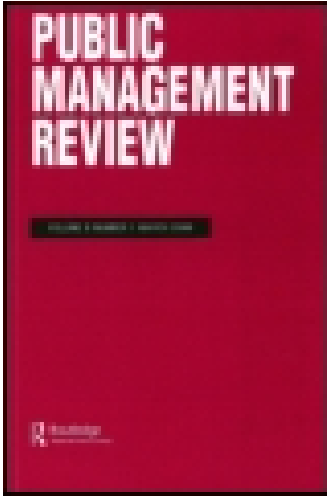


This article was downloaded by: [University of Central Florida]

On: 15 December 2014, At: 09:52

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH,
UK



CrossMark

[Click for updates](#)

Public Management Review

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rpxm20>

Unfulfilled Promise: Laboratory experiments in public management research

Derrick M. Anderson^a & Barry C. Edwards^b

^a School of Public Affairs, Arizona State University, Phoenix, AZ, USA

^b Department of Political Science, University of Georgia, Athens, GA, USA

Published online: 31 Jul 2014.

To cite this article: Derrick M. Anderson & Barry C. Edwards (2014): Unfulfilled Promise: Laboratory experiments in public management research, Public Management Review, DOI: [10.1080/14719037.2014.943272](https://doi.org/10.1080/14719037.2014.943272)

To link to this article: <http://dx.doi.org/10.1080/14719037.2014.943272>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities

whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Abstract

We make the case for increased laboratory experimentation in public management research. Laboratory experiments can generate useful knowledge, particularly in testing causal relationships among constructs of interest. The challenge in this regard is one of identifying the appropriate role for experiments in a greater knowledge production enterprise. Although laboratory experiments are underutilized, they have proffered important knowledge contributions to the field, especially in areas of decision-making and, increasingly, motivation. Because practical problems may pose a greater obstacle to laboratory experimentation in public management than epistemological issues, we address external validity and the cost of conducting laboratory experiments before concluding with suggestions for future research.

Key words

Laboratory experiments, experimental research, decision-making

UNFULFILLED PROMISE

Laboratory experiments in public management research

Derrick M. Anderson and
Barry C. Edwards

Derrick M. Anderson

School of Public Affairs
Arizona State University
Phoenix, AZ
USA
E-mail: derrick.anderson@asu.edu

Barry C. Edwards

Department of Political Science
University of Georgia
Athens, GA
USA
E-mail: bce@uga.edu

2 Public Management Review

The primary purpose of research in the field of public management is generating knowledge. Assessing causal relationships among variables of interest is central to this enterprise. Experiments are the optimal tool for testing causality, and, for this reason, they are a primary method of inquiry in many social science fields. Although interest in experiments appears to be increasing, public management lags other fields with respect to experimental research. We offer this article to encourage more experiments in the field of public management.

We are not alone in calling for more experimental research in the fields of public administration and public management (Margetts 2011; Brewer and Brewer 2011; Perry 2012; Bozeman and Scott 1992). Rather than reiterate the arguments ably advanced by others, this article focuses on the unique potential of one subset of experiments – laboratory experiments – in advancing public management research. ‘Laboratory experiments’ or ‘laboratory studies’ provide for the random assignment of individuals to treatment and control groups and the manipulation of experimental variables in a controlled physical setting.¹ We do not claim that laboratory experiments are the only or best way to conduct public management research, but rather that laboratory experiments are a useful method for examining certain questions about public management. In particular, laboratory experiments can help us responsibly study how public employees are motivated to work, make decisions and interact with the public. We contend that other research methods, including other types of experimental designs, could be inappropriate methods for studying ethically sensitive issues such as trust, stress, credibility, and legitimacy because the risks associated with manipulating public employees, the public, or service provision, for research purposes are too great.

This article begins with a brief discussion of knowledge production in public management research. Next, we highlight the importance of causal inference for scholarly and practical public management research. We review a selection of the field’s previously published laboratory experiments. Although public management’s experimental record is limited, these works demonstrate how scholars can study sensitive causal relationships responsibly in controlled laboratory settings. We also address whether experiments with student subjects generalize to other populations and the costs of conducting laboratory research before concluding with some suggestions for future research.

PUBLIC MANAGEMENT RESEARCH SHOULD GENERATE USEFUL KNOWLEDGE THROUGH EXPERIMENTATION

In ideal circumstances, public management research speaks directly to the practice of public management (Perry 2012). Accordingly, research in the field can be evaluated pursuant to the prospects of producing ‘useable knowledge’ (Perry and Kraemer 1986, 215). However, empirical evidence on social science knowledge utilization – and theories such as Caplan’s (1979) ‘two communities theory’ – suggests a gap between

social science knowledge production and its practical applications. Such observations have particular relevance in case of public administration and management where confusion as to the status of the field as an art, science, or profession has lingered for decades (Perry and Kraemer 1986; Raadschelders 2011; Lynn 1996). While this so-called gap can be attributed to many factors, neglect of certain modes of inquiry, including laboratory experiments, may inhibit the field's capacity to produce high-quality, useable knowledge.

The scientific aspiration of public management research is to establish the extent to which causal relationships exist among variables of interest. In turn, it has been observed that significant causal relationships may be exploited for management and policy purposes (Goggin 1986, 334–335). For example, public management research can generate useful knowledge about decision-making by assessing how different informational inputs affect decision outcomes. Research along these lines potentially helps practitioners make better use of the information and resources at their disposal, engage participants in a constructive manner, and design decision-making processes that optimize quality, accuracy, and confidence.

The prevailing framework under which scientific notions of causality are defined necessitates the simultaneous existence of three criteria when explaining causal relationships: correlation between variables, appropriate time order relationship between cause and effect (the cause precedes the effect in time), and non-spurious relationships between cause and effect (Cook and Campbell 1979). These are the standards by which research designs are measured even when their implementation is impossible or infeasible (Singleton, Straits, and Straits 2009).

Though some traditions and perspectives appear more frequently than others do, public management research does not follow a single methodological or epistemic tradition. As a result, public management has been broadly criticized for lacking methodological sophistication (Gill and Meier 2000). But the view taken here is that the absence of significantly entrenched modes of discovery can also be seen as an opportunity for rapid progress. This condition affords the field considerable methodological flexibility. This flexibility may allow the field to successfully incorporate laboratory experiments into the set of methods used to generate useful information.

Controlled experiments minimize the likelihood that research findings will be attributed to extraneous or omitted variables (Singleton, Straits, and Straits 2009, 195).² Experiments are widely viewed as the optimal research design for assessing causality (Falk and Heckman 2009; Dobbins, Lane, and Steiner 1988).³ For these reasons, experimental research is frequently published in fields related to public management, including business, political science, and economics (see Scandura and Williams 2000; Druckman et al. 2006; and Falk and Heckman 2009 for overview of experimentation in these fields).

Of course, there are different types of experiment, many of which have surfaced in the public management literature. For example, an increasing amount of recent

4 Public Management Review

scholarship in public administration and management utilizes survey and field experiments (e.g. Avellaneda 2013; James 2011a, 2011b; Jakobsen 2013; Jakobsen and Andersen 2013; Hvidman and Andersen 2014; Van Ryzin 2004, 2006; Clerkin, Paynter, and Taylor 2009; Bellé 2013). Researchers have demonstrated tremendous creativity in field experiments on subjects such as voter mobilization (Nickerson 2008; Gerber, Green, and Larimer 2008; Arceneaux and Nickerson 2009) and workplace discrimination (Pager 2003; Pager, Western, and Bonikowski 2009; Bertrand and Mullainathan 2004). Additionally, there are also a multitude of rigorous, robust quasi-experimental designs in the literature, including regression discontinuity designs, fixed effects models, instrumental variable models, and matching methods, which help scholars generate useful research.

Simon and Divine (1941) observed that well-executed experiments can generate useful insights into the practice of public management. In the business management literature, leading scholars have argued that reliable research on causal relationships can help scholars offer practical recommendations with greater confidence (Locke and Latham 1990; Colquitt 2008). In public administration, there is at least some evidence that practitioners value causal inferences derived from experimentation. For example, Gano and colleagues (2007) studied the criteria that government agencies use to evaluate information. In their study, some interview respondents expressed high regard for knowledge produced through controlled experiments and cite experiments as 'gold standards' in scientific research (2007, 49).

Historically, experiments have played a minor role in public management research despite the vital role they play in generating knowledge of causal relationships (Bozeman and Scott 1992; Margetts 2011). According to Margetts (2011), leading public administration and management journals have published, on average, less than one article featuring experimental research of any sort (field experiments, quasi-experiments, laboratory experiments, etc.) per year between 1960 and 2009. This trickle of publication led Margetts to observe: '[T]here is as yet little evidence that experimental research is penetrating the mainstream of public management' (195).

