiness), or because of an unanticipated non-project factor (e.g., girls refusing to attend school because their latrines are next to those of the boys, who taunt and humiliate them) that may be readily corrected (placing girls' and boys' latrines in separate places on the school grounds) is being overlooked. Absent such insights about these interventions, inaccurate conclusions regarding their effectiveness may be drawn and low-cost opportunities for learning may be lost.

In other instances, however, qualitative research might qualify or even contradict the findings emerging from quantitative approaches, in which case the research team needs to work together to resolve the anomalies; these deliberations, if done carefully, can serve to enhance the confidence the project team (and stakeholders in the reform process, including policy makers) has in the final conclusions and the policy implications to which they give rise (Woolcock, 2009; Rugh et al, 2011).

⁷ Quasi-experimental designs, for example, present the risk of selection bias due to unobservable factors that affect participation and outcomes which are neither easy to measure, are not known by the researcher or are time variant. Using qualitative methods enables researchers to identify potential instrumental variables or identify those time variant and invariant unobservable variables.

To summarize, the systematic combination of quantitative and qualitative methods helps evaluators to optimize the likelihood that their findings (and interpretations of those findings) will lead to accurate inferences about the effectiveness of interventions and the prospects of those interventions if implemented elsewhere (or at a larger scale). It achieves this primarily by using the strengths of one approach to offset the weaknesses of the other (Rao, 2002; Rao & Woolcock 2003). Specific instances where quantitative and qualitative methods can be combined in the evaluation process include:

- Generating hypotheses about an intervention's effectiveness from theory, experience and qualitative research and then testing their ability to be generalized with quantitative techniques.
- Identifying contextual factors, processes and causal mechanisms via qualitative methods and assessing them further via quantitative methods (e.g., Ludwig et al, 2011) and/or additional qualitative analysis.
- Applying quantitative sampling techniques to units of qualitative data collection, and/or findings from qualitative analysis and using them to inform the design of quantitative data collection tools (i.e., household or firm surveys).
- Using qualitative findings to see if they support, explain, qualify or refute quantitative findings regarding an intervention's impact (Ananthpur et al, 2014).

Qualitative methods are especially useful when the interventions to be evaluated increase in complexity (i.e., require many discretionary and face-to-face transactions, and are contentious⁸), when the 'context' itself is highly variable (and perhaps 'fragile'), when the quality and availability of existing data is poor, and when insights are sought on specific types of impacts on specific groups (e.g., the effectiveness of a project for ethnic minorities, informal firms or illegal immigrants, who may not be adequately represented in formal surveys). Qualitative methods can also be useful when evaluating small-N interventions such as regulatory reforms at the national level, or automation of procedures in one single agency.⁹

⁸ Thus delivering the mail is a 'simple' (logistical) task while promoting women's empowerment in rural Pakistan is a highly 'complex' one (see Andrews, Pritchett and Woolcock 2013).

⁹ Small N cases are those in which insufficient units are available to be assigned to comparison groups to get the sufficient statistical power to run an experimental or quasi-experimental design.

Even though the deployment of mixed-method approaches has been increasing in economic development impact evaluations, most notably in health, to date relatively few impact evaluations can be identified as truly using a mixed-method approach. For example, only three percent of 3ie's portfolio has used a mixed-method approach¹⁰, and neither J-PAL nor World Bank databases formally record whether mixed methods were used in a given evaluation. At present the systematic integration of qualitative and quantitative methods in the evaluation of investment climate related topics is rare. In the following sections we provide some examples of how qualitative methods have been deployed in each stage of the standard evaluation cycle. Although these studies did not use a systematic integration of methods, they are useful to showcase the value-added of qualitative methods in specific stages of the evaluation. We reiterate that, ideally, the most valid and useful findings are likely to emerge when both qualitative and quantitative methods can be integrated at different stages, enabling their systematic combination to exploit the strengths (and minimize the weaknesses) of using one method alone.

3. The purpose and value added of integrating qualitative methods at various stages of an impact evaluation

This section seeks to illustrate the purposes of a mixed-method approach and how qualitative data collection and analysis might be useful in each stage of an impact evaluation.

3.1 Purposes of a mixed-method approach

The examples below illustrate how qualitative analysis and data collection can be integrated and add value to each stage of an evaluation. When, how and to what extent they are integrated depends on the objective of the research and the type(s) of purposes the evaluation will serve. Four main purposes for a mixed-method approach are identified in the literature (see Greene, et.al. 1989, Shaffer 2013):

¹⁰ Better Evaluation Blog, August 2013. "Mixed methods in evaluation Part 3: Enough pick and mix; time for some standards on mixing methods in impact evaluation." <http://betterevaluation.org/blog/mixed-methods-part-3>

(a) Expansion: Extending the scope and breadth of the research. Qualitative methods contribute to assessing and explaining the causal mechanisms, the nature of the trajectories of change, and the conditions under which particular underlying factors shape observed outcomes.

(b) Development: Qualitative methods are used to help inform and refine the development of quantitative methods. It can do this, for example, by using ‘vignettes’ (short, familiar scenarios) within surveys to ensure respondents are answering subjective questions in comparable ways (Hopkins and King 2010), by ensuring the wording and sequencing of survey questions yields robust results, and by identifying (perhaps unexpected) instrumental variables to be used in econometric model specification.

(c) Triangulation: Different methodological approaches assess the same research questions from a range of vantage points to confirm, modify or contest the initial results. Triangulation contributes to the enhancement of internal validity and minimizes biases stemming from using a particular method.

(d) Complementarity: Even when the empirical findings derived from different methods align, qualitative and quantitative perspectives can contribute to a more comprehensive *interpretation* of the results – what they mean, and what their implications are for policy and practice (Shaffer, 2013).

3.2 Integrating qualitative methods at each impact evaluation stage

Hypothesis formulation and theory of change

Evaluators usually develop hypotheses and specify causal mechanisms based on literature reviews, existing theories or field experience. Hence, based on deduction, evaluators develop hypotheses from a “top-down” approach (Bamberger, 2012). When using qualitative analysis the development of hypotheses is often made inductively, starting from specific observations to build general theories or propositions. Integrating both approaches may lead to a more accurate theory of change and understanding of the underlying mechanisms. Put differently, an iterative dialogue between qualitative and quantitative approaches can help refine the initial hypotheses, and make necessary adjustments along the way in light of new or different evidence (Clark and Badiee, 2010).

Giving inadequate attention to changing circumstances and the possibility of non-linear impact trajectories can lead to claims about impact that turn out to be premature, and that form an inaccurate basis for future projections. For example, a study that evaluated the impact of an export promotion-matching grant for SMEs in Tunisia found that in the short term, beneficiary firms showed higher export growth and export diversification than those of the control group. However, in a subsequent study it was found that the effects were not sustained over time, an issue that the authors highlighted as commonly overlooked in the literature (Cadot et al, 2012).

