PAR PUBLIC ADMINISTRATION REVIEW

Sebastian Jilke
Rutgers University–Newark
Steven Van de Walle
Erasmus University Rotterdam, The Netherlands

Soonhee Kim
KDI School of Public Policy and Management, South Korea

Generating Usable Knowledge through an Experimental Approach to Public Administration

Symposium Introduction

Abstract: *This introduction to the symposium on experimental methods in public administration shows how using experimental methods generates not only research that is empirically credible, but also relates to the real world of public administration. The ten articles in the symposium subject classic public administration theories or hypotheses that have been generated in nonexperimental research to rigorous testing using experimental methods. The first group of articles consists of studies with citizens who interact with government. The second group consists of three studies with public officials.*

In a 2012 editorial, *Public Administration Review* (*PAR*) Editor in Chief James L. Perry called upon prospective authors to more frequently submit experimental research to *PAR*. He went on to argue: "Well-designed experiments, in combination with other methods, hold the prospect for advancing our pursuit of usable knowledge" (Perry 2012, 480). Perry was not alone in calling upon public administration scholars to consider experiments in their research, and we can witness a strong increase in the use of experimental methods in the field in recent years. One may even go as far as to speak of an experimental turn that public administration is currently experiencing. This turn will hopefully help public administration scholars to improve our science to generate more usable knowledge.

An experimental approach to public administration research is not merely about conducting an experiment. It is also a distinct logic of inquiry that can be applied to a variety of settings. The core of experimental reasoning is characterized by using exogenous (i.e., external) predictors that do not share variation with any error terms (Harrison and List 2004). This means that researchers try to minimize error in the estimation of causal effects by introducing some kind of variation that comes from outside the model. Within a randomized experiment, this is done by deliberately intervening in the real world. Remler and Van Ryzin (2010) call this *making things happen*. One example would be by introducing a performance-related pay (PRP) scheme into government agencies. The effects of PRP on, say, job satisfaction could now be easily measured, one would assume. However, here the *fundamental problem of causal inference* occurs because we cannot observe the same unit at the same time in a treated *and* an untreated state (Holland 1986). Thus researchers use the process of randomization to create statistically equivalent groups of treated and untreated units. This would mean that organizations, or offices, would need to get randomly assigned to operate under a PRP scheme. Differences between treatment and control groups would then arguably display the *average treatment effect* of introducing PRP.

Randomized experiments are not the only way of employing an experimental logic to the design and analysis of data. Yet they are commonly referred to as a benchmark—in terms of their empirical credibility—against which (non-)experimental studies can be evaluated (Angrist and Pischke 2010). As argued earlier, introducing exogenous (i.e., external) variation into one's model can also be done without deliberately intervening in the data-generating process. One can also simply search in the real world. This can be done through the use of quasi-experimental methods, and/or within the context of so-called natural experiments (Dunning 2012; Shadish, Cook, and Campbell 2001). For example, quasi-experimental techniques such as those using instrumental variables try to create an exogenous state that allows researchers to estimate the causal effect of a given variable (see, e.g., Sovey and Green 2011). Another example is the regression discontinuity design, which makes use of thresholds as treatment assignment, by, for example, comparing subjects who have just passed a threshold, or cut-point to those who just missed it (Lee and Lemieux 2010; see, e.g., Olsen 2015). Natural experiments, in turn, try to find situations where the occurrence of an event can be regarded as-if random (Dunning 2012; see, e.g., Jilke 2015). While not a randomized experiment

Sebastian Jilke is assistant professor in the School of Public Affairs and Administration at Rutgers University–Newark. He studies behavioral public administration and is interested in the politics and psychological underpinnings of citizen-state interactions, especially in relation to administrative reforms. Other areas of interest include work-related attitudes and behaviors of government workers (especially discrimination at the front line of public services), development administration, and research methodology (especially measurement and [quasi-] experimental techniques).
E-mail: sebastian.jilke@rutgers.edu

Steven Van de Walle is professor of public management at Erasmus University Rotterdam, The Netherlands. His research focuses on interactions between citizens and public services, and comparative public sector reform.
E-mail: vandewalle@fsw.eur.nl

Soonhee Kim is professor of public administration at the Korea Development Institute (KDI) School of Public Policy and Management.
E-mail: soonheekim@kdischool.ac.kr

Public Administration Review, Vol. 76, Iss. 1, pp. 69–72. © 2015 by The American Society for Public Administration.
DOI: 10.1111/puar.12488

in a narrow sense, these methods try to improve the empirical credibility of their findings by employing an experimental logic of inquiry to the study of nonexperimental data. Indeed, like in evidence-based medicine (e.g., Oxford Centre for Evidence-Based Medicine 2009), such methods can be arranged within a hierarchy of evidence that displays their empirical credibility when it comes to their ability to minimize systematic and unsystematic errors.