Since 2010, we are pleased to report that there has been a noticeable increase in the publication of experimental research in leading public administration and management journals. We counted the number of published experiments in the same journals reviewed by Margetts for the time period 2010 through April, 2014 (the last full month passed at the time of this writing).⁴ One observes in Figure 1a relative surge in the publication of experimental works in recent years. Indeed, 2013 represents a high point in the publication of experimental research work more such works published in that year alone (9) than during some entire decades (i.e., the 1960s, 1980s, or 1990s).

Whether journals sustain this recent increase in the publication of experimental works remains to be seen. However, proponents of experimental research should be encouraged by the fact that there is now some evidence that experimental research is penetrating the mainstream of public administration and management research.

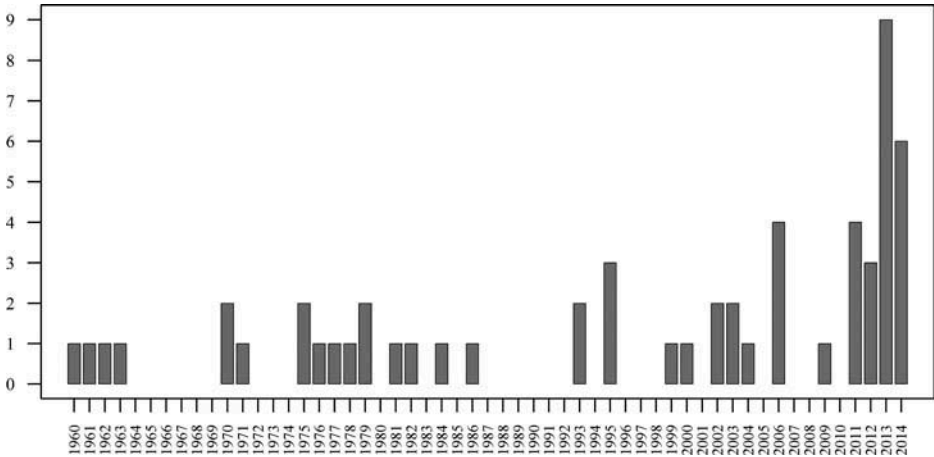


Figure 1: Experiments published in leading journals, 1960 to April, 2014

Notes: Data for 1960–2009 are from Margetts (2011).

It should be noted that some important experimental results have been published in journals of related fields where the tradition of experimentation is more robust (e.g. Miller and Whitford 2002; Bozeman and Shangraw 1988, 1989; Bozeman and McAlpine 1977; Bretschneider, Straussman, and Mullins 1988). To some extent, public management researchers can tap the wealth of experimental work in the field of business administration. This literature offers rich insights into important dynamics that transcend sector boundaries such as task performance (on individual and team levels) and organizational turnover. However, generic approaches to studying management are questionable; the relationship of public managers to the public is not entirely analogous to the relationship of private sector managers to consumers. Accordingly, there is a clear need to test how business management studies fare in the public sector context (see Nutt 2005; Brewer and Brewer 2011 for examples of this type of research).

THE UNIQUE PROMISE OF LABORATORY EXPERIMENTATION

The promise for social scientific experimentation in the domain of public management is robust. Until this point, we have made the case that public management research should rely on causal inference in generating useful knowledge, and experimental methods are ideal for advancing causal inferences. We now turn to the issue of laboratory experimentation. Within the context of promise for experimentation, we think laboratory experiments have a particularly important contribution that has been largely overlooked.⁵ Indeed, little has changed in the more than 20 years since Bozeman

and Scott observed that neglect for laboratory experiments in public administration research was so problematic and widespread that conditions of avoidance 'almost seemed studied' (1992, 293).

What is the role of laboratory experiments? A helpful starting point for answering this question is reflection upon the 'necessary evils' of experiments: experimental research is inherently manipulative, subjects are often deceived, and the experiment may be designed to generate undesirable outcomes, like poor judgements, loss of trust, and cognitive errors.

Deception is frequently used to obtain authentic responses to treatment conditions, but deception makes obtaining informed consent problematic. The subject cannot make a knowledgeable decision to participate in an experiment if he or she is not informed as to the nature of the research. Anyone conducting research using human test subjects must comply with applicable regulations. Although deception is allowed in some circumstances by federal guidelines for research institutions, some social scientists, including economists, oppose deception because it potentially contaminates the pool of research subjects (for further reading on the ethics of deception in experiments, see Sell 2008; Cook and Yamagishi 2008).⁶ This is an area where guidance from an Institutional Review Board is helpful, if not required.⁷

These negative aspects of experimental research – manipulation, deception, and bad outcomes – highlight why some causal relationships are better suited for analysis in the laboratory than the field. Although public managers may be interested in how manipulating employees would affect operations, they should be reluctant to experiment on their employees or the public in manner that may cause irreparable damage.⁸ The idea that public managers would experiment on employees or citizens participating in the democratic process raises serious ethical issues. Cooper (1987), for example, maintains that public administrators serve as fiduciaries for the citizenry and should act in manner consistent with public trust. These ethical obligations would tend to limit the public managers' ability to participate in exploratory field research: '[N]o act could be deemed acceptable on the grounds that it strengthened the organization or furthered the interests of practitioners unless it first produced significant public good' (325).⁹ In addition to ethical issues, manipulating public employees or citizens exercising their democratic rights for the purpose of field-testing theories would raise serious legal issues for public managers. Manipulating citizens' access to information and decision-making processes, for example, would almost certainly violate citizens' due process rights. There are, of course, limits to the types of manipulation and deceptions that are permissible in laboratory research, but the point is we can do things in a controlled laboratory environments that we could not, or would not dare to do, in the real world. In the next section, we demonstrate some of the potential insights of this method of research.

Scholars with experiment-based research ambitions can look to the exemplary features of previous studies. This section provides a review of several noteworthy

laboratory studies, emphasizing each study's design features and findings. The chief, though not exclusive, focus of existing laboratory experiments is decision-making – an important subject of many decades of public management research. Rather than discuss a random or representative sample of experimental works, we focus on works that highlight the particular contribution of laboratory studies in manipulating individual behaviour to explore how public managers can create or destroy trust, credibility, certainty, and other intangible resources.

Individuals and decision-making

Scientists who study decision-making have examined a range of decision-making processes in laboratory experiments. This research focuses on cognitive processes and individual characteristics as predictors of various aspects of decisions. In most of these studies, the dependent variable is an end product decision or an individual's attitude towards this decision.¹⁰

Coursey (1992) based his study on the observation that individual decision-makers express preferences for certain decision-making processes. Coursey proposed that cognitive response theory emerges as a prevailing mechanism only after credibility judgements are made.¹¹ To study the interaction of information and cognition, Coursey engaged graduate students in a hypothetical decision-making environment wherein they were required to consult various types of analyses in making recommendations for potential investments of government resources. Coursey manipulated the quality and credibility of proposals he supplied to test subjects. Findings from this study suggest a minimal influence of credibility in the presence of high perceptions of benefits. Similarly, in instances where benefits are perceived to be low, highly credibility information is preferred or at least more persuasive than information of low credibility. Attempting this type of analysis through observational research is confounded by the likelihood that organizations that produce high- and low-quality information differ in a multitude of dimensions that affect individual decision-making.

In practice, one would not want to knowingly withhold information or supply bad information to manipulate decision-makers; the damage on organizational and public trust would be significant. However, this line of inquiry can be pursued in controlled settings. Recently, James (2011b) engineered a series of experiments to assess how different performance measures affect public evaluation of governmental performance. The impact of truthful, positive information about governmental performance was assessed through a field experiment, but James astutely observed the impropriety of contaminating public perceptions with false information about bad government services (406). Laboratory experiments offered James the opportunity to expose research subjects to both positive and (false) negative performance information and provide participants correct performance information at the conclusion of the experiment. This approach utilized both the control of laboratory experiments and the realism of a field

experiment. Below, we discuss other strategies for designing experiments to produce general, useful information for practitioners.

In a related study of how individuals navigate decision-making scenarios, Landsbergen and colleagues (1992) examined the effects of perceived decision difficulty on the use of various decision criteria. They theorized that internal rationality-based models of decision-making operate amidst an assumption that a rational or good decision can be made. Subjects for this study included both public managers and graduate students. The laboratory protocol required subjects to evaluate a proposed telecommunications programme that would provide subsidized technologies to low-income households. Subjects were provided an analysis of the programme that touched on costs and benefits as well as expert evaluations that invoked different evaluative criteria. Subjects were asked to make policy recommendations and assess the influence of the various expert evaluations. The authors found that decision-makers may appear to make decisions in highly idiosyncratic and unsystematic ways; however, when the laboratory setting allows researchers to control the type of decision, the context of the decision, the quality of the information, and the type of information made available to the decision-maker, behaviours become more stable and predictable. Again, this type of research is better suited for the laboratory than the field because purposely manipulating employees is a risky proposition for public managers.