The authors of the follow-up study even mention that these types of reforms have not been explored in the long term, questioning the sustainability of what in the short-term was found to be “successful” (Cadot et al, 2012). Ravallion (2008) warns that the assessment of short-term impacts is common in impact evaluation, generating a “myopia bias” that can lead not only to erroneous conclusions but also to decisions to scale-up policies and programs without knowing the underlying factors of impact that can lead to negative spillovers. This dynamic can be seen in a World Bank-supported land reform project in Cambodia, which was hailed (rightly) as an initial success. But an inadequate theory of change caused by a poor understanding of the social, economic and political norms led to a mismatch between the reform’s expectations and the actual capacity of the administrative system to implement on a larger scale, especially in sensitive peri-urban areas. This generated stress on the demand side and weakened (almost collapsed) the capacity of the system (Adler et al, 2008, Biddulph, 2014). Hence, generating in-depth contextual information is key to identifying the factors that are shaping the nature and extent of an intervention’s impact trajectory (see Box 1), and to informing decisions about whether, when, where and how the intervention might be scaled up (or shut down, for that matter).

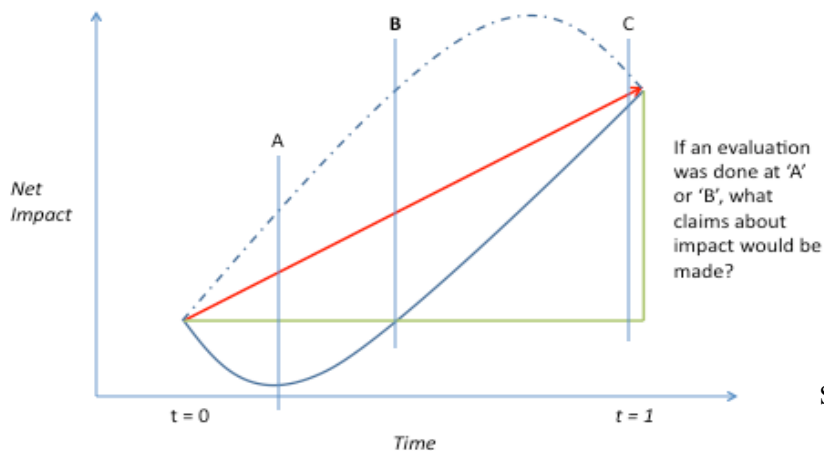
Box 1: Understanding Impact Trajectories

Four months after planting, we do not conclude that the growth of oak trees (which takes years) is ‘less effective’ than the growth of sunflowers (which takes weeks) because science and experience tell us what it is reasonable to expect by when. The same logic should apply to development interventions. The important implication is that when assessing an intervention at two points in time, evaluators must have (or build) a solid theory of change – on the basis of experience, evidence or theory – to specify the mechanisms (processes) by which they expect given inputs to generate observed outcomes, and over what time-frame and trajectory it is reasonable for these outcomes to emerge (Woolcock, 2009; 2013). Both qualitative and quantitative methods are needed to do this well. (Most complex to assess of all, of course, are those interventions that have no consistent impact trajectory.)

A central issue for both causal inference and policy extrapolation is that methods per se, no matter how ‘rigorously’ and comprehensively they are applied, do not on their own provide a clear basis for discerning whether an intervention is working or is likely to do so in the future; for that, the empirical findings must be guided by theory and experience. In short, *the implications of evidence are never self-evident.*

Consider the figure below, which exemplifies four different impact trajectories and three different points in time at which an evaluation could be conducted: without knowledge of the likely impact trajectory associated with a given intervention (say, roads versus schools versus immunization versus land titling), and thus knowledge of what it is reasonable to expect by when, wildly inaccurate conclusions regarding the intervention’s efficacy could be drawn. If the intervention was evaluated at point C, a fortuitously consistent story would emerge since all four trajectories converge on a similar net impact between ‘baseline’ ($t=0$) and follow-up ($t=1$). (And the timing of the follow-up is largely determined by political and administrative imperatives, not scientific ones.) But if the intervention was evaluated at point A, four very different conclusions regarding the intervention’s net impact – ranging from spectacular success to dismal failure – would be drawn, even if the intervention was being assessed via an RCT. The shape of the trajectories, when extended into the future, has correspondingly important implications for the claims we make about the intervention’s likely impacts down the road.

Understanding impact trajectories



Source: Woolcock (2013)

Hence, one important question that arises is: *when* should impacts be measured? By using qualitative methods to understand the context and by drawing on a range of experiences elsewhere, evaluators can derive informed knowledge of firms' behavior and motivations, and thus help to more accurately specify what outcomes the intervention can be expected to generate over a given time-frame. Integration of methods at this stage might prove beneficial for investment climate reforms given that assessments such as the Doing Business indicators or the Enterprise surveys can identify potential areas of reform. However, few details are provided regarding what and why firms need these reforms. Hence qualitative work can contribute to refining the intervention.

Evaluation design

Qualitative analysis and data collection can complement quantitative techniques in the evaluation design to address common challenges such as identification and inference of causal relations:

a) Identification: Qualitative data collection and analysis can be helpful in informing and selecting samples (whether of people, places or issues) of interest. For example, in-depth interviews or focus groups might be used to identify firms or individuals with “entrepreneurial” behavior, or to identify what constitutes entrepreneurial behavior according to the context and prevailing social norms. Once firms are identified, quantitative methods can be applied to the population of interest to make the sample (more) representative. Another common identification strategy is selecting samples (or even stratified samples) of interest from the sampling list with specific characteristics; qualitative research can then be conducted on those selected individuals or units of interest to help explain common or different characteristics, or to explain variance or outlier behavior (Tedlie et al, 2007). This technique is particularly useful when sample sizes are small.

Qualitative data collection methods have also been useful in refining the identification strategy and diminishing the risk of selection bias, especially for quasi-experimental studies where it is difficult to control for unobservable variables. An example is Bloom et al (2013), who assessed the impact of management practices in firms' performance in India by conducting retrospective interviews and observation assessments at the factories of a representative sample of firms. Data gathered was used to confirm that there was no significant difference between the project and

non-project firms. This study shows the importance of integrating both qualitative and quantitative sampling techniques to obtain a sample representative of the population with the specific desired characteristics. Such quantitative sampling techniques help to ensure that qualitative samples meet standards in being adequately representative, and contribute to ensuring that claims regarding the implications of these findings for wider populations are well founded.

b) Tailored concepts: Qualitative analysis may be useful to explore the dimensions of the indicators used in the design. Definitions of concepts such as ‘corruption’, ‘justice’ or ‘transparency’ may vary widely across individuals, locations or sectors. Exploring the meanings of indicators according to the context and incorporating them into the quantitative data collection methods is not only critical to obtain accurate data from surveys, but also plays a key role in establishing and explaining causation.

As an example, the concept of ‘delay’ in clearing goods in a border post may have different meanings depending on the sector and for people working at different points in the distribution channel. For importers of ultra-fresh products, a ‘delay’ might be understood as more than one day, while for other sectors (e.g., processed food), it might represent more than three days. Also, the definition may vary per location.