But what does this mean for the relation between experimental and nonexperimental methods, such as traditional cross-sectional surveys or case studies? When should we use the one, and when the other—or should we, given the often claimed empirical superiority of experimental research in terms of internal validity, abandon nonexperimental methods altogether? Empirical credibility alone, in our opinion, is not sufficient when deciding which method best fits our research questions. It is an important, but not the only consideration that should guide us. Lower levels of empirical credibility can be acceptable, for example, if the research question is innovative (or novel), or if additional knowledge is needed to complement results from experimental data to increase external validity or to generate new theory. A public administration approach to experimental research needs to balance the use of experimental methods with traditional nonexperimental methods. This way, an experimental approach to public administration would greatly improve our science and generate usable knowledge for scholars and public professionals alike.

The articles in this symposium show how using experimental methods generates not only research that is empirically credible, but also research that relates to the real world of public administration. Classic public administration theories or hypotheses that have been generated in nonexperimental research are subjected to rigorous testing using experimental methods. This exemplifies the complementarity of research designs in the process of knowledge development.

The Symposium Articles

This symposium consists of this introduction and eight experimental studies plus two articles that represent a tradition of experimentation as is used in experimental economics (yet, these articles are not experimental in a strict sense). The studies have been grouped according to their unit of analysis—i.e., the subjects they study. The first group includes seven studies with citizens who—in one way or another—interact with government institutions, and the second group consists of three studies with public officials (or students in their assumed roles as public officials). These studies include a broad range of methods, ranging from randomized survey and field experiments to more stylized studies that represent the experimental economics tradition (e.g., Kagel and Roth 1992). Not only do the articles reflect plurality in experimental methods, but they also display a great diversity of research areas (including topics such as performance management, representative bureaucracy, public service motivation [PSM], or discrimination), as well as a strong international scope (with ten contributions from six different countries).

Experiments with Citizens

In their opening article, Baekgaard and Serritzlew (2015) present two randomized survey experiments among a large and diverse adult population in Denmark. They challenge the fundamental assumption that citizens interpret objective, clear, and unambiguous governmental performance information in an unbiased manner. The results they present show that people instead interpret performance information in accordance with their prior ideological beliefs, thereby potentially misinterpreting government performance information metrics. This important finding adds to an increasing literature on decision-making and individual-level biases in interpreting (performance) information.

Information about poorly performing services can lead to a host of citizen responses such as leaving the service provider in question, complaining and voting against the incumbent, but also other forms of individual or collective action. Yet to do so, citizens need to attribute blame for poor outcomes to governmental actors. Especially when services are delivered by contracted units, this may prove challenging.

James, Jilke, Petersen, and Van de Walle (2015) conduct a large-scale survey experiment among a diverse adult population in the United Kingdom to investigate who citizens blame for poor public service outcomes when services have been delegated to a unit within government, or contracted out to a private provider. Their findings show that providing citizens information cues about contracted-out delivery does not shift blame away from political principals—but information about in-house delegation does so to some extent.

Similarly, Marvel and Girth (2015) conducted a randomized survey experiment among U.S.-based respondents to investigate whether contracting out to private providers shifts blame for poor outcomes away from politicians, and whether this blame-shifting is affected by the length of the accountability chain between service providers and politicians. They find that contracting out does not shift blame attributions away from politicians, and that the length of the accountability chain does not seem to matter. In sum, both experiments provide complementary evidence that it may be difficult for political actors to deliberately escape public blame for poor services by contracting out to private providers.