The impact of information technology

This section reviews experimental works which examine how variation in information and decision-making environments affect decision-makers. These studies are particularly relevant to applying information technology in the public sector.¹² The impact of information technology on public management is an important issue because information technology is likely to fundamentally transform the relationship between citizens and governments (Dunleavy et al. 2005, 2008). Manipulating these interactions in controlled environments where subjects can be debriefed before being dismissed is preferable to testing programmes in natural settings where false and misleading information can spread beyond the researcher's control. Public management research can help us understand how information technologies affect decision-making so that processes are designed to reinforce, rather than undermine, democratic institutions.

Landsbergen and colleagues (1997) critically examined information technologies that help in the decision-making process. They conducted controlled experiments to study the relationship between the quality, confidence, and commitment of decisions aided by expert systems (computer modelling technologies designed to simulate the human decision process). Test subjects, graduate students at several universities, were asked to select the top three job applicants from a pool of ten. Participants were randomly assigned to treatment groups including a group where information was presented on paper, a group where information was presented on computer, and a group where

information was provided on a computer-based expert system with expansion capabilities. After making their decisions, subjects were provided contradictory solutions and their commitment to their decisions was assessed. The authors found that decision-makers using expert systems were able to make higher-quality decisions but displayed lower confidence and commitment in their decisions.

More recently, Grimmelikhuijsen has published a series of studies on the effect of governmental transparency on public trust (2009, 2010, 2012). These studies demonstrate how laboratory experiments can make a unique contribution to public management research. Grimmelikhuijsen evaluated the claim that transparency about governmental decisions causes higher trust in government by manipulating information about air quality on mock versions of a local government website. Some subjects were purposely supplied outdated, fragmentary information that exaggerated air pollution. Surprisingly, varying transparency had little effect on subjects' trust in government. Imagine manipulating the content available to the public on a fully operational government website to see how people react! If the political fallout from intermittent problems accessing the health insurance marketplace on HealthCare.gov is any indication, manipulating governmental transparency outside of the laboratory is unthinkable.

Recent experimental research suggests that effect of procedural fairness on public attitudes is moderated by knowledge and certainty (Herian et al. 2012) and varies by policy issue area (de Fine Licht 2014). Rather than manipulate mock government websites like Grimmelikhuijsen, Herian et al. (2012) and de Fine Licht (2014) present survey respondents with media reports of governmental budgeting decisions that describe varying degrees of public participation and involvement in the decision-making process. This approach allows the researchers to investigate the effect of transparency without manipulating actual budgeting processes. Moreover, the experimental protocol provides researchers the opportunity to correct any potentially misleading reports before concluding data collection.¹³

Laboratory studies by Landsbergen et al., Grimmelikhuijsen, Herian et al., and de Fine Licht (2014) potentially offer insight into efforts to actively engage citizens in public affairs through deliberative democracy (which Elster 1998, 1 defines as 'decision making by discussion among free and equal citizens'). Scholars have articulated bold ideas for democratic engagement but few of these ideas have been rigorously tested. Reviewing the empirical findings on deliberative democracy, Delli Carpini, Cook, and Jacobs (2004, 328) found a 'good deal of support' for the democratic potential of public deliberation but cautioned that the positive contributions of public participation are 'highly context dependent and rife with opportunities for going awry.' Ryfe (2005, 64) offers a similar assessment on the state of research on deliberative democracy: '[A] great deal more research remains to be done. Despite its breadth, the empirical study of deliberation is not yet very rich or deep'.

Laboratory experiments in public management could play a useful role testing the impact of deliberative democracy proposals on decision-making. As discussed earlier, a

number of laboratory experiments in public management have examined individual decision-making and interactions between the government and public. This line of research is particularly appropriate because we should be reluctant to manipulate democratic rules in the real world to learn what works through trial and error. In a laboratory setting, the researcher may responsibly manipulate the terms of democratic engagement to generate useful information about effective processes.

ISSUES IN EXPERIMENTAL RESEARCH IN PUBLIC MANAGEMENT

The remainder of this article considers some pressing challenges and opportunities associated with laboratory experiments in the field of public management. In particular, we discuss issues of realism and external validity and strategies for estimating and managing the costs of laboratory research.

External validity and realism

Because public management research aims to generate useful knowledge, it is important to address issues of external validity. Public management scholars have displayed a tendency to trade internally valid knowledge produced in the laboratory for externally valid knowledge available in the field (Bozeman and Scott 1992). Unless issues of realism and external validity are resolved satisfactorily, researchers will be understandably reluctant to conduct laboratory research.

At the outset, it should be noted that an experiment may still yield important contributions even if its findings fare poorly in dimensions of generalizability. Lucas (2003, 240–241) advises that the researcher's goal is not always to reach generally applicable conclusions: 'When an experiment is designed to test theoretical principles, to ask whether the experiment's sample allows for generalization to a larger population is to ask the wrong question'. In some instances, the ability to make definitive causal statements may outweigh the ability to generalize settings to other findings. The researcher should clearly state his or her objectives and accept that there is some inherent trade-off between internal and external validity.¹⁴

The issue of external validity has been debated extensively in the social science (Lucas 2003; Garcia and Wantchekon 2010 and others). It is very common for academic researchers to use students as subjects in laboratory experiments. Across social science disciplines, the majority of experimental studies use students (Davis and Holt 1993; Roth and Kagel 1995).

The practice of using undergraduate or graduate students in experimental research in organizational studies has engendered substantial criticism (Dobbins, Lane, and Steiner 1988; Gordon, Allen Slade, and Schmitt 1986; Levitt and List 2007). For example, one

critic argued that ‘generations of colleges students have toiled in university laboratories solving problems they did not create, learning syllables they have never seen before, and selecting applicants for hire in nonexistent organizations’ (Gordon, Allen Slade, and Schmitt 1986, 191).

However, ample empirical evidence suggests that students and managers respond similarly to management-related experimental treatments (Remus 1986, 1989). According to Druckman and Leeper (2012, 878), ‘an increasing amount of evidence suggests results from [undergraduate] samples widely generalize’. To the extent that the causal mechanisms are based on general psychological factors that students and managers have in common, the case for generalizing results is very strong. The experimental responses of students are seldom different from those of other subject pools (Ball and Cech 1996; Plott 1987). Recent comparisons of student and non-student responses to identical experiments indicate that student responses in experiments are largely equivalent to those of non-students (Alm 2012). This suggests that the benefits of this approach (cost and recruitment efficiency) may outweigh the costs (external validity). Moreover, since many laboratory experiments in management research use task performance as a dependent variable (e.g. Bonner and Sprinkle 2002; Locke et al. 1981; Wageman and Baker 1997), there may be some benefit in drawing samples from student populations where variations in ability (a predictor of performance) are often constrained by college admissions requirements.

Nevertheless, the researcher should consider whether students differ from the general public or public employees in ways that are likely to bias experimental results. The researcher may also consider whether his or her intended audience will discredit, mistakenly perhaps, the results of experiments involving student subjects. If so, the researcher may supplement a student sample with non-student participants. The research by James (2011b) on the effect of performance measures on public evaluation of governmental performance demonstrates this approach by utilizing a field experiment with non-student samples to address external validity as well as a pair of laboratory experiments where manipulation is used to address causal relationships. As we discuss in the following, modern informational technology has lowered the barriers to non-student participation in laboratory experiments by academic researchers, making off-campus recruitment and participation more feasible.

Laboratory experiments can be evaluated along dimensions of realism, of which there are several subcomponents, including mundane and experimental realism (Dobbins, Lane, and Steiner 1988; Carlsmith, Ellsworth, and Aronson 1976; Singleton, Straits, and Straits 2009). Researchers can increase the externality validity of their work by addressing issues of realism in their experimental designs.

Mundane realism pertains to the extent to which experimental conditions approximate actual behavioural settings. Mundane realism can be enhanced by having subjects execute experimental tasks that are relevant to and representative of tasks conducted by public sector workers. For example, toy assembly, a common task used in psychology

research (Jenkins et al. 1998), is arguably less relevant to public administration than developing a budget recommendation as seen in Nutt (2005).

Experimental realism pertains to the extent to which subjects perceive the experiment to be realistic, as opposed to a trivial exercise. Subjects in experiments with high levels of experimental realism will respond to treatments naturally and honestly. Merely asking subjects to pretend they are public administrators or budget analysts confronted with a hypothetical situation may lead to filtered behavioural responses.

In an effort to enhance experimental realism, the laboratory analogue is proposed as an alternative to the hypothetical scenario. Schwartz-Shea (1991) adopted this approach in her study of group decision making. The author designed a multi-stage game consisting of several in-groups and out-groups. Variables in this small group experiment included levels of discussion permitted and the format in which discussion was allowed. Subjects were required to make real decisions and were provided real monetary compensation based on game outcomes. Subjects were able to maximize compensation through cooperation in environments where opportunities for discussion varied. This represents a significant departure from the body of experiments in public administration that typically rely on hypothetical and scenario-based protocols.

