

1 Using Case Studies to Enhance the Quality of Explanation and Implementation

Integrating Scholarship and Development Practice

Jennifer Widner, Michael Woolcock, and Daniel Ortega Nieto

1.1 Introduction

In recent years the development policy community has turned to case studies as an analytical and diagnostic tool. Practitioners are using case studies to discern the mechanisms underpinning variations in the quality of service delivery and institutional reform, to identify how specific challenges are addressed during implementation, and to explore the conditions under which given instances of programmatic success might be replicated or scaled up.¹ These issues are of prime concern to organizations such as Princeton University's Innovations for Successful Societies (ISS)² program and the Global Delivery Initiative (GDI),³ housed in the World Bank Group (from 2015–2021), both of which explicitly prepare case studies exploring the dynamics underpinning effective implementation in fields ranging from water, energy, sanitation, and health to cabinet office performance and national development strategies.

In this sense, the use of case studies by development researchers and practitioners mirrors their deployment in other professional fields. Case studies have long enjoyed high status as a pedagogical tool and research

The views expressed in this chapter are those of the authors alone, and should not be attributed to the organizations with which they are affiliated.

¹ For example, see Barma, Huybens, and Viñuela (2014); Brix, Lust, and Woolcock (2015); and Woolcock (2013).

² See <https://successfultsocieties.princeton.edu/>.

³ GDI's case studies are available (by clicking on "Case studies" under the search category "Resource type") at www.effectivecooperation.org/search/resources.

method in business, law, medicine, and public policy, and indeed across the full span of human knowledge. According to Google Scholar data reported by Van Noorden, Maher, and Nuzzo (2014), Robert Yin's *Case Study Research* (1984) is, remarkably, the sixth most cited article or book *in any field, of all time*.⁴ Even so, skepticism lingers in certain quarters regarding the veracity of the case study method – for example, how confident can one be about claims drawn from single cases selected on a nonrandom or nonrepresentative basis? – and many legitimate questions remain (Morgan 2012). In order for insights from case studies to be valid and reliable, development professionals need to think carefully about how to ensure that data used in preparing the case study is accurate, that causal inferences drawn from it are made on a defensible basis (Mahoney 2000; Rohlfling 2012), and that broader generalizations are carefully delimited (Ruzzene 2012; Woolcock 2013).⁵

How best to ensure this happens? Given the recent rise in prominence and influence of the case study method within the development community and elsewhere, scholars have a vital quality control and knowledge dissemination role to play in ensuring that the use of case studies both accurately reflects and contributes to leading research. To provide a forum for this purpose, the World Bank's Development Research Group and its leading operational unit deploying case studies (the GDI) partnered with the leading academic institution that develops policy-focused case studies of development (Princeton's ISS) and asked scholars and practitioners to engage with several key questions regarding the foundations, strategies, and applications of case studies as they pertain to development processes and outcomes:⁶

- What are the distinctive virtues and limits of case studies, in their own right and vis-à-vis other research methods? How can their respective strengths be harnessed and their weaknesses overcome (or complemented by other approaches) in policy deliberations?

⁴ Van Noorden et al. (2014) also provide a direct link to the dataset on which this empirical claim rests. As of this writing, according to Google Scholar, Yin's book (across all six editions) has been cited over 220,000 times; see also Robert Stake's *The Art of Case Study Research* (1995), which has been cited more than 51,000 times.

⁵ In addition to those already listed, other key texts on the theory and practice of case studies include Feagin, Orum and Sjöberg (1991), Ragin and Becker (1992), Bates et al. (1998), Byrne and Ragin (2009), and Gerring (2017). See also Flyvbjerg (2006).

⁶ As such, this volume continues earlier dialogues between scholars and development practitioners in the fields of history (Bayly et al. 2011), law (Tamanaha et al. 2012), and multilateralism (Singh and Woolcock, forthcoming).