Hence, understanding the dimensions of the indicators in their context (and for different personnel within a given context) is necessary to understand what is intended to be measured. Rao and Woolcock (2003) describe how a survey on the incidence of domestic violence in India generated rates far below expectations. Initial survey results suggested that the incidence of household violence in India was even lower than in the US, but when researchers conducted qualitative analysis of this issue they found that domestic violence was understood differently relative to the context (e.g., a slap would not be considered as domestic violence by the average Indian household). Hence, the survey questions and results were inaccurate and did not reflect an accurate domestic violence situation. Even though quantitative approaches can be applied to measure changes in these outcomes, understanding the definitions of concepts is key to establishing valid quantitative measures for these concepts (i.e., to ensuring high ‘construct validity’).

A mixed-method approach can also contribute to incorporating different dimensions of particular indicators (Shaffer, 2013). Standard quantitative measurements of poverty such as consumption per capita, for example, can be weighted according to local or contextual definitions or perceptions of what ‘poverty’ means (Kristjanson et al, 2010).

c) Causation and model specification: By having detailed knowledge of a particular context, qualitative work can be helpful in solving endogeneity problems¹¹ and can reveal the direction of causality by identifying instrumental variables (Ravallion, 2000; Rao and Woolcock, 2003). (Some qualitative researchers also argue that techniques such as process tracing can be used to make causal claims of their own—e.g., Bennett (2010)—and note that case study evidence is routinely the basis on which causal arguments are made and defended in ‘real world’ settings such as court rooms—Honore (2010). However, we shall not address such matters here.)

d) Quality and reliability of data collection: Understanding the context through qualitative analysis is not only useful with regards to knowing what should be assessed in a survey or what should be included in an equation. It also contributes insights as to how and to whom questions should be asked or assessed, given that the quality of the data obtained depends on the collection methods used with specific objectives in specific contexts. As an example, Sana et al. (2012) conducted a study in the Dominican Republic and found that respondents answered differently depending on the type of questions asked by type of interviewer (local or external). They found that respondents reported higher income and higher tolerance towards marginalized groups to external interviewers compared to the responses given to local interviewers. Hence, qualitative methods can help to improve the quality of the data by exploring the best ways in which a question should be asked, how and to whom it should be asked, and by whom.

Parallel qualitative data collection techniques such as those generated by participant observation or case studies can also help to assess the reliability and quality of the data collected through surveys. The IFC Lima Tracer Study, for example, which assessed the impact of firm formalization on the performance of micro firms in Lima, found significant divergence from survey responses when the team conducted in-depth interviews to try to understand the low

¹¹ In evaluation, endogeneity problems stem from biased estimates of impact due to issues such as omitted variables or measurement error, which weaken the claims of attribution. In principle, experimental designs greatly reduce these problems by ensuring that any such biases are at least equally present in the treatment and control groups.

demand for operating licenses. Researchers explain that this may happen because “questions involving a moral issue, such as complying with the law, tend to be answered ‘correctly’, but not necessarily honestly” (Alcazar et al, 2011).

Intervention

Qualitative data collection and analysis generally ask and answer different questions from quantitative approaches (when the aim of the mixed-method approach is not triangulation), in the process uncovering other factors that may be shaping observed impacts such as the institutional framework (i.e., formal laws and regulations, and informal customs and norms). Contextual analysis contributes to assessing the institutional capacity of local agencies involved in the project (i.e., financial resources, political support, power of implementation) the political economy and the forces supporting (or undermining) the reform, and so on. These factors, which are difficult to measure quantitatively, may influence the quality of implementation and outcomes/impact. Qualitative data collection assessing the process of implementation can provide insights of how and why outcomes and impact were achieved. One criticism of conventional impact evaluations is that when expected impacts are not found, given that there is a lack of process evaluation or monitoring, it cannot be inferred if the absence of impact was because of failure of the design/causal link or the failure of implementation (Bamberger et al, 2010; Ananthpur et al, 2014).

Qualitative methods can be especially useful with regards to assessing the process and quality of implementation. For example, in implementing competition reforms, it has been found that larger impact in selected outcomes is achieved when effective enforcement is implemented (Kitzmuller and Licetti, 2012). The implementation of effective enforcement could be analyzed starting from the political context through the analysis of secondary data such as newspapers, by conducting direct observation or process tracing.¹² Results obtained from qualitative data collection methods may be transformed to variables that reflect these issues and can be incorporated into the econometric study, or they can be used in parallel to explain quantitative results.

¹² Process tracing is a tool of qualitative analysis that contributes to drawing descriptive and causal inferences from diagnostic observations undertaken chronologically (Collier, 2011).

Data analysis and interpretation

Qualitative analysis can contribute to internal validity by verifying the connections between the causal mechanisms identified in a quantitative analysis. (Similarly, if its findings are contradictory, it may provide an alternative explanation or lead to further research.) As an example, in an evaluation assessing the demand for formalization among firms in Sri Lanka, researchers wondered if the large shifts in profits that few firms reported were attributed to formalization or were due to measurement error (De Mel et al, 2012). The researchers conducted case studies to ensure that the findings were not driven by measurement error and to articulate the mediating channels through which formalization helped the firms that benefitted most. The qualitative analysis supported the quantitative findings and confirmed the causal mechanisms demonstrating that formalization led to increased firm profits. The qualitative analysis shed light on how formalization helped firms by allowing them to issue receipts and thereby become suppliers in larger value chains—in a very effective way.

Another relevant example is again the Lima Tracer Study, in which researchers used as baseline data firms operating without a license and used incentives such as fee waivers for the treatment group. The analysis found no significant impact on outcome variables. In addition, it was noted that firms were not eager to take the incentives. Through a qualitative study applied to a smaller sample, it was possible to distinguish behavioral characteristics of entrepreneurs associated with license acquisition. Information obtained through in-depth interviews revealed that there are two distinct groups among the entrepreneurs—“typical entrepreneurs” and “survival entrepreneurs”—and that this distinction may be considered a determinant in the decision to obtain a license. In addition, managers from micro firms did not perceive important benefits from formalization and recognized that the cost of the license is a real barrier for the formalization process, but not the most important. These interviews led to the conclusion that, in fact, there is not a high demand for operating licenses, an issue that was not captured through surveys, which also explains the low take-up and impact obtained.