Brewer and Brewer (2011) also avoid the limitations of hypothetical scenarios in their test of the effects of sector designation on work motivation. In this study, university students were asked to perform a simple task on a computer that tested their ability to sustain attention. Subjects were randomly assigned to groups and told that their work was funded by either a government agency or a business firm. Brewer and Brewer find that 'when individuals believe their work is sponsored and funded by a government agency, they perform significantly faster, more accurately and more vigilantly' (358). This research design has low mundane realism but high experimental realism. The outcome of interest is not a hypothetical administrative decision, but rather subjects' ability to perform a task in the laboratory.

In some instances, incentives can motivate student-subjects to take experiments more seriously and increase experimental realism, but researchers should be aware of problems that incentives can introduce (for an overview see, Camerer and Hogarth 1999). Financial incentives do not incorporate all the factors a real manager would use and increase the cost of conducting experimental research.¹⁵

Cost and sample size

The cost of conducting laboratory experiments poses an obstacle for public management researchers. Laboratory studies are generally more expensive than studies of observational data. The majority of costs is generally associated with participation incentives for subjects. If the researcher hopes to recruit actively employed public sector managers as experimental subjects, the cost of collecting data is likely to increase. Therefore, resource management is an important issue for experimentalists.¹⁶

Because the cost of experimental research is directly related to sample size, researchers should estimate how many participants are needed in order to budget their resources effectively. The sample size required to generate a study with sufficient statistical power can be determined by following approaches such as those provided by Cohen (1962, 1988). Statistical power refers to the probability that confidence intervals of coefficients estimated with a given sample size will allow one to reject the null hypothesis. In the approach outlined by Cohen, sample size should be determined as a function of three factors: the Type I error rate (fixed by convention), the desired level of power (fixed by convention), and the estimated treatment effect size (often inestimable). Because the researcher must estimate an unknown quantity, it has been argued that sample size calculations are ‘mystical’ (Schulz and Grimes 2005). We attempt to demystify this valuable planning tool with an example from our own experience.¹⁷

One of the authors of this article recently conducted an experiment on individual decision-making, hypothesizing that a treatment would cause subjects to agree with a randomized (coin-flip) prior decision more often than a control group. Let π_t equal the proportion of agreement in the treatment group, π_c equal the proportion of agreement in the control group, and N_t and N_c equal the size of the treatment and control groups, respectively. We know that the standard error of the difference of proportions between two samples (the treatment effect) is equal to:

$$SE_{\pi_t - \pi_c} = \sqrt{\frac{\pi_t(1 - \pi_t)}{N_t} + \frac{\pi_c(1 - \pi_c)}{N_c}}$$

Based on theory and prior works, the author thought a treatment might increase the probability that subjects agree with the prior decision by 4 per cent. Because the prior decision is randomized between the two possible choices, π_c is set to 0.50 and π_t is set to 0.54. We want to estimate sample sizes N_t and N_c , but expect both groups to be the same size. Let $N = N_t + N_c$.

In repeated samples, the value of our estimated treatment effect will be normally distributed with mean equal to 0.04 and standard error = $\sqrt{\frac{0.4984}{N}}$. Figure 2 illustrates the relationship between sample size, standard error of our estimator, and the power of our research design.

If we sample 1,197 subjects, 95 per cent of the expected distribution of $\pi_t - \pi_c$ is greater than zero (upper left panel). However, the statistical power with this sample size is only 50 per cent. The 95 per cent confidence intervals of our estimated treatment effect will include zero in 50 per cent of repeated samples (see upper right panel). We need a larger sample for more statistical power. For 80 per cent power (a conventional figure), we need a sample size N , such that $\pi_t - \pi_c = 2.8 \sqrt{\frac{0.4984}{N}}$.¹⁸ Given a Type I error rate of 5 per cent, a desired power level of 80 per cent, and an estimated 4 per cent treatment effect, we would need to recruit 2,442 subjects to our

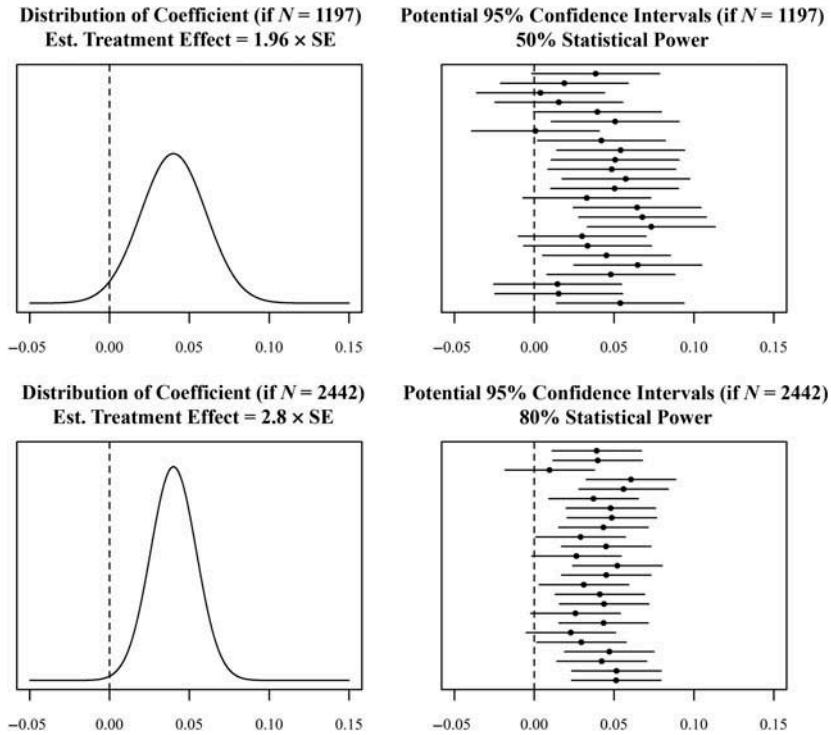


Figure 2: Relationship between sample size, standard error, and statistical power

experiment (bottom panels of Figure 2). These calculations indicated our experiment would be too costly. We needed to increase the strength of our treatment (or work on other projects). Because the necessary sample size decreases in proportion to the treatment effect squared, administering a stronger treatment helps the researcher manage sample sizes and research costs. In this case, we redesigned our instrument so that an expected treatment effect of 10 per cent seemed reasonable; this reduced the sample size needed for 80 per cent statistical power with a Type I error rate of 5 per cent to 392 test subjects.¹⁹

The view taken here with regard to sample size and resource management is twofold. First, public administration's aspiring experimentalists should be aware that the challenge of limited resources is not new to social science experimentation. Second, it is true that statistical power calculations require estimating unknown quantities and the approach for determining sample size is sometimes omitted from published works. Nevertheless, aspiring experimentalists should be very deliberate in determining how many observations they need to reach reliable conclusions. The researcher may need to strike a balance between realism and cost-effectiveness. Planning ahead can help prevent some problems from arising during experiment research.

Financial barriers to experimental research are, fortunately, declining as a result of technological advances. Many routine economic, managerial and other human interactions have moved to the Internet, and social scientists are also conducting research using virtual, online laboratories (Shen et al. 1999; Ma and Nickerson 2006; Birnbaum 2000). Recent literature has seen increasingly creative implementation of classic laboratory studies on the Internet, including studies of decision-making and collaboration (Edelman 2012). Indeed, the availability of research subjects online may account for much of the recent increase in experimental work evident in Figure 1. Nearly one-third of the experimental studies published in leading journals since 2010 were conducted online.²⁰

MOVING FORWARD WITH LABORATORY EXPERIMENTS

We believe that the potential for laboratory experiments to contribute to public management research is unfulfilled. However, a field's methodological advancement in the area of laboratory experimentation can occur across a relatively short period of time. In the early 1990s, laboratory experiment papers comprised a mere 3 per cent of articles published in leading economic journals (Falk and Heckman 2009). Today, laboratory studies in economics are prominent. A single institution can generate a significant positive impact on the field's adoption of laboratory studies.²¹ There are some signs that experimental research is finally penetrating the mainstream of public administration and management research.

Public management should not seek to become an experimental social science, but experimentation can complement existing and emerging methods. Few institutions will devote the resources necessary to constructing new facilities for laboratory research, but we can incrementally integrate laboratory research into the set of methods we use to conduct research. Consider how existing resources can be utilized for laboratory research.²² One of the authors of this article reserved a campus computer lab for several days to conduct a laboratory experiment. Creativity and persistence can often compensate for lack of resources.

