PART I

Context

1 *Introduction: Experiments in Public Management Research*

OLIVER JAMES, SEBASTIAN JILKE, AND GREGG G. VAN RYZIN

Introduction

There is an emerging experimental approach to public management research that is reflected in a substantial increase in published studies using this method. Experimental studies are becoming more common across a broad range of topic areas in public management and, relatedly, public organisations and public services.[1] This trend is, in part, a response to increasing recognition of the limitations of non-experimental, so-called observational, methods, including the analysis of surveys or administrative data. In many contexts observational methods risk providing ambiguous or even misleading evidence about causal relationships. In contrast, experiments of the kind we focus on in this book involve active interventions (sometimes termed *treatments*) by researchers, with randomly assigned treatment conditions to experimental subjects, accompanied by outcome measures, in order to produce more valid evidence about cause and effect.[2]

This book develops an approach to experimentation that recognises the distinctive set of issues about their use in public management as a discipline and area of professional practice.[3] In pursuit of this goal, this book takes stock of the current emerging interest in public

[1] We treat public management as synonymous with public administration whilst acknowledging that the term *public administration* is sometimes used to emphasise the role of democratic processes and constitutional procedures whereas the term *public management* is sometimes used to emphasise managerial structures and behaviour, often seen as generic across public and private sectors (Hood 2005). This book considers experiments and both sets of topic areas.

[2] Broader definitions of experiments do not require random assignment of the intervention and/or require only some form of exogenous treatment (rather than additionally requiring this intervention to be implemented by researchers). For a discussion, see Shadish, Cook, and Campbell (2002).

[3] We acknowledge that there is long-running debate about public management as a distinct discipline, although there has been increasing consensus in recent

management experiments and associated debates about their proper role. It seeks to show how experimental methods can be most suitably advanced in a way that reflects the distinctive topic areas, interest in theories, research practices, and ambitions of public management as a discipline. In particular, public management aspires to be a design science, informing policy making and the practice of public management, and evidence from experimentation is especially useful in this regard. However, the practices and institutions of public management need to be modified in order to take advantage of the opportunities offered by experimentation and this book sets out some reforms in support of this agenda.

Public management researchers develop and assess theory using different methods and present and discuss findings to build cumulative knowledge about public management topic areas. Looking at experiments in related disciplines is especially instructive because of public management's interdisciplinary characteristics, drawing as it does on management, political science, law, psychology, economics, and sociology. Some of the contemporary interest in experiments has been triggered by the use of the method in several of these disciplines. For example, books have been written about experiments as an increasing trend in political science (Morton and Williams 2010; Druckman et al. 2011) and economics (Frechette and Schotter 2015; Friedman 2010; Guala 2005). There has also been increased use of experiments in generic management research, although, despite calls for more, there are still relatively few experimental studies (Colquitt 2008). In contrast, the experimental method is long established in psychology (Field and Hole 2003) and the health sciences (Friedman, Furberg, and DeMets 2010; Matthews 2006). Insights from these disciplines are helpful, but can leave public management scholars wondering about how best to apply experiments to the issues of most direct interest to them. It is a good time for public management to consider the challenges and opportunities presented by the use of experimental methods.

Interest in the experimental approach to public management comes not only from the practice of current social sciences more generally, it has deep roots within our own discipline. Some form of experimentation has

decades about the need for use of systematic methods to gather evidence in order to generate cumulative knowledge(for an overview, see Wright 2015).

been part of the practice of public management and administration research from the beginning of academic study in this area. As Meier and Funk discuss in Chapter 3 on the 'classical roots' of public administration, Frederick W. Taylor's scientific management studies at the turn of the twentieth century were experimental, in the sense of intervening or manipulating various working conditions in order to measure their effects on productivity, although they predated the modern use of random assignment. The use of experiments is especially relevant to public management because of its ambitions to produce useable knowledge to inform policy making and the practice of public management. Herbert Simon (1946) advocated an experimental approach to administration as a design science, a term Simon himself coined to describe a science aimed at finding institutional, policy, and management design solutions to practical, real-world problems. Experiments make this contribution through producing reliable estimates of the causal effects of public management policy and practices. The relationship between experimental researchers and policy makers is often by necessity a close one, especially where experiments are undertaken in naturalistic 'field' contexts. Such field experiments typically require active collaboration to be able to make the interventions necessary for the experiment.

Despite the clear reference to experimentation in its classical roots, and its relevance for informing policy and practice, the history of public management has shown only limited use of experimental methods. In Chapter 2, Li and Van Ryzin's systematic review of the literature shows that only one or two experimental studies have been published on an annual basis for much of the past few decades in the 20 leading public management journals. However, in more recent years, the number of published experimental studies in these journals increased rapidly (see also Anderson and Edwards 2014; Bouwman and Grimmelikhuijsen 2016; Margetts 2011). Much of the work from this recent flourishing of experimental studies is represented in the pages of this volume. At the same time, calls for experimental work in public management have become more frequent (Anderson and Edwards 2015; Blom-Hansen, Morton, and Serritzlew 2015; Bozeman and Scott 1992; Brewer and Brewer 2011; Jilke, Van de Walle, and Kim 2016; Margetts 2011; Perry 2012). Experimental methods now appear more prominently in textbooks in the field than has been the case historically, for example in Van Thiel (2014), McNabb (2015), and Remler and Van Ryzin (2015). Furthermore, the

strengths and limitations of experiments are currently the subject of much active debate (Baekgaard et al. 2015; Jilke et al. 2016; Walker, James, and Brewer 2017).