The qualitative analysis was not initially contemplated; the original design was mainly a quantitative approach. As many companies did not accept the incentives, the research institute (GRADE) decided to conduct an in-depth study with a qualitative focus. Given the insightful

findings obtained from the qualitative analysis, GRADE started using mixed methods in its impact evaluations. The most common design now used is to initially conduct a qualitative study to understand the context and develop the questions for surveys and find insights regarding the outcome variables that should be taken into account. After the quantitative analysis is conducted, a second qualitative analysis is used to explain or dig deeper into the results found.¹³

3.3 Examples of mixed-methods in evaluation

One example that illustrates the iterative systematic approach is an assessment of the Kecamatan Development Project (KDP, a national community-driven development program) in Indonesia on local conflict dynamics (Barron et al, 2011). KDP's objective was to provide block grants to local communities, who would then allocate this money to those projects community members themselves deemed most pro-poor, sustainable and cost-effective. This allocation process took place in community forums, but not every proposal was funded, generating the potential for conflict if villagers perceived that outcomes were a function of non-merit-based procedures (or worse). The evaluation's objective was to assess whether and how these forums improved local governance; the hypothesis was that participating in KDP creates robust civic spaces and deliberative skills which enable local conflicts to be constructively addressed. One major challenge was that 'conflict' is notoriously hard to measure, and what little data there was had been collected from village leaders (who had obvious incentives to under-report the incidence of conflict on their watch). A mixed-method approach was used to find a novel way to measure conflict (which included a comprehensive analysis of local newspapers) and the mechanisms by which it is initiated or resolved (discerned via key informant interviews). In addition, it was critical for the evaluation to understand the causal chain of events, which was only possible with a deep qualitative analysis (which was generated by collecting dozens of cases of conflict pathways in program and comparable non-program villages).

An iterative strategy for integrating the quantitative and qualitative analysis was used. An initial period of qualitative fieldwork was pursued for three months. The villages were selected using a quantitative sampling frame [using propensity score matching (PSM) techniques derived from nationally representative household surveys], but the final selection of the best match of program

¹³ Information obtained from a telephone interview with Lorena Alcazar, November 2012.

and non-program villages was made using detailed contextual knowledge (since a well-understood weakness of PSM is that it only matches on ‘observable’ characteristics). This was critical to capture heterogeneity of the population and increase the validity of the results. This initial work contributed to the sampling of districts, research hypothesis formulation and design of adequate survey questions. Once the identification of a “counterfactual” was done using qualitative analysis and supported by quantitative methods, data was collected from a survey administered to a larger sample of households and used to assess the generality of the hypotheses emerging from the qualitative work. In addition to the quantitative analysis, the analysis of case studies of local conflict, interviews, surveys, key informant questionnaires and secondary data sources as newspaper evidence, provided a broad range of evidence to assess the validity of the hypotheses stating the conditions under which KDP could (and could not) contribute to solve local conflict.

Another common situation in which the usefulness of mixed methods can be seen is small N evaluations, such as the introduction of a business regulatory reform at the national or sub national level. Such reforms, by their very nature, make the construction of a counterfactual difficult or even impossible. In such circumstances, a process of elimination can be deployed to systematically identify and rule out alternative causal explanations of observed results. For example, firm performance could be attributed to the improvement of the business climate but this could be happening in ways unrelated to the actual business entry reform, such as via improvements in infrastructure or more information being available on business opportunities. A thorough qualitative analysis of the processes by which positive outcomes were attained could enable one to establish a detailed causal chain and define how the specific context interacts with the reform and outcomes. Quantitative approaches can be used in parallel for triangulation purposes, or can contribute by helping evaluators avoid some of the typical biases associated with qualitative analysis (such as selection bias), including selecting firms for in-depth analysis using randomization or purposive sampling techniques.

4. When is it useful to integrate a qualitative approach in investment climate evaluations?

This section seeks to identify when the systematic integration of qualitative and quantitative analysis is useful for evaluating investment climate interventions. As noted above, the distinctive contribution of qualitative data collection and analysis is its capacity to explore the role of local contexts and implementation dynamics in shaping variation in outcomes, and in interrogating how, why and for whom an intervention is working (or not). A mixed-method approach can potentially add value to all types of impact evaluations, but the importance and utility of this contribution increases with the complexity of the project and the diversity of the context within which the project is being implemented: assessing the economic impact of a business registration one-stop shop (streamlining of procedures) is hard enough, but even harder is a women's empowerment project, a firm formalization program, or a business regulatory reform initiative in a fragile state (see Bamberger, 2011; Woolcock, 2013).

To identify which type of projects might benefit most from the systematic integration of qualitative data collection and analysis, it helps to outline a description of the sources of complexity relevant to investment climate interventions and to disentangle them by (a) the outcomes/impacts expected and (b) the nature of the intervention. The relevance and role of qualitative techniques in the evaluation should be positively associated with the number of issues/sources of complexity an intervention has.

a) Outcomes/Impacts expected

- Investment climate interventions often seek a *behavioral change* from the target population. The trajectory of change for altering behavior is unlikely to be linear: for example, it may take many months (perhaps years) to convince beneficiaries to adopt a new practice (say, registering a small business) but once started by influential community leaders it may catch on quickly.¹⁴ In order to establish a theory of change connecting inputs with (hoped-for)

¹⁴ This phenomena is commonly observed with agricultural interventions: farmers may be slow to adopt new seeds and fertilizers that will clearly increase their yields (because they are highly risk averse), but once influential community members make the shift then others rapidly follow (on this see the classic work of Rogers 2010). Qualitative research can thus play a crucial role in enhancing both analysis and project effectiveness by helping identify who these influential community leaders are.

outcomes, evaluators must be aware of the factors underlying behavior and capture sources of influence: beliefs, perceptions, cultural, political and social norms. These determinant factors, which are often highly idiosyncratic, are not easily captured by quantitative data collection methods such as surveys with close-ended questions. The comparative advantage of qualitative data collection methods such as in-depth interviews, focus groups or observation assessments is that they are designed to explore contextual issues, motivations, social norms, political economy, and so on. For example, to establish a theory of change regarding changes in saving practices or increase in bank accounts registration, it is necessary to assess how individuals perceive their local banking system. Regardless of the intervention or type of incentives, if individuals do not trust the banking system, the amount of savings or individuals interested in accessing banking products may be lower than expected. Other examples of behavioral changes sought by investment climate interventions include increasing firms' tax compliance, increasing firms' formalization, and increasing participation in the banking system.

- Evaluations of investment climate interventions may also seek to document *changes in perceptions or “subjective” outcomes*, such as improved quality of a service, decreasing corruption, or increased transparency. As noted above, the definition and understanding of ‘quality’, ‘corruption’ or ‘transparency’ may vary considerably across individuals or firms based on location, culture, experience with government officials, and other factors.

As an example, in an evaluation that assessed the effect of management consulting on textile firms in India, researchers found that firms did not adopt management practices because managers considered their products to be of high quality (Bloom et. al., 2013). However, their products were not complying with international standards, hindering their market access and competitiveness. Through interviews and direct observation it was perceived that managers were defining “high quality” using as a comparison point other local firms in the sector, hence they never realized the need to adopt management practices that would bring their products up to *global* standards.