Hvidman and Calmar Andersen (2015) explore a related topic in a survey experiment. They look at people's stereotypes of public vs. private service delivery in a hospital setting. The effect of publicness on four dimensions of perceived performance is studied. Results show that sector cues change perceptions of performance, with public hospitals seen as more benevolent and private hospitals as more efficient.

In a randomized survey experiment among a diverse adult population in the United States, Riccucci, Van Ryzin, and Li (2015) study the effects of symbolic representation on citizens' willingness to coproduce. Adding to an established body of literature on representative bureaucracy, the theory of symbolic representation suggests that if public officials' demographics match their client characteristics, it will lead to an increase in citizen trust and cooperation. Consequently, the authors test whether the symbolic representation of clients' gender promotes coproduction. The study's results show that females are more likely to coproduce policy

outcomes when public announcements contain female instead of male names of public officials. The study is a great example of how experimental research can probe the microfoundations of an established body of research.

Whether citizen participation can also be increased through more direct policy interventions is tested in Arceneaux and Butler's (2015) experimental study. In their randomized survey experiment among citizens of a U.S. town, the authors assessed the effectiveness of social recognition and offering skills training on people's volunteering intentions (i.e., asking to be redirected to a web page where they could sign up to serve on a city committee). The interventions, however, did not increase citizens' willingness to participate. In fact, offering training to prospective volunteers actually decreased volunteering intentions. These results not only highlight the boundaries of nudge-like behavioral interventions (i.e., social recognition), but also point toward the importance of experimentally testing policy interventions at the local level before rolling them out nationally.

Adres, Vashdi, and Zalmanovutch (2015) argue that citizen participation in collective action depends on their individual level of globalism—this is the degree to which people are globalized. In their study, they test whether high levels of individual-level globalization are related to individuals' contributions to local public goods using a series of modified public good games across four countries. These behavioral games are a crucial component of an experimental economics tradition. The study's findings indicate that individual-level globalization is indeed negatively associated with contributions to the public good. On the basis of their results, the authors conclude that through individuals' level of globalization, participation in collective action can be decreased.

Experiments with Public Officials

The second set of articles uses public officials as subjects. In their large-scale randomized field experiment across German cities, Grohs, Adam, and Knill (2015) investigate whether municipal administrators respond to citizen requests differentially when they come from ethnic minorities. To do so, they build upon the audit-study tradition by creating fictitious e-mail requests using German- and Turkish-sounding e-mail aliases (including varying their gender and the nature of the request). The requests were sent to all German cities with a population larger than 25,000 inhabitants. Examining response speed, completeness, and service orientation, results show no systematic discrimination based on gender or ethnicity. The authors do find some evidence (positive and negative) for ethnic discrimination conditional on gender—but only in terms of service orientation. The study highlights the usefulness of audit-type field experiments to investigate potential discriminatory practices of public officials.

Hess, Hanmer, and Nickerson (2015) report from two randomized field experiments among public officials in two U.S. states. They look at whether simple interventions (e-mail reminders and brief staff trainings) increase agency officials' compliance with the National Voter Registration Act of 1993 (i.e., starting or increasing to register their clients to vote). Both types of interventions have been found effective in increasing voter registration, but only among agencies that already started efforts to register their clients in the past. The interventions did not work for offices that had low performance in terms of voter registration prior to the experiment. These findings show how relatively low-effort interventions can increase agency performance among organizations that already made some effort in the past, but that for those offices most in need of improvement more robust oversight may be needed.

In the final article of the symposium, Esteve, Urbig, van Witteloostuijn, and Boyne (2015) conduct two public good games among a sample of undergraduate students. Though not public officials, these students were questioned about their levels of PSM. The authors were interested in linking levels of PSM to individuals' pro-social behavior (i.e., their revealed contributions to the public good). Results show that subjects who report higher levels of PSM were more likely to contribute to the public good, and that this association depends on the overall contributions of the entire group. The findings suggest that for PSM to have an impact on behaviors of public officials, a fertile organizational context is needed. The study also shows how public good games can be used to deepen insights into a domain of public administration research largely dominated by cross-sectional survey research.

