

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); Brixi, Lust, and Woolcock (2015); and Woolcock (2013).

² See <https://successfulsocieties.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 Sjoberg (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 Tendler’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 turnover 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).

exogenous, but sometimes there are feedback effects and an outcome intensifies one of its causes or limits the range of values the outcome can later assume. In such situations, case studies can be helpful in parsing the structure of these causal relationships and identifying which conditions are prior. Some of the case studies that inform *Why States Fail* (Acemoglu and Robinson 2012), for example, perform this function. More detailed case studies of this type appear in political science and sociological writing in the “historical institutionalism” tradition (see Thelen and Mahoney 2009; Mahoney and Thelen 2015).

Case studies are also useful in other instances when both the design of a policy intervention and the way in which it is implemented affect the outcome. They help identify ways to distinguish the effects of policy from the effects of process, two things that most quantitative studies conflate. To illustrate, take another ISS case study series on rapid turnarounds observed in some types of public sector agencies: the quick development of pockets of effectiveness. The agencies at the focus of this project provided business licenses or identity documents – actions that required relatively little exercise of judgment on the part of the person dispensing the service and where the number of distribution points is fairly limited. Businesses and citizens felt the effects of delay and corruption in these services keenly, but not all governments put reformers at the helm and not all reformers improved performance. The ISS team was partly interested in the interventions that produced turnarounds in this type of activity: was there a secret recipe – a practice that produced altered incentives or outlooks and generated positive results? The literature on principal–agent problems offered hypotheses about ways to better align the interests of leaders and the people on the front-line who deliver a service, but many of these were inapplicable in low-resource environments or where removing personnel and modifying terms of service was hard to do. But ISS was also interested in how the mode of implementation affected outcomes, because solving the principal–agent problem often created clear losers who could block the new policies. How did the successful reformers win support?

The team refined and expanded its initial set of hypotheses through a detailed case study of South Africa’s Ministry of Home Affairs, and traced both the influence of the incentive design and the process used to put the new practices into effect. Without the second part, the case study team might have reasoned that the results stemmed purely from changed practices and tried to copy the same approach somewhere else, but in this instance, as in many cases, the mode of implementation was critical to success. The project leader

could not easily draw from the standard toolkit for solving principal–agent problems because he could not easily remove poorly performing employees. He had to find ways to win union acceptance of the new policies and get people excited about the effort. This case study was an example of using qualitative methods to identify a causal mechanism and to develop explanations we can evaluate more broadly by conducting other case studies.

An example from the GDI is a case on addressing maternal and child mortality in Argentina in the early 2000s.¹³ As a result of the 2001 economic crisis, thousands of people lost their jobs and hence were unable to pay for private healthcare; consequently, the public health system suddenly received a vast and unexpected influx of patients. Given that the Argentine public health system had been decentralized over the preceding decades and therefore the central government's role in the provinces was minor, policy-makers had to work around a set of conditions and do it fast, given the context. The case disentangled how the central government was able to design one of the first results-based finance programs in the health sector and how this design was critical in explaining the maternal and child mortality outcomes. Policy-makers had to react immediately to the pressure on the health system and were able to make use of a provincial coordination mechanism that had become mostly irrelevant. By reviving this mechanism and having access to international funds, the central government was able to reinstate its role in provincial health care and engage key local decision-makers. Through the case study, the authors were able to assess the relevance of the policy-making process and how it defined the stakeholders' choices, as well as the effect of the process in the Argentine healthcare system.

1.6 Testing a Theory in Novel Circumstances

Case study analysis is a relatively weak method for testing explanations derived from large sample sizes but it is often the only method available if the event is relatively uncommon or if sample sizes are small. Testing a theory against a small number of instrumentally chosen cases carries some peril. If we have only a few cases to study, the number of causal variables that potentially influence the outcome could overwhelm the number of observations, making it impossible to infer anything about the relationship between two variables, except through intensive tracing of processes.

¹³ Ortega Nieto and Parida (2015).

Usually theory testing with case studies begins with a “truth table” or matrix, with the key independent variable(s) arrayed on one axis and the outcome variable arrayed on the other. The researcher collects data on the same variables in each case. The names of the cases in the cells of the table are then arranged and comparisons made of expected patterns with the actual pattern. The proportion of cases in each cell will track expectations if there is support for the theory.