Collaboration represents another opportunity for incremental improvement. Appropriate facilities for experimental research may be available in other departments of one's college or university. Bozeman and Scott (1992) note that a lack of experience and knowledge of experiment-based research designs contributes to their under-representation in the literature. This condition can be improved though collaboration with scholars in fields more accustomed to experiment-based research.²³

We encourage other scholars to experiment with experiments. We wholeheartedly agree with Margetts (2011, 204) that 'experiments, in spite of all the logistical and ethical challenges discussed, are exciting and fun!' In this article, we have argued that laboratory experiments can generate useful information on particularly sensitive issues

facing public managers, such as the decision-making of public employees and the relationship of public agencies to the public. A few laboratory studies have addressed these issues successfully and demonstrated the promise of this method research, but that promise is largely unfulfilled. We encourage other scholars to consider using laboratory experiments to conduct public management research.

ACKNOWLEDGEMENTS

The authors would like to thank Barry Bozeman, Hal Rainey, Andrew Whitford, Robert Christensen, Justin Stritch, and Mogens Jin Pedersen for their feedback on prior versions of this article. They would also like to thank Azfar Nisar and Megan Lynn for their research assistance.

NOTES

- 1 The term 'laboratory' typically conjures up the image of an antiseptic environment stocked with exotic scientific instruments. Social science laboratories are comparatively modest, often resembling a small computer cluster. As we discuss in the text later, the financial cost of conducting laboratory experiments in public management, while not trivial, is not prohibitive.
- 2 Statistical control variables are useful, but as the saying goes 'you don't know what you don't know'. The problem is not knowing that one has omitted a variable. There is not a foolproof way to test for omitted variables. Random assignment to treatment conditions in laboratory experiment is a solution to lurking omitted variable problems.
- 3 This is not to suggest that causation is the only relevant issue, but rather that experiments are a useful method to employ when causation is the focal point of analysis. As Dobbins and colleagues note, 'if a researcher is interested in estimating the general satisfaction level of clerical workers, laboratory research is clearly inappropriate. If, on the other hand, the question is one of understanding factors which produce satisfaction in clerical works, then laboratory methodologies could be used' (1988, 282).
- 4 Consistent with Margetts (2011), the journals searched include *Public Administration Review*, *Public Management Review*, *Journal of Public Administration Research and Theory*, and *Public Administration*. We include field and survey experiments, in addition to laboratory experiments, but exclude observational studies of 'natural experiments'. Appendix 1 provides a comprehensive list of these studies from 2010 through mid-2014.
- 5 Some experimental designs potentially cross lines between laboratory, field, and survey research. For example, the researcher may recruit subjects to a laboratory to take a survey or administer the instrument in the field. Rather than focus on where an experiment is administered, it makes more sense to look at whether the research controls the data-generating process. Survey experiments often use real public sector managers to great advantage. Of course, the treatments one can administer via surveys are limited and the researcher cannot study dynamic processes.
- 6 Empirical research suggests the risk of contaminating the pool of research subjects through deception is limited (Barrera and Simpson 2012).
- 7 In our experience, a consent agreement may be drafted to describe the experiment in broad terms, so the subject can assess its risks and rewards omitting some detail about the treatment protocol. Following the experiment, the researcher should debrief subjects, disclosing the deception and providing subjects an opportunity to withdraw from the experiment.

- 8 It is conceivable that employees may perform better if they know their work is being studied as part of an experiment. This type of 'Hawthorne Effect' is a known source of experimental bias and may require the experimenter to withhold information about his or her experiment from test subjects. As we discuss in the text, the experimenter must be cognizant of ethical issues that arise in human subject research.
- 9 According to Cooper (1987, 326), the public manager's ethical obligations also extend to his or her colleagues within the organization.
- 10 Other public management laboratory studies that examine individual decision-making include Scott (1997), Schwartz-Shea (1991), Miller and Whitford (2002), Bozeman and McAlpine (1977), Bretschneider and Straussman (1992), Bretschneider, Straussman, and Mullins (1988), Kaufmann and Feeney (2014), and Christensen et al. (2013).
- 11 The theory of credibility logic (Bozeman 1986; Bozeman and Landsbergen 1989) asserts that decision-makers will subjectively assess the credibility of relevant information on a variety of dimensions including the source of the information. The central feature of credibility logic is the observation that perceptions of believability take precedence over scientific quality when determining the value of information or knowledge. At the same time, cognitive response theory asserts that a decision-maker's response to information is dependent upon a personal scheme of relevant experience.
- 12 Other public management studies that explore these issues using laboratory experiments include Shangraw (1986) and Bozeman and Shangraw (1988, 1989).
- 13 It should be noted that Herian et al. (2012) and de Fine Licht (2014) embed their manipulations in online surveys rather than recruit subjects to a laboratory. While the researcher loses some degree of control conducting an experiment online (the subject may walk away from his or her computer before completing the survey), these are not field experiments; the researcher must obtain consent and has the opportunity to debrief participants before concluding the survey.
- 14 While there may be some trade-off between internal and external validity, research is often cumulative. Demonstrating that a public management practice produces positive outcomes in a laboratory may justify conducting subsequent field experiments. Public managers could then participate in experimental research with reasonable expectations of doing good, rather than manipulating human subjects for exploratory research.
- 15 In our own experience, undergraduate subjects take their roles in laboratory experiments seriously, even in voluntary service. Students are accustomed to taking tests seriously without immediate financial reward for their efforts.
- 16 Financial constraints have often lead research to rely on small sample sizes and methodological innovations to correct for relatively low statistical power such as treatment-only experimental designs (Collins, Dziak, and Li 2009). Methodologists have, on the one hand, supported efforts to preserve resources (Collins, Dziak, and Li 2009; Edwards et al. 1997) and, on the other hand, argued that findings of low-powered studies are so problematic as to render their adoption unethical (Bacchetti et al. 2005; Halpern, Karlawish, and Berlin 2002; Janosky 2002).
- 17 For further guidance on calculating statistical power, including power calculation for variables of interest other than a simple difference of proportions, see Kraemer (1991), Legg and Nagy (2006), Nakazawa (2011) and Gelman (2007, Chapter 20).
- 18 With a Type I error of 0.05, we know a coefficient must be 1.96 or more standard errors away from zero to be statistically significant. To achieve statistical power P , the mean value of expected distribution of coefficient values must be $1.96 + \text{qnorm}(P)$ standard errors from zero. For 80 per cent statistical power, the mean of our expected treatment effect must be $1.96 + \text{qnorm}(0.8) = 2.8$ standard errors from zero. If we desired 95 per cent statistical power, we would need a sample size large enough that our estimated treatment effect is 3.6 standard errors from zero. Given an estimated 0.04 treatment effect, this high level of statistical power would require a sample size of 4,037. It should be clear that it is impossible to achieve 100 per cent statistical power.
- 19 The stronger treatment potentially gives rise to the issues of realism and external validity discussed earlier, but we do our best to produce the most useful results possible given our financial constraints.

18 Public Management Review

- 20 These studies include Van De Walle and Van Ryzin (2011), Herian et al. (2012), Feeney (2012), Charbonneau and Van Ryzin (2013), Nielsen and Baekgaard (2013), de Fine Licht (2014), and Ricucci, Van Ryzin, and Lavena (2014).
- 21 Many of the experimental works discussed in this article are products of Syracuse University's Maxwell School of Citizenship and Public Affairs. Bozeman, Bretschneider, and Straussman were faculty and Coursey, Whitmer, Thurmaier, Landsbergen, and Shangraw were doctoral students there during the late 1980s to the early 1990s.
- 22 Weibel, Rost, and Osterloh (2010) leverage the benefits of laboratory studies without as performing one. Their meta-analysis draws on studies from business management, psychology, and economics, where relevant experiment-based studies are readily available.
- 23 A nice example of this is the work of Brewer and Brewer (2011). In this case, one of the co-authors is a public administration scholar and the other is an experimental psychologist.