- Are there criteria for case study selection, research design, and analysis that can help ensure accuracy and comparability in data collection, reliability in causal inference within a single case, integrity in statements about uncertainty or scope, and something akin to the replicability standard in quantitative methods?
- Under what conditions can we generalize from a small number of cases? When can comparable cases be generalized or not (across time, contexts, units of analysis, scales of operation, implementing agents)?
- How can case studies most effectively complement the insights drawn from household surveys and other quantitative assessment tools in development research, policy, and practice?
- How can lessons from case studies be used for pedagogical, diagnostic, and policy-advising purposes as improvements in the quality of implementation of a given intervention are sought?
- How can the proliferation of case studies currently being prepared on development processes and outcomes be used to inform the scholarship on the theory and practice of case studies?

The remainder of this chapter provides an overview of the distinctive features (and limits) of case study research, drawing on “classic” and recent contributions in the scholarly literature. It provides a broad outline of the key claims and issues in the field, as well as a summary of the book’s chapters.

1.2 The Case for Case Studies: A Brief Overview

We can all point to great social science books and articles that derive from qualitative case study research. Herbert Kaufman’s (1960) classic, *The Forest Ranger*, profiles the principal-agent problems that arise in management of the US Forest Service as well as the design and implementation of several solutions. Robert Ellickson’s (1991) *Order Without Law* portrays how ranchers settle disputes among themselves without recourse to police or courts. Judith Tandler’s (1997) *Good Government in the Tropics* uses four case studies of Ceara, Brazil’s poorest state, to identify instances of positive deviance in public sector reform. Daniel Carpenter’s (2001) *The Forging of Bureaucratic Autonomy*, based on three historical cases, seeks to explain why reformers in some US federal agencies were able to carve out space free from partisan legislative interference while others were unable to do so. In “The Market for Public Office,” Robert Wade (1985) elicits the strategic structure

of a particular kind of spoiler problem from a case study conducted in India. In economics, a longitudinal study of poverty dynamics in a single village in India (Palanpur)⁷ has usefully informed understandings of these processes across the subcontinent (and beyond).

What makes these contributions stand out compared to the vast numbers of case studies that few find insightful? What standards should govern the choice and design of case studies, generally? And what specific insights do case studies yield that other research methods might be less well placed to provide?

The broad ambition of the social sciences is to forge general insights that help us quickly understand the world around us and make informed policy decisions. While each social science discipline has its own distinctive approach, there is broad agreement upon a methodological division of labor in the work we do. This conventional wisdom holds that quantitative analysis of large numbers of discrete cases is usually more effective for testing the veracity of causal propositions, for estimating the strength of the association between readily measurable causes and outcomes, and for evaluating the sensitivity of correlations to changes in the underlying model specifying the relationship between causal variables (and their measurement). By contrast, qualitative methods generally, and case studies in particular, fulfill other distinct epistemological functions and are the predominant method for:

1. Developing a theory and/or identifying causal mechanisms (e.g., working inductively from evidence to propositions and exploring the contents of the “black box” processes connecting causes and effects)
2. Eliciting strategic structure (e.g., documenting *how* interaction effects of one kind or another influence options, processes, and outcomes)
3. Showing how antecedent conditions elicit a prevailing structure which thereby shapes/constrains the decisions of actors within that structure
4. Testing a theory in novel circumstances
5. Understanding outliers or deviant cases

The conventional wisdom also holds that in an ideal world we would have the ability to use both quantitative and qualitative analysis and employ “nested” research designs (Bamberger, Rao, and Woolcock 2010; Goertz and Mahoney

⁷ The initial study in what has become a sequence is Bliss and Stern (1982); for subsequent rounds, see Lanjouw and Stern (1998) and Lanjouw, Murgai, and Stern (2013). This study remains ongoing, and is now in its seventh decade.

2012; Lieberman 2015). However, the appropriate choice of method depends on the character of the subject matter, the kinds of data available, and the array of constraints (resources, politics, time) under which the study is being conducted. The central task is to deploy those combinations of research methods that yield the most fruitful insights in response to a specific problem, given the prevailing constraints (Rueschemeyer 2009). We now consider each of these five domains in greater detail.