An example of this kind of use of case studies appears in Alejandro Portes’s collaborative project on institutional development in Latin America (Portes and Smith 2008). In each country, the project studied the same five agencies. The research team listed several organizational characteristics that prior theories suggested might be important. In the truth table, the characteristic on which the successful agencies clustered was having a merit system for making personnel decisions. Having a merit system distinguished the successful agencies from the unsuccessful agencies in each of the five country settings in which the research took place. (A slightly different design would have allowed the researchers to determine whether an antecedent condition shaped the adoption of merit systems in the successful cases and also exercised an independent effect on the outcome.)

In the ISS project about single-agency turnarounds, the aim was to make some tentative general statements about the robustness of a set of practices to differences in context. Specifically, the claim was that delays would diminish and productivity would rise by introducing a fairly standard set of management practices designed to streamline a process, increase transparency, and invite friendly group competition. In this kind of observational study, the authors had a before-and-after or longitudinal design in each individual case, which was married with a cross-sectional design.¹⁴ The elements of the intervention were arrayed in a truth table and examined to see which of them were present or absent in parallel interventions in a number of other cases. The team added cases with nearly identical interventions but different underlying country contexts. ISS then explored each case in greater detail to see whether implementation strategy or something else having to do with context explained which reforms were successful and which were not.

¹⁴ In the best of all possible worlds, we would want to draw the cases systematically from a known universe or population, but the absence of such a dataset meant we had to satisfice and match organizations on function while varying context means. Conclusions reached thus need to be qualified by the recognition that there could be more cases “out there,” which, if included in the analysis, might alter the initial results.

Small-scale observational studies (the only type of study possible in many subject areas) suffer from a variety of threats, including inability to control for large numbers of differences in setting. However, the interview data and close process tracing helped increase confidence in two respects. First, they helped reveal the connection between the outcomes observed and the practices under study. For example, it was relevant that people in work groups could describe their reactions when a poster showing how many identity documents they had issued had increased or decreased compared to the month before. Second, the information the interviews delivered about obstacles encountered and workarounds developed fueled hypotheses about robustness to changes in setting. In short, the deep dive that the case study permitted helped alleviate some of the inferential challenges that inevitably arise when there are only small numbers of observations and a randomized controlled trial is not feasible.

Rare events pose special problems for theory testing. Organizations must often learn from single cases – for example, from the outcome of a rare event (such as a natural disaster, or a major restructuring). In this circumstance it may be possible to evaluate impact across several units within the organization or influences across policy areas. However, where this approach is impossible few organizations decline to learn from experience; instead, they look closely at the history of the event to assess the sequence of steps by which prevailing outcomes obtained and how these might have been different had alternative courses of action been pursued.

1.7 Understanding Outliers or Deviant Cases

A common and important use of case studies is to explore the case that does not conform to expectations. An analysis comparing a large number of cases on a few variables may find that most units (countries, agencies, etc.) cluster closely around a regression line whose slope shows the relationship between the causal variables and the outcome. However, one or two cases may lie far from the line. We usually want to know what's different about those cases, and especially how and why they differ. For example, there is generally a quite robust relationship between a country's level of spending on education and the quality of outcomes that country's education system generates. Why is Vietnam in the bottom third globally in terms of its spending on education, yet in

the upper third globally in terms of outcomes (as measured by student performance on standardized examinations)? Conversely, why is Malaysia in the upper third on spending and bottom third on outcomes?

In the study of development, outliers such as these hold particular fascination. For example, several scholars whose contributions are ordinarily associated with use of quantitative methods have employed schematic case studies to ponder why Botswana seems to have stronger institutions than most other African countries (Acemoglu, Johnson, and Robinson 2003). Costa Rica and Singapore attract attention for the same reason.¹⁵ This same approach can be used to explore and explain subnational variation as a basis for deriving policy lessons. Brix, Lust, and Woolcock (2015), for example, deploy data collected from household surveys to map the wide range of outcomes in public service delivery across countries in the Middle East and North Africa – countries which otherwise have highly centralized line ministries, which means roughly the same policies regarding (say) health and education apply across any given country. The wide variation in outcomes is thus largely a matter of factors shaping policy *implementation*, which are often highly contextual and thus much harder to assess via standard quantitative instruments. On the basis of the subnational variation maps, however, granular case studies were able to be prepared on those particular locations where unusually high (and low) outcomes were being obtained; the lessons from these cases, in turn, became inputs for a conversation with domestic policy-makers about where and how improvements might be sought. Here, the goal was not to seek policy reform by importing what researchers deemed “best practices” (as verified by “rigorous evidence”) from abroad but rather to use *both* household surveys and case studies to endogenize research tools into the ways in which local practitioners make difficult decisions about strategy, trade-offs, and feedback, doing so in ways regarded as legitimate and useful by providers and users of public services.