REFERENCES

- Alm, J. 2012. "Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies." *International Tax and Public Finance* 19 (1): 54–77. doi:10.1007/s10797-011-9171-2.
- Arceneaux, K., and D. W. Nickerson. 2009. "Who Is Mobilized to Vote? A Re-Analysis of 11 Field Experiments." *American Journal of Political Science* 53 (1): 1–16. doi:10.1111/j.1540-5907.2008.00354.x.
- Avellaneda, C. N. 2013. "Mayoral Decision-Making: Issue Salience, Decision Context, and Choice Constraint? An Experimental Study with 120 Latin American Mayors." *Journal of Public Administration Research and Theory* 23 (3): 631–661. doi:10.1093/jopart/mus041.
- Bacchetti, P., L. E. Wolf, M. R. Segal, and C. E. McCulloch. 2005. "Ethics and Sample Size." *American Journal of Epidemiology* 161 (2): 105–110. doi:10.1093/aje/kwi014.
- Ball, S. B., and P.-A. Cech. 1996. "Subject Pool Choice and Treatment Effects in Economic Laboratory Research." *Research in Experimental Economics* 6 (3): 239–292.
- Barrera, D., and B. Simpson. 2012. "Much Ado About Deception: Consequences of Deceiving Research Participants in the Social Sciences." *Sociological Methods & Research* 41 (3): 383–413. doi:10.1177/0049124112452526.
- Belle, N. 2014. "Leading to Make a Difference: A Field Experiment on the Performance Effects of Transformational Leadership, Perceived Social Impact, and Public Service Motivation." *Journal of Public Administration Research and Theory* 24 (1): 109–136. doi:10.1093/jopart/mut033.
- Bellé, N. 2013. "Experimental Evidence on the Relationship between Public Service Motivation and Job Performance." *Public Administration Review* 73 (1): 143–153. doi:10.1111/j.1540-6210.2012.02621.x.
- Bertrand, M., and S. Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013. doi:10.1257/0002828042002561.
- Birnbaum, M. H. 2000. *Psychological Experiments on the Internet*. San Diego, CA: Academic Press.
- Bonner, S. E., and G. B. Sprinkle. 2002. "The Effects of Monetary Incentives on Effort and Task Performance: Theories, Evidence, and a Framework for Research." *Accounting, Organizations and Society* 27 (4–5): 303–345. doi:10.1016/S0361-3682(01)00052-6.
- Bozeman, B. 1986. "The Credibility of Policy Analysis: Between Method and Use." *Policy Studies Journal* 14 (4): 519–539. doi:10.1111/j.1541-0072.1986.tb00358.x.
- Bozeman, B., and D. Landsbergen. 1989. "Truth and Credibility in Sincere Policy Analysis: Alternative Approaches for the Production of Policy-Relevant Knowledge." *Evaluation Review* 13 (4): 355–379. doi:10.1177/0193841X8901300402.
- Bozeman, B., and W. E. McAlpine. 1977. "Goals and Bureaucratic Decision-Making: An Experiment." *Human Relations* 30 (5): 417–429. doi:10.1177/00187267703000502.

- Bozeman, B., and P. Scott. 1992. "Laboratory Experiments in Public Policy and Management." *Journal of Public Administration Research and Theory: J-PART 2* (3): 293–313.
- Bozeman, B., and R. Shangraw. 1989. "Computers and Commitment to a Public Management Decision: An Experiment." *Knowledge in Society* 2 (3): 42–56. doi:10.1007/BF02687206.
- Bozeman, B., and R. F. Shangraw. 1988. "Computers as a Public Management Decision Tool: An Experimental Analysis of the Effects of Cost and Access on Information Choice." *Science Communication* 10 (2): 111–139. doi:10.1177/0164025988010002002.
- Bretschneider, S., and J. Straussman. 1992. "Statistical Laws of Confidence versus Behavioral Response: How Individuals Respond to Public Management Decisions under Uncertainty." *Journal of Public Administration Research and Theory: J-PART 2* (3): 333–345.
- Bretschneider, S., J. J. Straussman, and D. Mullins. 1988. "Do Revenue Forecasts Influence Budget Setting? A Small Group Experiment." *Policy Sciences* 21 (4): 305–325. doi:10.1007/BF00138306.
- Brewer, G. A., and G. A. Brewer Jr. 2011. "Parsing Public/Private Differences in Work Motivation and Performance: An Experimental Study." *Journal of Public Administration Research and Theory* 21 (Supplement 3): i347–362. doi:10.1093/jopart/mur030.
- Camerer, C. F., and R. M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19 (1/3): 7–42. doi:10.1023/A:1007850605129.
- Caplan, C. F. 1979. "The Two-Communities Theory and Knowledge Utilization." *American Behavioral Scientist* 22 (3): 459–470. doi:10.1177/000276427902200308.
- Carlsmith, J. M., P. C. Ellsworth, and E. Aronson. 1976. *Methods of Research in Social Psychology*. Reading, MA: Addison-Wesley.
- Charbonneau, É., and G. G. Van Ryzin. 2013. "Benchmarks and Citizen Judgments of Local Government Performance: Findings from a Survey Experiment." *Public Management Review* no. ahead-of-print 1–17. doi:10.1080/14719037.2013.798027.
- Christensen, R. K., S. W. Whiting, T. Im, E. Rho, J. M. Stritch, and J. Park. 2013. "Public Service Motivation, Task, and Non-Task Behavior: A Performance Appraisal Experiment with Korean MPA and MBA Students." *International Public Management Journal* 16 (1): 28–52. doi:10.1080/10967494.2013.796257.
- Clerkin, R. M., S. R. Paynter, and J. K. Taylor. 2009. "Public Service Motivation in Undergraduate Giving and: Volunteering Decisions." *The American Review of Public Administration* 39 (6): 675–698. doi:10.1177/0275074008327512.
- Cohen, J. 1962. "The Statistical Power of Abnormal-Social Psychological Research: A Review." *The Journal of Abnormal and Social Psychology* 65 (3): 145–153. doi:10.1037/h0045186.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Collins, L. M., J. J. Dziak, and R. Li. 2009. "Design of Experiments with Multiple Independent Variables: A Resource Management Perspective on Complete and Reduced Factorial Designs." *Psychological Methods; Psychological Methods* 14 (3): 202–224. doi:10.1037/a0015826.
- Colquitt, J. A. 2008. "From the Editors Publishing Laboratory Research in AMJ: A Question of When, Not If." *Academy of Management Journal* 51 (4): 616–620.
- Cook, K. S., and T. Yamagishi. 2008. "A Defense of Deception on Scientific Grounds." *Social Psychology Quarterly* 71 (3): 215–221. doi:10.1177/019027250807100303.
- Cook, T. D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Chicago, IL: Houghton Mifflin.
- Cooper, T. L. 1987. "Hierarchy, Virtue, and the Practice of Public Administration: A Perspective for Normative Ethics." *Public Administration Review* 47: 320–328. doi:10.2307/975312.
- Coursey, D. H. 1992. "Information Credibility and Choosing Policy Alternatives: An Experimental Test of Cognitive-Response Theory." *Journal of Public Administration Research and Theory: J-PART 2* (3): 315–331.
- Davis, D. D., and C. A. Holt. 1993. *Experimental Economics*. Princeton, NJ: Princeton University Press.

- De Fine Licht, J. 2014. "Policy Area as a Potential Moderator of Transparency Effects: An Experiment." *Public Administration Review* 74 (3): 361–371. doi:10.1111/puar.12194.
- Delli Carpini, X. M., F. L. Cook, and L. R. Jacobs. 2004. "Public Deliberation, Discursive Participation, and Citizen Engagement: A Review of the Empirical Literature." *Annual Review of Political Science*. 7: 315–344. doi:10.1146/annurev.polisci.7.121003.091630.
- Dobbins, G. H., I. M. Lane, and D. D. Steiner. 1988. "A Note on the Role of Laboratory Methodologies in Applied Behavioural Research: Don't Throw out the Baby with the Bath Water." *Journal of Organizational Behavior* 9 (3): 281–286. doi:10.1002/job.4030090308.
- Druckman, J. N., D. P. Green, J. H. Kuklinski, and A. Lupia. 2006. "The Growth and Development of Experimental Research in Political Science." *American Political Science Review* 100 (4): 627–635. doi:10.1017/s0003055406062514.
- Druckman, J. N., and T. J. Leeper. 2012. "Learning More from Political Communication Experiments: Pretreatment and Its Effects." *American Journal of Political Science* 56 (4): 875–896. doi:10.1111/j.1540-5907.2012.00582.x.
- Dunleavy, P., H. Margetts, S. Bastow, and J. Tinkler. 2005. "New Public Management Is Dead—long Live Digital-Era Governance." *Journal of Public Administration Research and Theory* 16 (3): 467–494. doi:10.1093/jopart/mui057.
- Dunleavy, P., H. Margetts, S. Bastow, and J. Tinkler. 2008. *Digital Era Governance: IT Corporations, the State, and E-Government*. Oxford: Oxford University Press.
- Edelman, B. 2012. "Using Internet Data for Economic Research." *The Journal of Economic Perspectives* 26 (2): 189–206. doi:10.1257/jep.26.2.189.
- Edwards, S. J. L., R. J. Lilford, D. Braunholtz, and J. Jackson. 1997. "Why 'underpowered' Trials Are Not Necessarily Unethical." *The Lancet* 350 (9080): 804–807. doi:10.1016/S0140-6736(97)02290-3.
- Elster, J. 1998. *Deliberative Democracy*. Cambridge: Cambridge University Press.
- Falk, A., and J. J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952): 535–538. doi:10.1126/science.1168244.
- Feeney, M. K. 2012. "Organizational Red Tape: A Measurement Experiment." *Journal of Public Administration Research and Theory* 22 (3): 427–444. doi:10.1093/jopart/mus002.
- Gano, G. L., J. E. Crowley, and D. Guston. 2006. "'Shielding' the Knowledge Transfer Process in Human Service Research." *Journal of Public Administration Research and Theory* 17 (1): 39–60. doi:10.1093/jopart/muj013.
- Garcia, F. M., and L. Wantchekon. 2010. "Theory, External Validity, and Experimental Inference: Some Conjectures." *The Annals of the American Academy of Political and Social Science* 628 (1): 132–147. doi:10.1177/0002716209351519.
- Gelman, A. 2007. *Data Analysis Using Regression and Multilevel/hierarchical Models*. New York: Cambridge University Press.
- Gerber, A. S., D. P. Green, and C. W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102 (1): 33. doi:10.1017/S000305540808009X.
- Gill, J., and K. J. Meier. 2000. "Public Administration Research and Practice: A Methodological Manifesto." *Journal of Public Administration Research and Theory* 10 (1): 157–199. doi:10.1093/oxfordjournals.jp.art.a024262.
- Goggin, M. L. 1986. "The 'Too Few Cases/Too Many Variables' Problem in Implementation Research." *The Western Political Quarterly* 39 (2): 328–347. doi:10.2307/448302.
- Gordon, M. E., L. Allen Slade, and N. Schmitt. 1986. "The 'Science of the Sophomore' Revisited: From Conjecture to Empiricism." *The Academy of Management Review* 11 (1): 191–207.
- Grimmelikhuijsen, S. 2009. "Do Transparent Government Agencies Strengthen Trust?" *Information Polity* 14 (3): 173–186.
- Grimmelikhuijsen, S. 2010. "Transparency of Public Decision-Making: Towards Trust in Local Government?" *Policy & Internet* 2 (1): 4–34. doi:10.2202/1944-2866.1024.