- Evaluations of investment climate reforms may involve the assessment of outcomes with complex definitions or meanings may vary relative to context and location. As mentioned in the previous section, a deep understanding of the concept in its context helps avoid substantive miscommunication that can result in a range of errors and incorrect measurement and interpretation of results.

b) Nature of the intervention

- *Target population is difficult to identify.* Evaluation is difficult when targeting depends on unobservable behaviors or characteristics, or when the target population is marginalized or vulnerable—and thus, by definition, unlikely to be adequately represented in formal surveys. In a recent blog post, David McKenzie highlights the difficulty of identifying a sample of informal firms in Sri Lanka¹⁵ to explore the demand for formalization, given that (a) there were no data available on informal firms and (b) firms usually lie about their legal status. Another common target is informal firms with an “entrepreneurial” profile. Focus groups and in-depth interviews are useful to uncover the desired characteristics, and the suitability of their match with the prevailing economic conditions. In these latter circumstances, quantitative techniques can be applied to ensure a representative sample while qualitative work can be focused on the particular ways in which target populations display entrepreneurial behavior consistent with local market conditions.
- Implementation is *transaction intensive* when many agents are required to interact face-to-face over an extended period of time. When several stakeholders participate in the implementation of the intervention, whether at the individual level or institutional level, it is difficult to assess what happened during the process of project implementation and how it might have affected outcomes. Qualitative techniques such as ‘process tracing’ (George and Bennett, 2005) and ‘contribution analysis’ (Mayne, 2001) enable evaluators to document, over time, the ways in which various actors and combinations of actors yielded observed outcomes.

¹⁵ <http://blogs.worldbank.org/impactevaluations/was-de-soto-wrong-impacts-of-a-formalization-experiment-in-sri-lanka>

- *Some investment climate interventions depend on voluntary participation/take-up.* When the quality of implementation relies heavily on the participation of firms or individuals, it is important to understand the perceptions, ways and means by which firms/individuals are incentivized to participate (or not). As an example, in the Lima Tracer study previously mentioned, researchers found through interviews that firms were not willing to accept financial incentives for formalization since they did not perceive any benefit from formalization. More dramatically, a study in Brazil deploying a mixed methods design showed that business owners incorrectly perceived that a large-scale initiative to register firms was in fact a thinly veiled attempt by the state to expropriate their assets; they thus refused to sign up, even though, importantly, the initiative had worked well as a pilot (Bruhn and McKenzie, 2013). Qualitative research thus yielded key insights as to why a successful pilot project had failed to scale up.
- *Small N interventions.* Qualitative approaches such as, process tracing, general elimination methodology or contribution analysis are methodologies that help to establish causal relations and test underlying mechanisms. White and Phillips propose a useful framework to systematically tackle causal relations through qualitative methods and provide recommendations on how to diminish the risk of biases that are common in qualitative research (White et al, 2012).

5. Data collection techniques and analysis

There are many different types of qualitative analysis techniques.¹⁶ Table 2 lists and briefly defines the most common methods of qualitative analysis, which we provide for the sake of completeness and for those seeking to expand their knowledge of different approaches. For the purposes of applied analyses of development interventions, however, and particularly for investment climate evaluations, more detailed descriptions of the three most common data

¹⁶ For a more applied discussion of how different qualitative methods can be used in field settings to assess the nature and salience of social norms and networks, see Dudwick et al (2006). For a more general overview of mixed methods research, see Cresswell (2014).

collection techniques (focus groups, in-depth interviews, and case studies) that are likely to be used are outlined in the section immediately following Table 2.

Table 2: Some Common Methods of Qualitative Analysis

Type of Analysis	Short Description of Analysis
Constant comparison analysis	Systematically reducing data to codes, then developing themes from codes
Classical content analysis	Counting the number of codes
Word count	Counting the total number of words used or the number of times a particular word (or concept) is used
Keywords-in-context	Identifying keywords and using the surrounding words to understand the underlying meaning of the keyword
Domain analysis	Using the relationships between symbols and referents to identify domains
Taxonomic analysis	Crating a system of classification that inventories the domains into a flowchart or diagram to help the researcher understand the relationships among the domains
Componential analysis	Using matrices and/or tables to discover the differences among the subcomponents of domains
Conversation analysis	Using the behavior of speakers to describe people's methods for producing orderly social interaction
Discourse analysis	Selecting representative or unique segments of language use, such as several lines of an interview transcript, and then examining the selected lines in detail for rhetorical organization, variability, accountability, and positioning
Secondary data analysis	Analyzing non-naturalistic data or artifacts that were derived from previous studies
Membership categorization analysis	Utilizing the role that interpretations play in making descriptions and the consequences of selecting a particular category (e.g. baby, sister, brother, mother, father=family)
Semiotics	Using talk and text as systems of signs under the assumption that no inherent meaning can be attached to a single term
Manifest content analysis	Describing observed (i.e., manifest) aspects of communication via objective, systematic, and empirical means
Latent content analysis	Uncovering underlying meaning of text
Qualitative comparative analysis	Systematically analyzing similarities and differences across cases, typically being used as a theory-building approach, allowing the analyst to make connections among previously built categories, as well as to test and to develop the categories further
Narrative analysis	Considering the potential of stories to give meaning to

	individuals' lives, and treating data as stories, enabling researchers to take account of research participants' own evaluations
Text mining	Analyzing naturally occurring text to discover and capture semantic information
Micro-interlocutor analysis	Analyzing information stemming from one or more focus groups about which participants respond to each question, the order in which participants respond, the characteristics of each response, the nonverbal communication used, and the like

Source: Bergman (2010)

The most common qualitative data collection techniques used in impact evaluations

The most common qualitative data collection techniques are in-depth interviews, focus groups and case studies. Unfortunately, the details on data collection techniques are commonly underreported in the few private sector development evaluation reports that have used a mixed methods approach. As such, it is difficult to provide detailed examples of how these methods were selected to answer specific questions or how they proved helpful in dealing with challenges that emerged in the course of conducting the research. The aim of this concluding section is thus to give an overview of the main characteristics, advantages and disadvantages of each method.

Focus Groups

Focus groups are used to collect different points of view through a facilitated group discussion. Focus groups allow interaction among individuals who discuss and listen with others; participants assess their positions or opinions about a specific topic in the light of the views of others, and convey insights regarding their particular knowledge of a specific social context. As such, focus groups are often used to analyze sensitive and/or complex issues (on which, by definition, a range of views are likely to be held), and to explore how people talk about a topic. When carefully moderated, focus groups lead to self-reflection and to participants feeling sufficiently comfortable to articulate and refine their points of views. This dynamic also contributes to eliciting perspectives less influenced by the interaction of the researcher (Finch and Lewis, 2003), or which the researcher may initially have not even considered.

There are two basic types of focus groups, but both usually comprise some 6 to 8 individuals.¹⁷ The first and most common type gathers people with similar characteristics, especially if the topic of discussion is specific to a particular group (e.g., by age, gender, or occupation). In such circumstances, to include individuals with very different characteristics might be a hindrance to the flow of the discussion and to individuals expressing their true opinions and ideas. (The second type of focus group explicitly seeks people with diverse backgrounds, which may be desirable when rapid information is required or when one is seeking to assess how strongly people hold their views in the presence of those very different from them.) Researchers should select the variables along which the population is being selected according to the research topic (i.e., occupation, level of income, sector) and select the number of focus groups according to the variables selected for variation (i.e., age, gender, education level, occupation, etc.).