¹⁵ The ISS program began with a similar aim. The questions at the heart of the program were “What makes the countries that pull off institutional transformation different from others? What have they done that others could do to increase government capacity? What can be learned from the positive deviants, in particular?” For a variety of reasons having to do with the nature of the subject matter, the program disaggregated the subject and focused on responses to particular kinds of strategic challenges within countries and why some had negotiated these successfully in some periods and places but not in others.

1.8 Ensuring Rigor in Case Studies: Foundations, Strategies, and Applications

There is general agreement on some of the standards that should govern qualitative case studies. Such studies should:¹⁶

- respond to a clear question that links to an important intellectual debate or policy problem
- specify and define core concepts, terms, and metrics associated with the explanations
- identify plausible explanations, articulating a main hypothesis and logical alternatives
- offer data that allow us to evaluate the main ideas or discriminate between different possible causal mechanisms, including any that emerge as important in the course of the research
- be selected according to clear and transparent criteria appropriate to the research objective
- be amenable to replication – that is, other researchers ought to be able to check the results

Together, this book's three parts – on Internal and External Validity Issues, Ensuring High-Quality Case Studies, and Applications to Development Practice – explore how the content and realization of these standards can be applied by those conducting case studies in development research and practice, and how, in turn, the fruits of their endeavors can contribute to a refinement and expansion of the “ecologies of evidence” on which inherently complex decisions in development are made.

We proceed as follows. Part I focuses on the relative strengths and weaknesses of qualitative cases versus frequentist observational studies (surveys, aggregate data analysis) and randomized controlled trials (RCTs). Its constituent chapters explore the logic of causal inference and the logic of generalization, often framed as problems of internal and external validity.

In Chapter 2, philosopher of science Nancy Cartwright walks us through the logic behind RCTs on the one hand, and qualitative case studies on the other. RCTs have gained considerable prominence as a ‘gold standard’ for

¹⁶ These general standards, importantly, are consistent with a recent interdisciplinary effort to define rigor in case study research, which took place under the auspices of the US National Science Foundation. See Report on the Workshop on Interdisciplinary Standards for Systematic Qualitative Research. Available at: https://oconnell.fas.harvard.edu/files/lamont/files/issqr_workshop_rpt.pdf.

establishing whether a given policy intervention has a causal effect, but what do these experiments actually tell us and how useful is this information for policy-makers? Cartwright draws attention to two problems. First, an RCT only establishes a claim about average effects for the population enrolled in an experiment; it tells us little about what lies behind the average. The policy intervention studied might have changed nothing in some instances, while in others it triggered large shifts in behavior or health or whatever is under study. But, second, an RCT also tells us nothing about when we might expect to see the same effect size in a different population. To assess how a different population might respond requires other information of the sort that qualitative case studies often uncover. RCTs may help identify a cause, but identifying a cause is not the same as identifying something that is generally true, Cartwright notes. She then considers what information a policy-maker would need to predict whether a causal relationship will hold in a particular instance, which is often what we really want to know.

The singular qualitative case study has a role to play in addressing this need. Cartwright begins by asking what are the support factors that enable the intervention to work, and are they present in a particular situation? She suggests we should use various types of evidence, both indirect and direct. In the “direct” category are many of the elements that case studies can (and should) document: 1) Does O occur at the time, in the manner, and of the size to be expected that T caused it? 2) Are there symptoms of cause – by-products of the causal relationship? 3) Were requisite support factors present? (i.e., was everything in place that needed to be in order for T to produce O?), and 4) Were the expected intermediate steps (mediator variables) in place? Often these are the key elements we need to know in order to decide whether the effects observed in an experiment will scale.