1.3 Developing a Theory and/or Identifying Causal Mechanisms

Identifying a causal mechanism and inferring an explanation or theory are important parts of the research process, especially in the early stages of knowledge development. The causal mechanism links an independent variable to an outcome, and over time may become more precise: to cite an oft-used example, an initial awareness that citrus fruits reduced scurvy became more refined when the underlying causal mechanism was discovered to be vitamin C. For policy purposes, mechanisms provide the basis for a compelling storyline, which can greatly influence the tone and terms of debate – or the space of what is “thinkable,” “say-able,” and “do-able” – which in turn can affect the design, implementation, and support for interventions. This can be particularly relevant for development practitioners if the storyline – and the mechanisms it highlights – provides important insights into how and where implementation processes unravel, and what factors enabled a particular intervention to succeed or fail during the delivery process.

In this way, qualitative research can provide clarity on the factors that influence critical processes and help us identify the mechanisms that affect particular outcomes. For example, there is a fairly robust association, globally, between higher incomes and smaller family sizes. But what is it about income that would lead families to have fewer children – or does income mask other changes that influence child-bearing decisions? To figure out the mechanism, one could conduct interviews and focus groups with a few families to understand decision-making about family planning. Hypotheses based on these family case studies could then inform the design of survey-based quantitative research to test alternative mechanisms and the extent to which one or another predominates in different settings. Population researchers have done just that (see Knodel 1997).

Case studies carried out for the purpose of inductive generalization or identifying causal mechanisms are rarely pure “soak and poke” exercises

uninformed by any preconceptions. Indeed, approaching a case with a provisional set of hypotheses is vitally important. The fact that we want to use a case to infer a general statement about cause and effect does not obviate the need for this vital intellectual tool; it just means we need to listen hard for alternative explanations we did not initially perceive and be highly attentive to actions, events, attitudes, etc., that are at odds with the reasoned intuition brought to the project.

An example where having an initial set of hypotheses was important comes from a GDI case on scaling-up rural sanitation. In this case, the authors wanted to further understand how the government of Indonesia had been able to substantially diminish open defecation, which is the main cause of several diseases in thousands of villages across the country.⁸ The key policy change was a dramatic move from years of subsidizing latrines that ended up not being used to trying to change people's behavior toward open defecation, a socially accepted norm. The authors had a set of hypotheses with respect to what triggered this important policy shift: a change in cabinet members, the presence of international organizations, adjustments in budgets, etc. However, the precise mechanism that triggered the change only became clear after interviewing several actors involved in the process. It turns out that a study tour taken by several Indonesian officials to Bangladesh was decisive since, for the first time, they could see the results of a different policy "with their own eyes" instead of just reading about it.⁹

There are some situations, however, in which we may know so little that hypothesis development must essentially begin from scratch. For example, consider an ISS case study series on cabinet office performance. A key question was why so many heads of government allow administrative decisions to swamp cabinet meetings, causing the meetings to last a long time and reducing the chance that the government will reach actual policy decisions or priorities. One might have a variety of hypotheses to explain this predicament, but without direct access to the meetings themselves it is hard to know which of these hypotheses is most likely to be true (March, Sproul, and Tamuz 1991). In the initial phases, ISS researchers deliberately left a lot of space for the people interviewed to offer their own explanations. They anticipated that not all heads of state might want their cabinets to work as forums for decision-making and coordination, because ministers who had a lot of political and military clout might capture the stage or threaten vital interests of weaker members – or because the head of state benefited from the

⁸ Glavey and Haas (2015). ⁹ Glavey and Haas (2015).

dysfunction. But as the first couple of cases unfolded, the research team realized that part of the problem arose from severe under-staffing, simple lack of know-how, inadequate capacity at the ministry level, or rapid turn-over in personnel. In such situations, as March, Sproul, and Tamuz (1991: 8) aptly put it,

[t]he pursuit of rich experience . . . requires a method for absorbing detail without molding it. Great organizational histories, like great novels, are written, not by first constructing interpretations of events and then filling in the details, but by first identifying the details and allowing the interpretations to emerge from them. As a result, openness to a variety of (possibly irrelevant) dimensions of experience and preference is often more valuable than a clear prior model and unambiguous objectives.