- Grimmelikhuijsen, S. 2012. "Linking Transparency, Knowledge and Citizen Trust in Government: An Experiment." *International Review of Administrative Sciences* 78 (1): 50–73. doi:10.1177/0020852311429667.
- Grimmelikhuijsen, S., G. Porumbescu, B. Hong, and T. Im. 2013. "The Effect of Transparency on Trust in Government: A Cross-National Comparative Experiment." *Public Administration Review* 73 (4): 575–586. doi:10.1111/puar.12047.
- Grimmelikhuijsen, S. G., and A. J. Meijer. 2014. "Effects of Transparency on the Perceived Trustworthiness of a Government Organization: Evidence from an Online Experiment." *Journal of Public Administration Research and Theory* 24 (1): 137–157. doi:10.1093/jopart/mus048.
- Halpern, S. D., J. H. T. Karlawish, and J. A. Berlin. 2002. "The Continuing Unethical Conduct of Underpowered Clinical Trials." *JAMA: The Journal of the American Medical Association* 288 (3): 358–362. doi:10.1001/jama.288.3.358.
- Herian, M. N., J. A. Hamm, A. J. Tomkins, and L. M. Pytlik Zillig. 2012. "Public Participation, Procedural Fairness, and Evaluations of Local Governance: The Moderating Role of Uncertainty." *Journal of Public Administration Research and Theory* 22 (4): 815–840. doi:10.1093/jopart/mur064.
- Hock, S., S. Anderson, and M. Potoski. 2013. "Invitation Phone Calls Increase Attendance at Civic Meetings: Evidence from a Field Experiment." *Public Administration Review* 73 (2): 221–228. doi:10.1111/j.1540-6210.2012.02627.x.
- Hvidman, U., and S. C. Andersen. 2014. "The Impact of Performance Management in Public and Private Organizations." *Journal of Public Administration Research and Theory* 24 (1): 35–58. doi:10.1093/jopart/mut019.
- Jakobsen, M. 2013. "Can Government Initiatives Increase Citizen Coproduction? Results of a Randomized Field Experiment." *Journal of Public Administration Research and Theory* 23 (1): 27–54. doi:10.1093/jopart/mus036.
- Jakobsen, M., and S. C. Andersen. 2013. "Intensifying Social Exchange Relationships in Public Organizations: Evidence from a Randomized Field Experiment." *Journal of Policy Analysis and Management* 32 (1): 60–82. doi:10.1002/pam.21662.
- James, O., and A. Moseley. 2014. "Does Performance Information about Public Services Affect Citizens' Perceptions, Satisfaction, and Voice Behaviour? Field Experiments with Absolute and Relative Performance Information." *Public Administration* 92 (2): 493–511. doi:10.1111/padm.12066.
- James, O. 2011a. "Managing Citizens' Expectations of Public Service Performance: Evidence from Observation and Experimentation in Local Government." *Public Administration* 89 (4): 1419–1435. doi:10.1111/j.1467-9299.2011.01962.x.
- James, O. 2011b. "Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments." *Journal of Public Administration Research and Theory* 21 (3): 399–418. doi:10.1093/jopart/muq057.
- Janosky, J. E. 2002. "The Ethics of Underpowered Clinical Trials." *JAMA: The Journal of the American Medical Association* 288 (17): 2118–2119.
- Jenkins, G. D. Jr, A. Mitra, N. Gupta, and J. D. Shaw. 1998. "Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research." *Journal of Applied Psychology* 83 (5): 777–787. doi:10.1037/0021-9010.83.5.777.
- Kaufmann, W., and M. K. Feeney. 2014. "Beyond the Rules: The Effect of Outcome Favourability on Red Tape Perceptions." *Public Administration* 92 (1): 178–191. doi:10.1111/padm.12049.
- Kraemer, H. C. 1991. "To Increase Power in Randomized Clinical Trials without Increasing Sample Size." *Psychopharmacology Bulletin* 27 (3): 217–224.
- Landsbergen, D., B. Bozeman, and S. Bretschneider. 1992. "'Internal Rationality' and the Effects of Perceived Decision Difficulty: Results of a Public Management Decisionmaking Experiment." *Journal of Public Administration Research and Theory: J-PART* 2 (3): 247–264.

- Landsbergen, D., D. H. Coursey, S. Loveless, and R. F. Shangraw. 1997. "Decision Quality, Confidence, and Commitment with Expert Systems: An Experimental Study." *Journal of Public Administration Research and Theory* 7 (1): 131–158. doi:10.1093/oxfordjournals.jpart.a024336.
- Legg, C. J., and L. Nagy. 2006. "Why Most Conservation Monitoring Is, but Need Not Be, a Waste of Time." *Journal of Environmental Management* 78 (2): 194–199. doi:10.1016/j.jenvman.2005.04.016.
- Levitt, S. D., and J. A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *The Journal of Economic Perspectives* 21 (2): 153–174. doi:10.1257/jep.21.2.153.
- Locke, E. A., and G. P. Latham. 1990. "Work Motivation and Satisfaction: Light at the End of the Tunnel." *Psychological Science* 1 (4): 240–246. doi:10.1111/j.1467-9280.1990.tb00207.x.
- Locke, E. A., K. N. Shaw, L. M. Saari, and G. P. Latham. 1981. "Goal Setting and Task Performance: 1969–1980." *Psychological Bulletin* 90 (1): 125–152. doi:10.1037/0033-2909.90.1.125.
- Lucas, J. W. 2003. "Theory-Testing, Generalization, and the Problem of External Validity." *Sociological Theory* 21 (3): 236–253. doi:10.1111/1467-9558.00187.
- Lynn, L. E. 1996. *Public Management as Art, Science, and Profession*. Chatham, NJ: Chatham House Publishers.
- Ma, J., and J. V. Nickerson. 2006. "Hands-On, Simulated, and Remote Laboratories: A Comparative Literature Review." *ACM Computing Surveys (CSUR)* 38 (3): 7. doi:10.1145/1132960.1132961.
- Margetts, H. Z. 2011. "Experiments for Public Management Research." *Public Management Review* 13 (2): 189–208. doi:10.1080/14719037.2010.532970.
- Miller, G. J., and A. B. Whitford. 2002. "Trust and Incentives in Principal-Agent Negotiations: The 'Insurance/Incentive Trade-Off'." *Journal of Theoretical Politics* 14 (2): 231–267. doi:10.1177/095169280201400204.
- Nakazawa, M. 2011. "Examining the Advantages and Disadvantages of Pilot Studies : Monte-Carlo Simulations". Diss., University of New Mexico, New Mexico.
- Nickerson, D. W. 2008. "Is Voting Contagious? Evidence from Two Field Experiments." *American Political Science Review* 102 (1): 49. doi:10.1017/S0003055408080039.
- Nielsen, P. A., and M. Baekgaard 2013. "Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an experiment." *Journal of Public Administration Research and Theory*. doi:10.1093/jopart/mut051.
- Nutt, P. C. 2005. "Comparing Public and Private Sector Decision-Making Practices." *Journal of Public Administration Research and Theory* 16 (2): 289–318. doi:10.1093/jopart/mui041.
- Pager, D. 2003. "The Mark of a Criminal Record." *American Journal of Sociology* 108 (5): 937–975. doi:10.1086/374403.
- Pager, D., B. Western, and B. Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74 (5): 777–799. doi:10.1177/000312240907400505.
- Perry, J. L. 2012. "How Can We Improve Our Science to Generate More Usable Knowledge for Public Professionals?" *Public Administration Review* 72 (4): 479–482. doi:10.1111/j.1540-6210.2012.02607.x.
- Perry, J. L., and K. L. Kraemer. 1986. "Research Methodology in the 'Public Administration Review,' 1975–1984." *Public Administration Review* 46: 215–226. doi:10.2307/3110436.
- Plott, C. R. 1987. "Dimensions of Parallelism: Some Policy Applications of Experimental Methods." In *Laboratory Experimentation in Economics: Six Points of View*, edited by A. E. Roth, 193–219. Cambridge: Cambridge University Press.
- Raadschelders, J. C. N. 2011. "The Future of the Study of Public Administration: Embedding Research Object and Methodology in Epistemology and Ontology." *Public Administration Review* 71 (6): 916–924. doi:10.1111/j.1540-6210.2011.02433.x.
- Remus, W. 1986. "Graduate Students as Surrogates for Managers in Experiments on Business Decision Making." *Journal of Business Research* 14 (1): 19–25. doi:10.1016/0148-2963(86)90053-6.
- Remus, W. 1989. "Using Students as Subjects in Experiments on Decision Support Systems." In *System Sciences: Decision Support and Knowledge Based Systems Track, Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences*. Vol. 3, 176–180. New York: IEEE. doi:10.1109/HICSS.1989.49239.