Political scientist Christopher Achen also weighs the value of RCTs versus qualitative case studies with the aim of correcting what he perceives as an imbalance in favor of the former within contemporary social science. In Chapter 3 he shows that “the argument for experiments depends critically on emphasizing the central challenge of observational work – accounting for unobserved confounders – while ignoring entirely the central challenge of experimentation – achieving external validity.” Using the mathematics behind randomized controlled trials to make his point, he shows that once this imbalance is corrected, we are closer to Cartwright’s view than to the current belief that RCTs constitute the gold standard for good policy research.

As a pivot, Achen takes a 2014 essay, a classic statement about the failure of observational studies to generate learning and about the strengths of RCTs. The authors of that essay argued that

[t]he external validity of an experiment hinges on four factors: 1) whether the subjects in the study are as strongly influenced by the treatment as the population to which a generalization is made, 2) whether the treatment in the experiment corresponds to the treatment in the population of interest, 3) whether the response measure used in the experiment corresponds to the variable of interest in the population, and 4) how the effect estimates were derived statistically. (Gerber et al. 2014, 21)

But Achen finds this list a little too short: “The difficulty is that those assumptions combine jaundiced cynicism about observational studies with gullible innocence about experiments,” he writes. “What is missing from this list are the two critical factors emphasized in the work of recent critics of RCTs: heterogeneity of treatment effects and the importance of context.” For example, in an experiment conducted with Michigan voters, there were no Louisianans, no Democrats, and no general election voters; “[h]ence, no within-sample statistical adjustments are available to accomplish the inferential leap” required for generalizing the result.

Achen concludes: “Causal inference of any kind is just plain hard. If the evidence is observational, patient consideration of plausible counterarguments, followed by the assembling of relevant evidence, can be, and often is, a painstaking process.” Well-structured qualitative case studies are one important tool; experiments, another.

In Chapter 4, Andrew Bennett help us think about what steps are necessary to use case studies to identify causal relationships and draw contingent generalizations. He suggests that case study research employs Bayesian logic rather than frequentist logic: “Bayesian logic treats probabilities as degrees of belief in alternative explanations, and it updates initial degrees of belief (called ‘priors’) by using assessments of the probative value of new evidence vis-à-vis alternative explanations (the updated degree of belief is known as the ‘posterior’).”

Bennett’s chapter sketches four approaches: generalization from ‘typical’ cases, generalization from most- or least-likely cases, mechanism-based generalization, and typological theorizing, with special attention to the last two. Improved understanding of causal mechanisms permits generalizing to individuals, cases, or contexts outside the initial sample studied. In this regard, the study of deviant, or outlier, cases and cases that have high values

on the independent variable of interest (theory of change) may prove helpful, Bennett suggests, aiding the identification of scope conditions, new explanations, and omitted variables.

In “Will it Work Here?” (Chapter 5), Michael Woolcock focuses on the utility of qualitative case studies for addressing the decision-maker’s perennial external validity concern: What works there may not work here. He asks how to generate the facts that are important in determining whether an intervention can be scaled and replicated in a given setting. He focuses our attention on three categories. The first he terms causal density, or whether 1) there are numerous causal pathways and feedback loops that affect inputs, actions, and outcomes, and 2) there is greater or lesser openness to exogenous influence. Experiments are often helpful when causal density is low – deworming, use of malaria nets, classroom size – but they fail when causal density is high, as in parenting. To assess causal density, Woolcock suggests we pay special attention to how many person-to-person transactions are required; how much discretion is required of front-line implementing agents; how much pressure implementing agents face to do something other than respond constructively to the problem; and the extent to which implementing agents are required to deploy solutions from a known menu or to innovate *in situ*.

Woolcock’s two other categories of relevant fact include implementation capability and reasoned expectations about what can be achieved by when. With respect to the first, he urges us not to assume that implementation capacity is equally available in each setting. Who has the authority to act? Is there adequate management capacity? Are there adequately trained front-line personnel? Is there a clear point of delivery? A functional supply chain? His third category, reasoned expectations, focuses on having a grounded theory about what can be achieved by when. Should we anticipate that the elements of an intervention all show results at the same time, as we usually assume, or will some kinds of results materialize before others? Will some increase over time, while others dissipate? Deliberation about these matters on the basis of analytic case studies, Woolcock argues, are the main method available for assessing the generalizability of any given intervention. Woolcock supplements his discussion with examples and a series of useful summary charts.