The moderator's skills in designing and planning of the focus group are key to achieving the desired objectives. To this end, a moderator should ideally be fully conversant with local languages and political dynamics, and experienced in managing the flow of a collective conversation that might sometimes become contentious. It is recommended that the focus group is conducted by both a moderator and an assistant: the moderator will facilitate the discussion while the assistant will take notes and record the discussion. The moderator plays a neutral role, so s/he must refrain from giving personal opinions and should control reactions towards participants (verbal and nonverbal), neutralize conflicts, summarize complex comments, and ensure that all questions are adequately understood.

It is recommended that a set of no more than 10 open-ended questions are prepared beforehand (Krueger, 2008). These questions can be usefully categorized as:

1. Opening questions: to break the ice and make participants comfortable
2. Introductory questions: to introduce participants to the topic of discussion
3. Key questions: the focus of the research
4. Ending questions: to check if anything important has been left out in the discussion

¹⁷ In village settings, however, it is not uncommon to have quite large group discussions since this is often a community norm (and thus to insist on a smaller number would potentially offend community members or compromise the integrity of the discussion).

The advantage of a focus group for data collection is that is useful to:

- Get in-depth information related to perceptions, processes, normative influences, social context, and interactions and shared meanings
- Focus on ‘how’ and ‘why’ participants perceive given outcomes to have occurred
- Contrast, confirm and clarify data collected through quantitative methods or other qualitative techniques as well.
- Address conceptual, abstract or technical topics

Focus groups are not recommended when (Morgan, 1997):

- Social interactions among participants are difficult (e.g., when discussing civil conflict, raising potential ethical considerations)
- If the topic is sensitive and people in general are not interested or feel uncomfortable talking about it (e.g., gender relations within a household).

Disadvantages of the focus group might be:

- If not conducted properly, discussion can be dominated by a few individuals, resulting in biased responses.
- Responses generated from individuals in a group discussion may differ from responses they would give if interviewed alone (i.e., when less influenced by peers or influential figures).
- Analysis is complex and time consuming; many pages of textual information from different groups need to be synthesized and interpreted.

More information on how to conduct and analyze focus groups can be found in:

Krueger, Richard and Mary Anne Casey. *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage Publications 4th edition. 2008.

Finch, Helen and Jane Lewis. “*Focus Group*”, in Jane Ritchie and Jane Lewis (eds.) *Qualitative Research Practice*. Thousand Oaks, CA: Sage Publications, 1st edition. 2003.

Morgan, David L. *The Focus Group Guidebook*. Thousand Oaks, CA: Sage Publications. 1997.

In-depth Interviews

In contrast with surveys, in-depth interviews are conversations based on questions, many of which may be open-ended; the goal is strike the right balance between structure and flexibility. In-depth interviews are useful to investigate individual points of view, beliefs and experiences, yielding detailed insights into—and understandings of—personal context. They are especially appropriate for exploring sensitive topics and issues that are not easy (and/or ethically undesirable) to discuss in a group.

There are three types of interviews: structured, semi-structured and unstructured. Structured interviews follow a questionnaire with predetermined questions with little variation. Hence, they limit the extent to which the participant can go into detail on specific topics or issues that are not specifically addressed by the questions. However, they can be administered relatively easily and quickly (e.g., customer satisfaction surveys conducted over the phone). On the other hand, semi-structured interviews use key questions as a guide to explore topics for discussion. These questions provide guidance during the conversation, allowing certain flexibility to the respondent to go into greater detail on certain topics. Finally, unstructured interviews start with an open question and continue without any predetermined list of questions. They last relatively longer and need certain expertise from the interviewer. This type of interview is recommended when in-depth information is required or the researcher has little information about a specific topic (Legard et al, 2003).

Semi and unstructured interviews are the most useful for obtaining in-depth information, especially if the interaction between the researcher and the respondent is respectful and open. In these circumstances, in-depth interviews allow “the researcher to explore fully all the factors that underpin participants’ answers: reasons, feelings, opinions and beliefs. This furnishes the explanatory evidence that is an important element of qualitative research.”¹⁸

After identifying and selecting the respondents, it is recommended that an interview protocol be developed in case different persons are implementing the interviews. The protocol includes guidelines to conduct the interview, how to act during the interview and what should be done after the interview to ensure that responses can be adequately compared and contrasted (Boyce

¹⁸ Legard et al. (2003).

and Neale, 2006). Even if the type of interview is semi structured or unstructured the interviewer should still prepare a list of questions beforehand, including examples to explain the questions or to provide further explanations. The interviewer should explain the purpose of the interview, its duration and confidentiality status, and notify the respondent as to how notes will be taken and if a recording will be made. Interviewers should of course explicitly seek the informed consent of those being interviewed, and clearly indicate the purpose(s) to which their responses will be put.

The interviewer's attitude should be neutral towards the topic. The interviewer should be skillful at listening and quickly understand the participant's answers to deepen on it or pursue the next question. The data collector should start asking factual questions to be followed by questions that require a more personal point of view or judgment. In addition, the first questions should be focused in the present and the following ones in the past and future (if required). Moreover, given the complexities that may be inherited to the use of translators, it is also recommended to brief the translator on the purposes of the study and the topic to reduce misunderstandings and ensure a better translation quality (USAID, 1996).

To learn more about best practices on how to conduct and analyze in-depth interviews, readers should refer to:

Legard, Robin, Jill Keegan and Kit Ward. "In-depth Interviews", in Jane Ritchie and Jane Lewis (eds,) *Qualitative Research Practice*. Thousand Oaks, CA: Sage Publications, 1st edition, 2003.

Boyce, Carolyn and Palena Neale. *Conducting In-Depth Interviews: A Guide for Designing and Conducting In-Depth Interviews for Evaluation Input, Monitoring and Evaluation* Pathfinder International, 2006. Available at http://www2.pathfinder.org/site/DocServer/m_e_tool_series_indepth_interviews.pdf?docID=6301

United States Agency for International Development, Center for Development Information and Evaluation. *Conducting Key Informant Interviews*. Performance Monitoring and Evaluation TIPS, 1996. Available at: http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf

Case Studies

Case studies are used in a range of disciplines (including medicine, law and business) for evaluative, diagnostic and pedagogical purposes. They are selected on the basis of a considered awareness of the broader context, such that their status as being ‘representative’ of or an ‘outlier’ in this broader context is well understood. (Without such knowledge, case studies are vulnerable to the critique of “selecting on the dependent variable”; that is, of failing to recognize that the same factors seemingly driving success may elsewhere have no—or even the opposite—effect.) Case studies are usually conducted comparatively, either by assessing different cases (‘between case’ analysis) or by exploring constituent aspects of a single case (‘within case’ analysis) (see Gerring, 2007).