- Riccucci, N. M., G. G. Van Ryzin, and C. F. Lavena. 2014. "Representative Bureaucracy in Policing: Does It Increase Perceived Legitimacy?" *Journal of Public Administration Research and Theory* 24 (3): 537–551. doi:10.1093/jopart/muu006.
- Roth, A. E., and J. H. Kagel. 1995. *The Handbook of Experimental Economics*. Vol. 1. Princeton, NJ: Princeton University Press.
- Ryfe, D. M. 2005. "Does Deliberative Democracy Work?" *Annual Review of Political Science* 8: 49–71. doi:10.1146/annurev.polisci.8.032904.154633.
- Scandura, T. A., and E. A. Williams. 2000. "Research Methodology in Management: Current Practices, Trends, and Implications for Future Research." *The Academy of Management Journal* 43 (6): 1248–1264. doi:10.2307/1556348.
- Schulz, K. F., and D. A. Grimes. 2005. "Sample Size Calculations in Randomised Trials: Mandatory and Mystical." *The Lancet* 365 (9467): 1348–1353. doi:10.1016/S0140-6736(05)61034-3.
- Schwartz-Shea, P. 1991. "Understanding Subgroup Optimization: Experimental Evidence on Individual Choice and Group Processes." *Journal of Public Administration Research and Theory: J-PART* 1 (1): 49–74.
- Scott, P. G. 1997. "Assessing Determinants of Bureaucratic Discretion: An Experiment in Street-Level Decision Making." *Journal of Public Administration Research and Theory* 7 (1): 35–58. doi:10.1093/oxfordjournals.jpart.a024341.
- Sell, J. 2008. "Introduction to Deception Debate." *Social Psychology Quarterly* 71 (3): 213–214. doi:10.1177/019027250807100302.
- Shangraw, Jr., R. F. 1986. "How Public Managers Use Information: An Experiment Examining Choices of Computer and Printed Information." *Public Administration Review* 46 (Special Issue): 506–515. doi:10.2307/975572.
- Shen, H., Z. Xu, B. Dalager, V. Kristiansen, O. Strom, M. S. Shur, T. A. Fjeldly, J.-Q. Lu, and T. Ytterdal. 1999. "Conducting Laboratory Experiments over the Internet." *IEEE Transactions on Education* 42 (3): 180–185. doi:10.1109/13.779896.
- Simon, H. A., and W. R. Divine. 1941. "Controlling Human Factors in an Administrative Experiment." *Public Administration Review* 1 (5): 485–492. doi:10.2307/972865.
- Singleton, R. A., B. C. Straits, and M. M. Straits. 2009. *Approaches to Social Research*. New York: Oxford University Press.
- Van De Walle, S., and G. G. Van Ryzin. 2011. "The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons from a Split-Ballot Experiment." *Public Administration* 89 (4): 1436–1450. doi:10.1111/j.1467-9299.2011.01922.x.
- Van Ryzin, G. G. 2004. "Expectations, Performance, and Citizen Satisfaction with Urban Services." *Journal of Policy Analysis and Management* 23 (3): 433–448. doi:10.1002/pam.20020.
- Van Ryzin, G. G. 2006. "Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government." *Journal of Public Administration Research and Theory* 16 (4): 599–611. doi:10.1093/jopart/mui058.
- Wageman, R., and G. Baker. 1997. "Incentives and Cooperation: The Joint Effects of Task and Reward Interdependence on Group Performance." *Journal of Organizational Behavior* 18 (2): 139–158. doi:10.1002/(SICI)1099-1379(199703)18:2<139::AID-JOB791>3.0.CO;2-R.
- Weibel, A., K. Rost, and M. Osterloh. 2010. "Pay for Performance in the Public Sector—Benefits and (Hidden) Costs." *Journal of Public Administration Research and Theory* 20 (2): 387–412. doi:10.1093/jopart/mup009.
- Yackee, S. W. 2013. "Assessing Regulatory Participation by Health Professionals: A Study of State Health Rulemaking." *Public Administration Review* 73 (s1): S105–114. doi:10.1111/puar.12080.
- Yackee, S. W. 2014. "Participant Voice in the Bureaucratic Policymaking Process." *Journal of Public Administration Research and Theory*. doi:10.1093/jopart/muu007.

Appendix 1. Public administration and management experiments published in leading journals (2010 to mid-2014)*

<i>Author(s) and date</i>	<i>Journal</i>	<i>Topic</i>	<i>Experiment type</i>	<i>Sample (n)</i>
Grimmelikhuijsen et al. (2013)	PAR	Trust in government	Laboratory	Students (381 + 279)
Hock, Anderson, and Potoski (2013)	PAR	Civic engagement	Field	Residents and business owners (277)
de Fine Licht (2014)	PAR	Transparency	Laboratory	Citizens (1,032)
Bellé (2013)	PAR	Transformational leadership	Field	Nurses (90)
Jakobsen and Andersen (2013)	PAR	Coproduction and equity in public service delivery	Field	5-year-old children (284)
Yackee (2013)	PAR	Professional participation in policymaking	Survey	Health professionals (380)
Charbonneau and Van Ryzin (2013)	PMR	Attitudes towards performance	Survey	Citizens (595)
Van De Walle and Van Ryzin (2011)	PA	Citizen satisfaction	Split-ballot survey	Citizens (830 + 808)
James (2011a)	PA	Citizen attitudes towards performance	Field	Citizen Internet users (9,500)
Kaufmann and Feeney (2014)	PA	Perceptions of red tape	Survey	MPA students (81)
James and Moseley (2014)	PA	Citizen perceptions of performance	Field	Citizens (332 + 292)
Grimmelikhuijsen and Meijer (2014)	JPART	Trust in governments	Online	Citizens (570)
Jakobsen (2013)	JPART	Citizen participation	Field	Immigrant families (334 + 280)
Belle (2014)	JPART	Leadership, social impact, motivation	Field	Nurses (138)
Nielsen and Baekgaard (2013)	JPART	Political decision making	Survey	Danish city councillors (844)

(continued)

Appendix 1. (Continued)

<i>Author(s) and date</i>	<i>Journal</i>	<i>Topic</i>	<i>Experiment type</i>	<i>Sample (n)</i>
James (2011b)	<i>JPART</i>	Perceptions of performance	Field & laboratory	Citizens (field 351 + 413 + 427; lab 100 + 100)
Brewer and Brewer (2011)	<i>JPART</i>	Motivation and performance	Laboratory	Undergraduate students (40)
Riccucci, Van Ryzin, and Lavena (2014)	<i>JPART</i>	Representative bureaucracy and legitimacy	Online survey	Adults in the CivicPanel Project (789)
Herian et al. (2012)	<i>JPART</i>	Public perceptions of government, fairness	Survey	Citizens (197)
Yackee (2014)	<i>JPART</i>	Participation and policymaking	Survey	Citizens (388)
Avellaneda (2013)	<i>JPART</i>	Local government decision-making	Survey	Incumbent mayors (120)

Note: *Journals include *Public Administration Review (PAR)*, *Public Administration (PA)*, *Public Management Review (PMR)*, and *Journal of Public Administration Research and Theory (JPART)*.