Part II of the book builds upon these methodological concerns to examine practical strategies by which case studies in international development (and elsewhere) can be prepared to the highest standards. Although not exhaustive, these strategies, presented by three political scientists, can help elevate

the quality and utility of case studies by focusing on useful analytical tools that can enhance the rigor of their methodological foundations.

In Chapter 6, Jennifer Widner, who directs Princeton University's Innovations for Successful Societies program, reflects on what she and others have learned about gathering reliable information from interviews. Case study researchers usually draw on many types of evidence, some qualitative and some quantitative. For understanding motivation/interest, anticipated challenges, strategic choices, steps taken, unexpected obstacles encountered, and other elements of implementation, interviews with people who were "in the room where it happens" are usually essential. There may be diary entries or meeting minutes to help verify personal recall, but often the documentary evidence is limited or screened from view by thirty-year rules. Subject matter, proximity to elections or other sensitive events, interviewer self-presentation, question sequence, probes, and ethics safeguards are among the factors that shape the reliability of information offered in an interview. Widner sketches ways to improve the accuracy of recall and the level of detail, and to guard against "spin," drawing on her program's experience as well as the work of survey researchers and anthropologists.

Political scientist Tommaso Pavone analyzes how our evolving understanding of case-based causal inference via process tracing should alter how we select cases for comparative inquiry (Chapter 7). The chapter explicates perhaps the most influential and widely used means to conduct qualitative research involving two or more cases: Mill's methods of agreement and difference. It then argues that the traditional use of Millian methods of case selection can lead us to treat cases as static units to be synchronically compared rather than as social processes unfolding over time. As a result, Millian methods risk prematurely rejecting and otherwise overlooking (1) ordered causal processes, (2) paced causal processes, and (3) equifinality, or the presence of multiple pathways that produce the same outcome. To address these issues, the chapter develops a set of recommendations to ensure the alignment of Millian methods of case selection with within-case sequential analysis. First, it outlines how the use of processualist theories can help reformulate Millian case selection designs to accommodate ordered and paced processes (but not equifinal processes). Second, it proposes a new, alternative approach to comparative case study research: the method of inductive case selection. By selecting cases for comparison after a causal process has been identified within a particular case, the method of inductive case selection enables researchers to assess (1) the generalizability of the causal sequences, (2) the logics of scope conditions on the causal argument,

and (3) the presence of equifinal pathways to the same outcome. A number of concrete examples from development practice illustrate how the method of inductive case selection can be used by both scholars and policy practitioners alike.

One of the common criticisms of qualitative research is that a case is hard to replicate. Whereas quantitative researchers often share their research designs and their data and encourage one another to rerun their analyses, qualitative researchers cannot as easily do so. However, they can enhance reliability in other ways. In Chapter 8, Andrew Moravcsik introduces new practices designed to enhance three dimensions of research transparency: *data transparency*, which stipulates that researchers should publicize the data and evidence on which their research rests; *analytic transparency*, which stipulates that researchers should publicize how they interpret and analyze evidence in order to generate descriptive and causal inferences; and *production transparency*, which stipulates that social scientists should publicize the broader set of design choices that underlie the research. To respond to these needs, Moravcsik couples technology with the practice of discursive footnotes common in law journals. He discusses the rationale for creating a digitally enabled appendix with annotated source materials, called Active Citation or the Annotation for Transparency Initiative.

Part III – this volume’s concluding section – explores the ways in which case studies are being used today to learn from and enhance effectiveness in different development agencies.

In Chapter 9, Andrew Bennett explores how process tracing can be used in program evaluation. “Process tracing and program evaluation, or contribution analysis, have much in common, as they both involve causal inference on alternative explanations for the outcome of a single case,” Bennett says:

Evaluators are often interested in whether one particular explanation – the implicit or explicit theory of change behind a program – accounts for the outcome. Yet they still need to consider whether exogenous nonprogram factors . . . account for the outcome, whether the program generated the outcome through some process other than the theory of change, and whether the program had additional or unintended consequences, either good or bad.