By highlighting the contributions and prospects of experimental methods in public management research, and by emphasising their advantages for probing causal effects, this book might risk conveying the message that experiments are a generally preferred or superior approach in contrast to all other methods. This is not our position, and different methods, including experimental methods, have their strengths and weaknesses for different types of research (see also Haverland and Yanow 2012). Indeed, the chapters in Part IV discuss the main limitations of using experiments in public management research. Moreover, public management research addresses a remarkably wide range of issues and problems – including normative and public value issues, historical traditions, legal foundations, and organisational culture – that require a diverse range of methods to fully understand and explain. Observational methods of both quantitative and qualitative character are often required, including using historical methods or case studies. Sometimes description rather than identifying cause and effect relations is the main focus of research, whether this is done by traditional methods or by complex correlations between multiple variables in the analysis of so-called big data. The increased use of experiments does not rule out these forms of analysis when best suited to the research question. We do, however, maintain that experiments are currently under-utilised and have much potential to add to the set of methods employed to investigate key issues in public management.

We advocate an approach to the use of experiments that views them as an important additional research method. However, there are several methods that are often described using the term *experiment*, as set out in Table 1.1. The experiments that we primarily focus on in this book, and whose design and analysis are discussed more fully in Chapter 4, have three main elements. First, an intervention/treatment on experimental units/participants that is undertaken by the researcher. Second, random allocation of the treatment to a group receiving it and a 'control' group not receiving it (or the use of multiple different treatment, or placebo groups). Third, the measurement of outcomes and comparison across the different groups. We particularly focus on elaborating the use of three different types of experiment, those conducted in the laboratory, field experiments

Table 1.1: ***Methods containing the term*** **experiment** ***(those in bold are the main focus of***

	Intervention by researcher and comparison with control or across interventions	Random allocation of intervention
Laboratory experiment	Yes	Yes
Field experiment	Yes	Yes
Survey experiment	Yes	Yes
Natural experiment*	No	Yes
Quasi-experiment**	No	No

Notes: *A strict definition requires random allocation but conducted by someone other than resear
and Green 2012). An alternative definition allows allocation to be equivalent to random (Shadish, C
** Sometimes the comparison is between policies implemented or not in different places in a process a
typically not, with some authors emphasising research focused on policy interventions that could
2002).

conducted in more naturalistic domains, and experiments with treatments embedded in surveys.

A distinction is often made between the experimental method and non-experimental, observational methods. However, increasingly, similarities between aspects of experiments and a variety of related types of method are recognised. Table 1.1 sets out the relationship between the three main types of experiment in this book and so-called natural and quasi-experiments. There is a degree of inconsistency in the use of these latter two terms in the literature, but they are closely related to experiments.

Natural experiments are usually considered as not involving an intervention by researchers but where an event can be considered to be allocated at random. Some authors (Gerber and Green 2012: 15) insist on a strict definition of natural experiments, such that the process has to be exactly random but not conducted by researchers themselves, for example when the US government conducted the Vietnam draft lottery. Others use the term to describe a naturally occurring contrast between a treatment and comparison condition which allows a comparison but does not necessarily involve random allocation (Shadish, Cook, and Campbell 2002: 12–17). These latter authors also stress the non-manipulability of the cause of a naturally occurring 'experiment', for example an earthquake creates a shock that can be used to contrast areas that are affected with those that are unaffected. This form of natural experiment does not involve researchers intervening in the world to make something happen, but instead the research design takes advantage of events that occur in ways argued to closely resemble randomisation.

Quasi-experiments involve near random processes, although not random assignment by researchers or policy makers, that cause some units under investigation to receive a treatment but not others. For example, the narrow winners or losers of an election are sometimes considered separated as if random, although whether this is equivalent to random allocation is often debated. Shadish, Cook, and Campbell (2002) discuss the historical use of quasi-experiments and stress that the interest is often in causes that potentially could be manipulated to inform policy making. They also emphasise that, whilst quasi-experiments lack true randomisation, researchers often have considerable control over how they define comparison groups and measure outcomes.