In another ISS case study on the factors shaping the implementation and sustainability of “rapid results” management practices (e.g., setting 100-day goals, coupled with coaching on project management), a subquestion was when and why setting a 100-day goal improved service delivery. In interviews, qualitative insight into causal mechanisms surfaced: some managers said they thought employees understood expectations more clearly and therefore performed better as a result of setting a 100-day goal, while in other instances a competitive spirit or “game sense” increased motivation or cooperation with other employees, making work more enjoyable. Still others expected that an audit might follow, so a sense of heightened scrutiny also made a difference. The project in question did not try to arbitrate among these causal mechanisms or theories, but using the insight from the qualitative research, a researcher might well have proceeded to decipher which of these explanations carried most weight.

In many instances it is possible and preferable to approach the task of inductive generalization with more intellectual structure up front, however. As researchers we always have a few “priors” – hunches or hypotheses – that guide investigation. The extent to which we want these to structure initial inquiry may depend on the purpose of our research, but also on the likely causal complexity of the outcome we want to study, the rapidity of change in contexts, and the stock of information already available.

1.4 Eliciting Strategic Structure

A second important feature of the case study method, one that is intimately related to developing a theory or identifying causal mechanisms, is its ability to elicit the strategic structure of an event – that is, to capture the interactions

that produce an important outcome. Some kinds of outcomes are “conditioned”: they vary with underlying contextual features like income levels or geography. Others are “crafted” or choice-based: the outcome is the product of bargaining, negotiating, deal-cutting, brinkmanship, and other types of interaction among a set of specified actors. Policy choice and implementation fall into this second category. Context may shape the feasible set of outcomes or the types of bargaining challenges, but the only way to explain outcomes is to trace the process or steps and choices as they unfold in the interaction (see Bennett and Checkel 2015).

In process tracing we want to identify the key actors, their preferences, and the alternatives or options they faced; evaluate the information available to these people and the expectations they formed; assess the resources available to each to persuade others or to alter the incentives others face and the expectations they form (especially with regard to the strategies they deploy); and indicate the formal and informal rules that govern the negotiation, as well as the personal aptitudes that influence effectiveness and constrain choice. The researcher often approaches the case with a specific type of strategic structure in mind – a bargaining story that plausibly accounts for the outcome – along with a sense of other frames that might explain the same set of facts.

In the 1980s and 1990s, the extensive literature on the politics of structural adjustment yielded many case studies designed to give us a better understanding of the kinds of difficulties ministers of finance faced in winning agreement to devalue a currency, sell assets, or liberalize trade or commodity markets, as well as the challenges they encountered in making these changes happen (e.g., Haggard 1992). Although the case studies yielded insights that could be used to create models testable with large-N data, in any individual case the specific parameters – context or circumstance – remained important for explaining particular outcomes. Sensitivity to the kinds of strategic challenges that emerged in other settings helped decision-makers assess the ways their situations might be similar or different, identify workarounds or coalitions essential for winning support, and increase the probability that their own efforts would succeed. It is important to know what empirical relationships seem to hold across a wide (ideally full) array of cases, but the most useful policy advice is that which is given in response to specific people in a specific place responding to a specific problem under specific constraints; as such, deep knowledge of contextual contingencies characterizing each case is vital.¹⁰

¹⁰ For example, if it can be shown empirically that, in general, countries that exit from bilateral trade agreements show a subsequent improvement in their “rule of law” scores, does this provide warrant for