Bennett discusses how to develop a process-tracing case study to meet these demands and walks the reader through several key elements of this enterprise, including types of confounding explanations and the basics of Bayesian analysis.

In Chapter 10, with a focus on social services in the Middle East, political scientist Melani Cammett takes up the use of positive deviant cases – examples of sustained high performance in a context in which good results are uncommon – to identify and disentangle causal complexity and understand the role of context. Although the consensus view on the role of deviant cases is that they are most useful for exploratory purposes or discovery and theory building, Cammett suggests they can also generate insights into the identification and operation of causal mechanisms. She writes that “analyses of positive deviant cases among a field of otherwise similar cases that operate in the same context . . . can be a valuable way to identify potential explanatory variables for exceptional performance.” The hypothesized explanatory variables can then be incorporated in subsequent quantitative or qualitative studies in order to evaluate their effects across a broader range of observations. The chapter discusses how to approach selection of positive deviant cases systematically and then works through a real example.

In Chapter 11, on “Analytical Narratives and Case Studies,” Margaret Levi and Barry Weingast focus on a particular type of case in which the focus is on an outcome that results from strategic interaction, when one person’s decision depends on what another does. “A weakness of case studies per se is that there typically exist multiple ways to interpret a given case,” they begin. “How are we to know which interpretation makes most sense? What gives us confidence in the particular interpretation offered?” An analytic narrative first elucidates the principal players, their preferences, key decision points and possible choices, and the rules of the game. It then builds a model of the sequence of interaction including predicted outcomes and evaluates the model through comparative statics and the testable implications the model generates. An analytic narrative also models situations as an extensive-form game. “The advantage of the game is that it reveals the logic of why, in equilibrium, it is in the interest of the players to fulfill their threats or promises against those who leave the equilibrium path,” the authors explain. Although game theory is useful, there is no hard rule that requires us to formalize. The particular findings do not generalize to other contexts, but an analytic narrative points to the characteristics of situations to which a similar strategic logic applies.

The book’s final chapters focus on the use of case studies for refining development policy and practice – in short, for learning. In Chapter 12, Sarah Glavery and her coauthors draw a distinction between explicit knowledge, which is easily identified and shared through databases and reports, and tacit knowledge – the less easily shared “know how” that comes with having

carried out a task. The chapter explores ways to use case study preparation, as well as a case itself, as a vehicle for sharing “know how,” specifically with respect to program implementation. It considers the experiences of four different types of organizations that have used case studies as part of their decision-making as it pertains to development issues: a multilateral agency (the World Bank), a major bilateral agency (Germany’s GIZ), a leading think tank (Brookings), and a ministry of a large country (China’s Ministry of Finance), which are all linked through their involvement in the GDI.

Finally, in Chapter 13, Maria Gonzalez and Jennifer Widner reflect more broadly on the intellectual history of a science of delivery and adaptive management, two interlinked approaches to improving public services, and the use of case studies to move these endeavors forward. They emphasize the ways in which case studies have become salient tools for front-line staff whose everyday work is trying to solve complex development challenges, especially those pertaining to the implementation of policies and projects, and how, in turn, case studies are informing a broader turn to explaining outcome variation and identifying strategies for responding to complex challenges and ultimately seeking to enhance development effectiveness. The chapter discusses seven qualities that make a case useful to practitioners, and then offers reflections on how to use cases in a group context to elucidate core ideas and spark innovation.

1.9 Conclusion

In both development research and practice, case studies provide unique insights into implementation successes and failures, and help to identify why and how a particular outcome occurred. The data collected through case studies is often richer and of greater depth than would normally be obtained by other research designs, which allows for (potentially) richer discussions regarding their generalizability beyond the defined context of the case being studied. The case study method facilitates the identification of patterns and provides practical insights on how to navigate complex delivery challenges. Case studies can also capture the contextual conditions surrounding the delivery case, trace the detailed dynamics of the implementation process, provide key lessons learned, and inform broader approaches to service delivery (e.g., by focusing attention on citizen outcomes, generating multidimensional responses, providing usable evidence to enhance real-time implementation, and supporting leadership for change).