For present purposes, the comparative advantage of case studies is that they can be (a) a vehicle for integrating insights derived from different methodological perspectives and units of analysis into a single, coherent narrative; and (b) can be used to trace *how*—i.e., to identify the conditions under which—particular combinations of events and processes yielded particular outcomes. They are currently being deployed to great effect in understanding the bases for effective service delivery in otherwise unlikely settings¹⁹, the better to gain insights on how domestic reformers elsewhere might proceed. In those instances where investment climate reforms are being launched at the national (or sub-national) level, case studies are especially well suited to investigating in detail how, where and why they are being experienced by different groups.

Readers interested in learning more about preparing case studies should consult:

George, Alexander and Andrew Bennett. *Case Studies and Theory Development in the Social Sciences* Cambridge, MA: MIT Press, 2005.

Gerring, John. *Case Study Research: Principles and Practices*. New York: Cambridge University Press, 2007.

Weller, Nicholas and Jeb Barnes. *Finding Pathways: Mixed-Method Research for Studying Causal Mechanisms*. New York: Cambridge University Press, 2014.

¹⁹ See in particular the evolving library of cases being compiled (and made freely available) by Princeton University’s ‘Innovations for Successful Societies’ (<http://successfulsocieties.princeton.edu/>).

Yin, Robert K. *Case Study Research: Design and Methods* (5th ed). Thousand Oaks, CA: Sage Publications, 2014.

6. Implications

Researchers, policymakers and donors are increasingly calling for high-quality evaluations of development interventions; this is to be welcomed and encouraged. In responding to such imperatives, however, it is important to appreciate that all methods have strengths and weaknesses, and that particular methods (or combinations of methods) should be selected to the extent they meet ethical criteria, can be faithfully implemented within the prevailing time, budget, political, and human resource constraints, and are an optimal response to the *type* of problem under consideration. As we have argued, the empirical material derived from any given methodological strategy must also be interpreted in the light of theory and experience; the significance and policy implications of evidence are never self-evident.

A mixed method approach can contribute to improving the quality of the evaluation—and indeed the program design itself—by enhancing the accuracy of the empirical claims being made regarding the intervention’s impact (or lack thereof), explanations of *why* and *how* such impact was achieved (or not), and the likelihood that others implementing a ‘similar’ project could expect comparable results elsewhere, in the future, or at a larger scale. The particular ‘comparative advantage’ of qualitative methods is understanding the idiosyncrasies of the context in which an intervention is being implemented; how outcomes are influenced by this context provides a crucial entry point for policy makers and practitioners seeking to more accurately assess and improve the effectiveness of development initiatives, of all kinds.

Having an array of methodological tools in one’s kit is the best response to the many contingencies, problems and setbacks that are an evaluator’s constant companion. One need not be an ‘expert’ in all of these methods—indeed, this is highly unlikely even after a full research career—but one can at least recognize the strengths and weaknesses, and virtues and limits, of any singular approach, and seek to offset these by calling upon others who have the complementary skills (and temperaments) required. Just as professionals ranging from carpenters

to surgeons seek to use the right instrument for the right problem—and do not regard, for example, a hammer as inherently “more rigorous” than a screwdriver—so too must evaluators of development interventions seek to cultivate a comprehensive awareness of how to craft the right combinations of methods for the task at hand. This sensibility, as much as the technical skill associated with mastering a particular method, is what is required to generate robust, useful and useable insights regarding the effectiveness of investment climate reforms. The use of qualitative and mixed methods approaches has made a promising start in this domain; it is our hope that this summary note can encourage and inform the important next steps.

References

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. “Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program.” *Journal of the American Statistical Association* 105(490): 493-505, 2010.

Adler, Daniel, Doug Porter and Michael Woolcock. “Legal pluralism and equity: Some reflections on land reform in Cambodia.” World Bank, 2008. Available at <http://siteresources.worldbank.org/INTJUSFORPOOR/Resources/J4PBriefingNoteVolume2Issue2.pdf>

Alcázar Lorena and Miguel Jaramillo. “The real barriers for municipal formalization: A qualitative case study in Lima Cercado.” GRADE, 2011.

Alcázar Lorena and Miguel Jaramillo. “Panel /tracer study on the impact of business facilitation processes on microenterprises and identification of priorities for future business enabling environment projects in Lima, Peru” GRADE, 2011.

Ananthpur, Kripa, Kabur Malik and Vijayendra Rao. “The anatomy of failure: An ethnography of a randomized trial to deepen democracy in rural India.” Policy Research Working Paper No. 6958, Washington, DC: World Bank, 2014.

Bamberger, Michael. "Introduction to mixed methods in impact evaluation." Interaction.org 2012. Available at <http://www.interaction.org/impact-evaluation-notes>.

Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. "Using mixed methods in monitoring and evaluation: experiences from international development", in Abbas Tashakkori and Charles Teddlie (eds.) *Handbook of Mixed Methods in Social and Behavioral Research* (2nd revised edition) Thousand Oaks, CA: Sage Publications, pp. 613-641, 2010.

Bamberger, Michael. "Conducting quality impact evaluations under budget, time and data constraints." World Bank, Independent Evaluation Group/Poverty Analysis, Monitoring and Impact Evaluation Thematic Group, PREM Network, 2006.

Bamberger, Michael. "Integrating quantitative and qualitative research in development projects." World Bank Publications, 2000.

Barron, Patrick, Rachael Diprose and Michael Woolcock. *Contesting Development: Participatory Projects and Local Conflict Dynamics in Indonesia*. New Haven: Yale University Press, 2011.

Bennett, Andrew. "Process tracing and causal inference", in Henry Brady and David Collier (eds.) *Rethinking Social Inquiry* (2nd ed.) Rowman and Littlefield, 2010.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts. "Does management matter? Evidence from India" *Quarterly Journal of Economics* 128(1): 1-51, 2013.

Biddulph, Robin. "Cambodia's Land Management and Administration project" WIDER Working Paper No. 2014/086. Helsinki: UNU-WIDER, 2014.

Bruhn, Miriam and David McKenzie. "Using administrative data to evaluate municipal reforms: an evaluation of the impact of Minas Fácil Expresso." Washington DC: World Bank Policy Research Working Paper No. 6368, 2013.

Boyce, C and Neale, P. *Conducting In-Depth Interviews: A Guide for Designing and Conducting In-Depth Interviews for Evaluation Input, Monitoring and Evaluation – 2*, 2006. Pathfinder International. Available at http://www2.pathfinder.org/site/DocServer/m_e_tool_series_indepth_interviews.pdf?docID=6301

Cadot, Olivier, Ana M. Fernandes, Julien Gourdon and Aaditya Mattoo. "Are the benefits of export support durable? Evidence from Tunisia." World Bank, 2012. <https://openknowledge.worldbank.org/handle/10986/12189>

Chung, Kimberly. "Issues and approaches in the use of integrated methods," in Michael Bamberger (ed.) *Integrating Quantitative and Qualitative Research in Development Projects*, Washington, DC: World Bank, pp. 37–46, 2000.

Clark, Vicki Plano and Manijeh Baidee. "Research questions in mixed methods research", in Abbas Tashakkori and Charles Teddlie (eds.) *Handbook of Mixed Methods in Social and Behavioral Research* (2nd revised edition) Thousand Oaks, CA: Sage Publications, pp. 275-304, 2010.