For example, consider the challenge of improving rural livelihoods during an economic crisis in Indonesia. In “Services for the People, By the People,” ISS researchers profiled how Indonesian policy-makers tried to address the problem of “capture” in a rural development program. Officials and local leaders often diverted resources designed to benefit the poor. The question was how to make compliance incentive compatible. That is, what did program leaders do to alter the cost–benefit calculus of the potential spoiler? How did they make their commitment to bargains, deals, pacts, or other devices credible? In most cases, the interaction is “dynamic” and equilibria (basis for compliance) are not stable. Learning inevitably takes place, and reform leaders often have to take new steps as circumstances change. Over time, what steps did a reformer take to preserve the fragile equilibrium first created or to forge a new equilibrium? Which tactics proved most effective, given the context?

In this instance, leaders used a combination of tactics to address the potential spoiler problem. They vested responsibility for defining priorities in communities, not in the capital or the district. They required that at least two of three proposals the communities could submit came from women’s groups. They set up subdistrict competitions to choose the best proposals, with elected members of each community involved in selection. They transferred money to community bank accounts that could only be tapped when the people villagers elected to monitor the projects all countersigned. They created teams of facilitators to provide support and monitor results. When funds disappeared, communities lost the ability to compete. Careful case analysis helped reveal not only the incentive design, but also the interaction between design and context – and the ways in which the system occasionally failed, although the program was quite successful overall.

A related series of ISS cases focused on how leaders overcame the opposition of people or groups who benefited from dysfunction and whose institutional positions enabled them to block changes that would improve service delivery. The ambition in these cases was to tease out the strategies reform leaders could use to reach an agreement on a new set of rules or practices; if they were able to do so, case studies focused on institutions where spoiler traps often appear: anticorruption initiatives, port reform (ports, like banks, being “where the money is”), and infrastructure. The strategies or tactics at the focus in these studies included use of external agencies of restraint (e.g., the Governance and Economic Management Assistance Program [GEMAP] in Liberia); “coalitions

advising (say) Senegal that if it wants to improve its “rule of law” then it should exit from all its bilateral trade agreements? We think not.

with the public” to make interference more costly in social or political terms; persuading opponents to surrender rents in one activity for rewards in another; pitting strong spoilers against each other; and altering the cost calculus by exposing the spoiler to new risks. The cases allowed researchers both to identify the strategies used and to weigh the sensitivity of these to variations in context or shifts in the rules of the game or the actors involved. The hope was that the analysis the cases embodied would help practitioners avoid the adoption of strategies that are doomed to fail in the specific contexts they face. It also enabled policy-makers to see how they might alter rules or practices in ways that make a reformer’s job (at least to a degree) easier.

A couple of GDI cases provide further illustration of how to elicit strategic structure. In a case on how to shape an enabling environment for water service delivery in Nigeria,¹¹ the authors were able to identify the political incentives that undermine long-term commitments and overhaul short-run returns, and which generate a low-level equilibrium trap. This has led to improvements in investments in rehabilitation and even an expansion of water services, yet it has not allowed the institutional reforms needed to ensure sustainability to move forward. In the case of Mexico, where the government had been struggling to improve service delivery to Indigenous communities, a World Bank loan provided a window of opportunity to change things. A number of reformers within the government believed that catering services to these populations in their own languages would help decrease the number of dropouts from its flagship social program, Oportunidades.¹² However, previous efforts had not moved forward. A World Bank loan to the Mexican government triggered a safeguards policy on Indigenous populations and it became fundamental for officials to be able to develop a program to certify bilingual personnel that could service these communities. Interviews with key officials and stakeholders showed how the safeguards policy kick-started a set of meetings and decisions within the government that eventually led to this program, changing the strategic structures within government.

1.5 Showing How an Antecedent Condition Limits Decision-Makers’ Options

Some types of phenomena require case study analysis to disentangle complex causal relationships. We generally assume the cause of an outcome is

¹¹ Hima and Santibanez (2015). ¹² Estabridis and Nieto (2015).