The core idea behind recent initiatives seeking to expand, formalize, and catalogue case studies of development practice is that capturing implementation processes and building a cumulative body of operational knowledge and know-how can play a key role in helping development practitioners deliver better results. Systematically investigating delivery in its own right offers an opportunity to distill common delivery challenges, and to engage constructively with the nontechnical problems that often hinder development interventions and prevent countries and practitioners from translating technical solutions into results on the ground.

Doing this well, however, requires drawing on the full array of established and leading approaches to conducting case study research. As this volume seeks to show, the last twenty years have led to considerable refinements and extensions of prevailing practice, and renewed confidence among scholars of case study methods that they have not merely addressed (or at least identified defensible responses to) long-standing concerns regarding the veracity of case studies but actively advanced those domains of inquiry in which case studies enjoy a distinctive epistemological ‘comparative advantage’. In turn, the veritable explosion of case studies of development processes now being prepared by academic groups, domestic governments, and international agencies around the world offers unprecedented opportunities for researchers to refine still further the underlying techniques, methodological principles, and theory on which the case study itself ultimately rests. As such, the time is ripe for a mutually beneficial dialogue between scholars and practitioners of development – a dialogue we hope this volume can inspire.

References

- Acemoglu, D., Johnson, S., and Robinson, A. (2003) “Botswana: An African success story” in Rodrik, D. (ed.) *In search of prosperity: Analytic narratives on economic growth*. Princeton, NJ: Princeton University Press, pp. 80–119.
- Acemoglu, D. and Robinson, J. A. (2012) *Why nations fail: The origins of power, prosperity and poverty*. New York: Crown Business.
- Bamberger, M., Rao, V., and Woolcock, M. (2010) “Using mixed methods in monitoring and evaluation: Experiences from international development” in Tashakkori, A. and Teddlie, C. (eds.) *Handbook of mixed methods in social and behavioral research* (2nd revised edition). Thousand Oaks, CA: Sage Publications, pp. 613–641.
- Barma, N., Huybens, E., and Viñuela, L. (2014) *Institutions taking root: Building state capacity in challenging contexts*. Washington, DC: World Bank.

- Bates, R. H., Avner Greif, M. L., Rosenthal, J. L., and Weingast, B. (eds.) (1998) *Analytic narratives*. Princeton, NJ: Princeton University Press.
- Bayly, C. A., Rao, V., Szreter, S., and Woolcock, M. (eds.) (2011) *History, historians and development policy: A necessary dialogue*. Manchester: Manchester University Press.
- Bennett, A. and Checkel, J. T. (eds.) (2015) *Process tracing: From metaphor to analytic tool*. New York: Cambridge University Press.
- Bliss, C. and Stern, N. (1982) *Palanpur: The economy of an Indian village*. Oxford: Oxford University Press.
- Brixi, H., Lust, E., and Woolcock, M. (2015) *Trust, voice and incentives: Learning from local success stories in service delivery in the Middle East and North Africa*. Washington, DC: World Bank.
- Byrne, D. and Ragin, C. C. (eds.) (2009) *The Sage handbook of case-based methods*. Los Angeles: Sage Publications.
- Carpenter, D. (2001) *The forging of bureaucratic autonomy: Reputations, networks and policy innovation in executive agencies, 1862–1928*. Princeton, NJ: Princeton University Press.
- Ellickson, R. (1991) *Order without law: How neighbors settle disputes*. Cambridge, MA: Harvard University Press.
- Estabridis, C. A. and Nieto, D. O. (2015) *How to overcome communications and cultural barriers to improve service provision to indigenous populations*. Washington, DC: World Bank, Global Delivery Initiative Case Study.
- Feagin, J. R., Orum, A. M., and Sjoberg, G. (eds.) (1991) *A case for the case study*. Chapel Hill: University of North Carolina Press.
- Flyvbjerg, B. (2006) "Five misunderstandings about case-study research," *Qualitative Inquiry*, 12(2), 219–245.
- George, A. and Bennett, A. (2005) *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- Gerring, J. (2017) *Case study research: Principles and practices, 2nd ed.* New York: Cambridge University Press.
- Glavey, S. and Haas, O. (2015) *How to scale up rural sanitation service delivery in Indonesia*. Washington, DC: World Bank.
- Goertz, G. and Mahoney, J. (2012) *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton, NJ: Princeton University Press.
- Haggard, S. (1992) *The politics of economic adjustment: International constraints, distributive conflicts, and the state*. Princeton, NJ: Princeton University Press.
- Hima, H. and Santibanez, C. (2015) *Against the current: How to shape an enabling environment for sustainable water service delivery in Nigeria*. Washington, DC: World Bank.
- Kaufman, H. (1960) *The forest ranger: A study in administrative behavior*. Washington, DC: Resources for the Future.
- Knodel, J. (1997) "A case for nonanthropological qualitative methods for demographers," *Population and Development Review*, 23(4), 847–853.
- Lanjouw, P., Murgai, R., and Stern, N. (2013) "Nonfarm diversification, poverty, economic mobility, and income inequality: A case study in Village India," *Agricultural Economics*, 44 (4–5), 461–473.
- Lanjouw, P. and Stern, N. (1998) *Economic development in Palanpur over five decades*. Oxford: Oxford University Press.