Collier, David. "Understanding process tracing." *PS: Political Science and Politics* 44(4): 823-30, 2011.

Creswell, John. *A Concise Introduction to Mixed Methods Research*. Thousand Oaks, CA: Sage Publications, 2014.

De Mel, Suresh, David McKenzie and Woodruff, Christopher. "The demand for, and consequences of, formalization among informal firms in Sri Lanka." Policy Research Working Paper Series 5991, World Bank, 2012.

Department for International Development (DFID). "Broadening the range of designs and methods for impact evaluations" Working Paper 38, April 2012

Dudwick, Nora, Kathleen Kuehnast, Veronica Nyhan Jones, and Michael Woolcock. "Analyzing social capital in context." Washington DC: World Bank Institute Working Paper No. 37260, 2006.

Ensminger, Jean. "Experimental economics in the bush." *Engineering and Science* 65(2): 6-16, 2000.

Finch, Helen and Jane Lewis. "Focus groups", in Jane Ritchie and Jane Lewis (eds.) *Qualitative Research Practice* (1st edition). Thousand Oaks, CA: Sage Publications, 2003.

Garbarino, Sabine, and Jeremy Holland. "Quantitative and qualitative methods in impact evaluation and measuring results." Discussion Paper. University of Birmingham, Birmingham, UK, 2009.

George, Alexander and Andrew Bennett. *Case Studies and Theory Development in the Social Sciences* Cambridge, MA: MIT Press, 2005.

Gerring, John. *Case Study Research: Principles and Practices*. New York: Cambridge University Press, 2007.

Goertz, Gary and James Mahoney. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press, 2012.

Hentschel, Jesko. "Contextuality and data collection methods: A framework and application to health service utilisation." *Journal of Development Studies* 35(4): 64-94, 1999.

Honore, Anthony. "Causation in the law", in *Stanford Encyclopedia of Philosophy*, 2010. Available at: <http://stanford.library.usyd.edu.au/entries/causation-law/>

Hopkins, Daniel J. and Gary King. "Improving anchoring vignettes: Design surveys to correct interpersonal incomparability." *Public Opinion Quarterly* 74(2): 201-222, 2010.

Kitzmuller, Markus and Martha Martinez Licetti. "Competition policy: Encouraging thriving markets for development." World Bank View Point Series, September 2012. Available at http://fpdweb.worldbank.org/units/fpdvp/ficdr/Impact/Documents/Encouraging%20Thriving%20Markets%20for%20Development_Competition-Policy_Viewpoint.pdf

Kristjanson, Patti, Nelson Mango, Anirudh Krishna, Maren Radeny and Nancy Johnson. "Understanding poverty dynamics in Kenya." *Journal of International Development* 22(7): 978-996, 2010.

Krueger, Richard and Mary Anne Casey. *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage Publications 4th edition, 2008.

Legard, Robin, Jill Keegan and Kit Ward. "In-depth interviews," in Jane Ritchie and Jane Lewis (eds.) *Qualitative Research Practice* (1st edition). Thousand Oaks, CA: Sage Publications, 2003.

Ludwig, Jens, Jeffrey R. Kling and Sendhil Mullainathan. "Mechanism experiments and policy evaluations" *Journal of Economic Perspectives* 25(3): 17-38, 2011.

Mayne, John. "Addressing attribution through contribution analysis: using performance measures sensibly." *Canadian Journal of Program Evaluation* 16(1): 1-24, 2001.

Morgan, David L. *The Focus Group Guidebook*. Thousand Oaks, CA: Sage Publications, 1997.

Onwuegbuzie, Anthony and Julie P. Combs. *Emergent Data Analysis Techniques in Mixed Methods Research*. Thousand Oaks, CA: Sage Publications, 2010.

Rao, Vijayendra and Michael Woolcock. "Integrating qualitative and quantitative approaches in program evaluation", in Francois J. Bourguignon and Luiz Pereira da Silva (eds.) *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools* New York: Oxford University Press, pp. 165-90, 2003.

Rao, Vijayendra. "Mixed methods for poverty analysis". World Bank, 2002. Available at http://siteresources.worldbank.org/INTPAME/Resources/Training-Materials/Training_2002-06-19_Rao_MixedMethods_pres.pdf

Ravallion, Martin. "The mystery of the vanishing benefits: An introduction to impact evaluation." *World Bank Economic Review* 15(1): 115-140, 2000.

Ravallion, Martin. "Evaluation in the practice of development." *World Bank Research Observer* 24(1): 29-53, 2009.

Ravallion, Martin. "The two poverty enlightenments: Historical insights from digitized books spanning three centuries." *Poverty & Public Policy* 3(2): 1-46, 2011.

Rogers, Everett. *The Diffusion of Innovations* (4th ed.) New York: Simon and Schuster, 2010.

Rugh, Jim, Michael Bamberger, and Linda Mabry. *RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints*. Thousand Oaks, CA: Sage Publications, 2011.

Sana, Mariano; Gu Stecklov and Alexander A. Weinreb. "Local or outsider interviewer? An experimental evaluation." 2012. Available at <http://paa2012.princeton.edu/papers/122313>

Shaffer, Paul. *Q-Squared: Combining Qualitative & Quantitative Approaches in Poverty Analysis*. New York: Oxford University Press, 2013.

Teddle Charles and Fen Yu. "Mixed methods sampling: A typology with examples." *Journal of Mixed Methods Research* 1(1): 77-100, 2007.

Thelan, Kathleen and James Mahoney (eds.) *Explaining Institutional Change: Ambiguity, Agency, and Power* New York: Cambridge University Press, 2010.

United States Agency for International Development's Center for Development Information and Evaluation. (1996). Conducting Key Informant Interviews. (Performance Monitoring and Evaluation TIPS) Available at http://www.usaid.gov/pubs/usaid_eval/pdf_docs/pnabs541.pdf

Weller, Nicholas and Jeb Barnes. *Finding Pathways: Mixed-Method Research for Studying Causal Mechanisms*. New York: Cambridge University Press, 2014.

White Howard and Daniel Phillips. "Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework" International Initiative for Impact Evaluation

Working Paper 15, June 2012. Available at http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf

Woolcock Michael. "Using qualitative and mixed methods in project evaluation." Innovations in Investment Climate Reforms, an Impact Evaluation Workshop. Paris, November 2012. Available at <https://www.wbginvestmentclimate.org/results/workshop-agenda-and-materials.cfm>

Woolcock, Michael. "Toward a plurality of methods in project evaluation: A contextualised approach to understanding impact trajectories and efficacy." *Journal of Development Effectiveness* 1(1): 1-14, 2009.

Woolcock, Michael. "Using case studies to explore the external validity of complex development interventions." *Evaluation* 19(3): 229-248, 2013.

Yin, Robert K. *Case Study Research: Design and Methods* (5th ed.). Thousand Oaks, CA: Sage Publications, 2014.