We do not include extensive discussion of natural or quasi-experimental methods, or studies that utilise them, in this book because of the variety of approaches. The methods raise distinctive issues of their own (for a more in depth discussion, Dunning 2012; Gerber and Green 2012; Shadish et al. 2002). However, the methods are referred to at several points in this book, including in Meier and Funk's chapter on classical roots (Chapter 3) and John's chapter on the transformative potential of experiments in public management (Chapter 23). These discussions show that methods often use some but not all elements of the logic of an experiment. A solid foundation in experimental methods is useful for considering the strength of such quasi-experimental designs and evidence, especially as an emerging literature on modern causal inference treats the randomised experiment as a benchmark to assess and interpret non-experimental, observational research more generally (Angrist and Pischke 2014; Imbens and Rubin 2015; Morgan and Winship 2014; Pearl 2000). In this way, experimental methods can play a role in public management research, even when they are not themselves directly used, by alerting researchers and reviewers to issues involved in trying to draw valid conclusions about causes and effects from different research designs.

In the rest of this introduction we develop the rationale for using experiments in public management research. First, we set out the important contribution of experiments for assessing public management theory and contributing to theory development across a wide range of topics. Second, we describe how experiments can enhance public management's role as a design science providing guidance to policy makers about policies, management practices, and programmes. Third, we discuss implications for the conduct of the discipline and its institutions. The concluding section provides an overview of the structure of this book.

1.1 The Contribution to Public Management Theory

Public management researchers are interested in many different types of questions. These include normative questions about values and ethical conduct (Dobel 2005) or questions of due process and legality (Drewry 2003). However, much research in public management is focused on empirical questions about causal effects (Remler and Van

Ryzin 2015). Typical examples include: Does a management training programme improve managerial performance? What is the effect of a system of performance-pay on public employees' work performance? Such research questions about causal effects are very well suited to be answered by experiments that, if properly implemented, can produce evidence with strong internal validity in the sense of convincing inferences about the causal relationship from the manipulated cause to the measured effect (Shadish et al. 2002: 38). Indeed, the experimental turn in public management research in part reflects growing unease about the internal validity of evidence about causality from observational studies, on which public management traditionally has been heavily reliant. As discussed in Chapter 4 on causal inference and the design and analysis of experiments, observational studies use data that is not generated from researchers' interventions in experiments with randomisation, but instead use data from measurement or observation of existing phenomena. This data is often incorporated in large-N quantitative analysis, especially using regression, or qualitative case study analysis, to generate empirical knowledge and assess theory. But such studies often have difficulty in isolating or identifying causal effects because of the complex patterns of influence among observed and unobserved variables at work in causal relationships in the social and political world.

Experiments can assist with these difficulties in establishing causal effects to evaluate the empirical implications of a potentially broad and diverse set of public management theories. Public management theories that can be assessed by experimentation have a broad range of objects, for example they can focus on individuals or organisations. At the organisational level of analysis, characteristics of organisational structures, management routines, or less formal practices of working can be varied, for example systems of training or for rewarding public employees. Many theories about individual-level perceptions, beliefs, attitudes, and behaviour, such as the multitude of theories informed by psychology or behavioural economics, can be assessed in experiments with public managers, service users, or citizens as participants (Grimmelikhuijsen et al. 2017; Tummers et al. 2016). Indeed, many of the contemporary experiments in public management directly use or are informed by individual-level psychological theories, for example in a recent journal special issue (Jilke et al. 2016). There are, however, many sources of theory to which experimental methods can be applied

and this current association should not discourage a much broader focus of investigation.

Many of the requirements that theories have to meet in order to make them suitable for evaluation using experiments are consistent with the typical concerns of public management researchers, such as the requirement that theories have clear empirical implications. However, the need to think about experimental design *ex ante* is particularly valuable in focusing researchers' attention on such issues before the research is conducted. The experimental approach helps clarify the causal mechanisms being subject to empirical investigation because researchers need to know what they should be manipulating and how this should be achieved. The mechanisms can help give clear expectations about the direction of effects caused by an intervention and their magnitude. Theories in public management can be formally stated with axioms that allow the development of expectations using logical rules, as is often the case in experimental economics and is discussed at more length in Chapter 19. However, depending on the context and so long as expectations about effects are clear, verbally expressed theories about causal relations are often sufficient and experimentation is not necessarily linked to strong formalism.

Theories have domains for their applicability which set the contexts where particular effects of causes are expected. The need for public management theories to be more explicit about context is increasingly noted (O'Toole and Meier 2015). Discussions in experimental research often refer not only to the internal validity of a study relating to inference about the causal effect within the experiment, but also to the external validity of experimental findings to populations of interest beyond the immediate sample. The findings may apply to a particular sample, more broadly to a population where the sample came from, for example public managers in a given sector in one country, or may apply to still broader populations. Cross-national or cross-contextual replication experiments may help to increase the external validity of experimental evidence, as Chapter 21 discusses, and can help test and refine theories of the institutional or cultural boundary conditions of expected findings (but see also Jilke et al. 2016).

Experiments differ in the domains in which they are conducted. A major distinction is between laboratory experiments and those that are conducted in naturalistic contexts, often called *field experiments*. Field experiments, as discussed in Chapter 5, seek to produce relevant