Conclusions

The symposium provides an introduction and overview into the use of experimental methods within public administration. It gives us the opportunity to reflect on the appropriate relationship between experimental and nonexperimental research within the field. Clearly, an experimental public administration is in its infancy and will require more time and maturation to locate its place within the methodological toolbox of public administration scholars and practitioners. We hope the symposium will contribute to this development. The results from the contributions within the symposium provide credible evidence in many important areas. At the same time, we hope they will also encourage and inspire public administration scholars and professionals to more frequently *experiment* within the studies they conduct, thereby generating usable knowledge.

References

Adres, Eitan, Dana R. Vashdi, and Yair Zalmanovitch. 2015. Globalization and the Retreat of Citizen Participation in Collective Action: A Challenge for Public Administration. *Public Administration Review* 76(1): 142–52.

Angrist, Joshua D., and Joern-Steffen Pischke. 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics. *Journal of Economic Perspectives* 24(2): 3–30.

Arceneaux, Kevin, and Daniel M. Butler. 2015. How Not to Increase Participation in Local Government: The Advantages of Experiments When Testing Policy Interventions. *Public Administration Review* 76(1): 131–39.

Baekgaard, Martin, and Søren Serritzlew. 2015. Interpreting Performance Information: Motivated Reasoning or Unbiased Comprehension. *Public Administration Review* 76(1): 73–82.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences*. Cambridge: Cambridge University Press.

Esteve, Marc, Diemo Urbig, Arjen van Witteloostuijn, and George Boyne. 2015. Prosocial Behavior and Public Service Motivation. *Public Administration Review* 76(1): 177–87.

Grohs, Stephan, Christian Adam, and Christoph Knill. 2015. Are Some Citizens More Equal than Others? Evidence from a Field Experiment. *Public Administration Review* 76(1): 155–64.

Harrison, Glenn W., and John A. List. 2004. Field Experiments. *Journal of Economic Literature* 42(4): 1009–55.

Hvidman, Ulrik, and Simon Calmar Andersen. 2015. Perceptions of Public and Private Performance: Evidence From a Survey Experiment. *Public Administration Review* 76(1): 111–20.

Hess, Douglas R., Michael J. Hanmer, and David W. Nickerson. 2015. Encouraging Local Compliance with Federal Civil Rights Laws: Field Experiments with the National Voter Registration Act. *Public Administration Review* 76(1): 165–74.

Holland, Paul W. 1986. Statistics and Causal Inference. *Journal of the American Statistical Association* 81(396): 945–60.

James, Oliver, Sebastian Jilke, Carolyn Petersen, and Steven Van de Walle. 2015. Citizens' Blame of Politicians for Public Service Failure: Experimental Evidence about Blame Reduction through Delegation and Contracting. *Public Administration Review* 76(1): 83–93.

Jilke, Sebastian. 2015. Job Satisfaction and Regime Change: Evidence from a Natural Experiment. *International Public Management Journal*, DOI: 10.1080/10967494.2015.1043168.

Kagel, John H., and Alvin E. Roth, eds. 1992. *The Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Lee, David S., and Thomas Lemieux. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature* 48: 281–355.

Marvel, John D., and Amanda M. Girth. 2015. Citizen Attributions of Blame in Third-Party Governance. *Public Administration Review* 76(1): 96–108.

Olsen, Asmus L. 2015. Naming Bad Performance: Does Naming and Shaming Work? Paper presented at the Midwest Political Science Association Meeting 2013. Chicago.

Oxford Centre for Evidence-Based Medicine. 2009. *Levels of Evidence*. http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009.

Perry, James L. 2012. How Can We Improve Our Science to Generate More Usable Knowledge for Public Professionals? *Public Administration Review* 72(4): 479–82.

Remler, Dahlia, and Gregg G. Van Ryzin. 2010. *Research Methods in Practice: Strategies for Description and Causation*. Thousand Oaks, CA: SAGE Publications.

Riccucci, Norma M., Gregg G. Van Ryzin, and Huafang Li. 2015. Representative Bureaucracy and the Willingness to Coproduce: An Experimental Study. *Public Administration Review* 76(1): 121–30.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Cengage Learning.

Sovey, Allison J., and Donald P. Green. 2011. Instrumental Variables Estimation in Political Science: A Reader's Guide. *American Journal of Political Science* 55(1): 188–200.

Public Administration Review wants you to be the first to know about

upcoming research in public administration.

Sign up to receive content alerts of recently released articles by going to

http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6210.