- Lieberman, E. S. (2015) "Nested analysis: Toward the integration of comparative-historical analysis and other social science methods" in Mahoney, J. and Thelen, K. (eds.). *Advances in comparative-historical analysis*. New York: Cambridge University Press, pp. 240–263.
- Mahoney, J. (2000) "Strategies of causal inference in small-n analysis," *Sociological Methods & Research*, 28(4), 387–424.
- Mahoney, J. and Thelen, K. (eds.) (2015) *Advances in comparative-historical analysis*. New York: Cambridge University Press.
- March, J. G., Sproull, L. S., and Tamuz, M. (1991) "Learning from samples of one or fewer," *Organization Science*, 2(1), 1–13.
- Morgan, M. (2012) "Case studies: One observation or many? Justification or discovery?" *Philosophy of Science*, 79(5), 667–677.
- Ortega Nieto, D. and Parida, I. (2015) *How to ensure quality healthcare and coverage to uninsured populations: Argentina's Plan Nacer/Programa Sumar*. Washington, DC: World Bank.
- Portes, A. and Smith, L. D. (2008) "Institutions and development in Latin America: A comparative analysis," *Studies in Comparative International Development*, 43(2), 101–128.
- Ragin, C. and Becker, H. (1992) *What is a case? Exploring the foundations of social inquiry*. New York: Cambridge University Press.
- Rohlfing, I. (2012) *Case studies and causal inference: An integrative framework*. Basingstoke: Palgrave Macmillan.
- Rueschemeyer, D. (2009) *Useable theory: Analytic tools for social and political research*. Princeton, NJ: Princeton University Press.
- Ruzzene, A. (2012) "Drawing lessons from case studies by enhancing comparability," *Philosophy of the Social Sciences*, 42(1), 99–120.
- Singh, J. P. and Woolcock, M. (eds.) (forthcoming) Special issue of *Global Perspectives* on the future of multilateralism and global development.
- Stake, R. E. (1995) *The art of case study research*. Thousand Oaks, CA: Sage Publications.
- Tamanaha, B., Sage, C., and Woolcock, M. (eds.) (2012) *Legal pluralism and development policy: Scholars and practitioners in dialogue*. New York: Cambridge University Press.
- Tendler, J. (1997) *Good government in the tropics*. Washington, DC: Johns Hopkins University Press.
- Thelen, K. and Mahoney, J. (eds.) (2009) *Explaining institutional change: Ambiguity, agency and power*. New York: Cambridge University Press.
- Van Noorden, R., Maher, B. and Nuzzo, R. (2014) "The top 100 papers: Nature explores the most cited research of all time," *Nature*, October 29. Available at www.nature.com/news/the-top-100-papers-1.16224 (accessed January 19, 2020).
- Wade, R. (1985) "The market for public office: Why the Indian state is not better at development," *World Development*, 13(4), 467–497.
- Woolcock, M. (2013) "Using case studies to explore the external validity of 'complex' development interventions," *Evaluation*, 19(3), 229–248.
- Yin, R. K. (1984) *Case study research: Design and methods*. Beverly Hills, CA: Sage Publications. [Now in its sixth edition, 2017.]