Replication, experiments and knowledge in public management research

Walker, Richard M.; James, Oliver; Brewer, Gene A.

Published in:
Public Management Review

Published: 01/01/2017

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Published version (DOI):
10.1080/14719037.2017.1282003

Publication details:
Walker, R. M., James, O., & Brewer, G. A. (2017). Replication, experiments and knowledge in public management research. *Public Management Review*, *19*(9), 1221-1234. Advance online publication. https://doi.org/10.1080/14719037.2017.1282003

Replication, Experiments and Knowledge in Public Management Research

Richard M. Walker,
Laboratory for Public Management and Policy,
Department of Public Policy,
College of Liberal Arts and Social Sciences,
City University of Hong Kong,
Tat Chee Avenue, Kowloon Tong, Hong Kong
+852 3442 8760
rmwalker@cityu.edu.hk

Oliver James,
Department of Politics,
University of Exeter, UK
o.james@exeter.ac.uk

Gene A. Brewer,
Department of Public Administration and Policy,
School of Public and International Affairs,
University of Georgia, USA
geneabrewer@uga.edu

Acknowledgement: This article was supported by RGC Project No.: 9042434 (CityU 11611516) project title: Organizational design and public management: Extending the experimental methods agenda and through City University of Hong Kong Department of Public Policy Conference Grant and College of Liberal Arts and Social Sciences Capacity Building Grant.

Replication, Experiments and Knowledge in Public Management Research

Abstract

Replication is increasingly recognized as an important part of knowledge production in the social sciences, especially for experimental research. However, despite growing use of experiments, replication is little discussed or practiced in public management. We review the approach to replication taken by research in leading public management journals and note its scarcity. We then use a typology developed by Tsang and Kwan (1999) to classify the experimental replications undertaken by the articles in this special issue, which reveals a substantial variety of approach. We conclude by suggesting that replication is undertaken for different purposes and present a protocol about replication for experimental public management research.

Introduction

Replication is increasingly recognized as an important issue in the social sciences, from exact replication of a study that operationalizes the same design, intervention or measure of the original study to broader forms of replication that relax these constraints or seek to examine new populations or contexts. These notions of replication sit at the heart of scientific progress, which involves building cumulative knowledge by following appropriate process (e.g. Kuhn 1996; Freese 2007; Francis 2012; Nosek and Lakens 2014; Van Witteloostuijn 2016). An important aspect of replication is assessing the empirical implications of theories under similar and dissimilar conditions so as to build confidence in (or falsify) the theory, thus advancing the frontiers of knowledge. Replication is discussed and undertaken, sometimes controversially, in many social science disciplines (Freese 2007; Nosek and Lakens 2014). In public management, questions about confirming knowledge are as old as attempts to study it scientifically. However, interest in replication has grown especially as scholars have expanded their repertoire of research designs to include experimental methods.

Replication of studies using experimental methods is important because it helps check the validity of knowledge from previous research and enables questions concerning generalization across populations or contexts to be discussed. Experiments offer especially good potential for undertaking replications because the experimental treatments and the context in which they are conducted can be clearly specified by researchers, enabling them to be repeated. In public management, the growth in studies employing experimental methods has not been matched by replication; very few experiments have been replicated to date. The current situation of limited replications arises in part because there is a tendency to publish research about novel theory and procedures, and also positive results—rather than to publish studies that repeat procedures or fail to reject the null hypothesis. Overall, presenting novel ideas and rejecting the null hypothesis tends to be more valued by journal editors, reviewers,

and many researchers themselves than the incremental accumulation of knowledge that comes through replication and its potential for producing null findings (Van Witteloostuijn 2016). A possible outcome is that false positives are likely to be disproportionately reported because of these preferences for original studies that produce large and statistically significant results.

This article contributes to the debate on replication in the context of experimental research designs in public management. First, the article assembles the key arguments in the wider social sciences for replication. It then hones down and examines the way in which authors use the term replication in three leading public management journals, finding that its use and shared meaning is not commonplace. Second, the debate on replication in the wider social sciences is typically associated with a discussion of the inability to exactly replicate a substantial proportion of prior studies. In the field of public management, context and timing is likely to influence the findings. Non-replicability may be common and the reasons for this may be complex and directly related to the type of replication attempted (Freese 2007; Lewin 1943; Morrell and Lucas 2012). To this end. we present Tsang and Kwan's (1999) replication classification scheme as a valuable tool for the study of public management. The studies presented in this special issue are classified according to this typology and the results reveal that in public management, scholars use several types of replication. Finally, we suggest that replication performs different functions so it is important not to frame a single set of rigid prescriptions. However, good practice guidance can be given contingent on the type of replication being undertaken. In this way, public management would benefit from considering a protocol for conducting different types of replication studies.

Experimental Methods and Replication

Public management research draws on the full methodological repertoire of the social sciences in the search for valid and reliable evidence. It has favoured observational

techniques, especially because they are considered to be the most suitable form of research for delivering practically relevant research findings. The repertoire has grown to embrace the use of experimental methods and the number public management studies adopting various types of experimental methods (including field, laboratory and survey types) has grown substantially in recent years (Blom-Hansen and Serrizlew 2015; Bouwman and Grimmelikhuijsen 2016; Margetts 2011; James, Jilke and Van Ryzin 2017). Recent studies implementing experimental methods have focused on a range of important public management topics including citizen-government relations (James 2011; Walker et al. 2013; Van Ryzin 2013; Grimmelikhuijsen and Meijer 2014; James and Petersen 2017), equity (Jakobsen and Andersen 2013), rules (Kaufman and Feeney 2014), and decision-making (Avellaneda 2011).

Properly implemented experiments hold an attraction because they offer strong internal validity for establishing causal effects. The researcher manipulates the variable of interest through an intervention with subjects typically being randomly assigned to receive the intervention treatment or not. A simple two-group study in this way establishes treatment and control groups that are, on average, the same except for the presence or absence of the intervention. Measures of outcome variables enable researchers to observe differences between the two groups and estimate the average treatment effect. Estimating the causal effects of public management interventions using experimental data avoids many of the problems associated with observational studies that attempt to use control variables, but are unable to eliminate all rival causes, which can include a multitude of confounding and spurious variables (Blom-Hansen and Serrizlew 2015; James, Jilke and Van Ryzin 2017).

Like all research methods, experiments have their limitations, but carrying them out within a structured approach to scientific investigation, which should include replication, has increasingly been recognized as part of more general concerns about the credibility of social

science findings. The debate in psychology is illustrative of this trend where major initiatives have been undertaken to replicate studies (for example, see Nosek 2015). Much attention has focused on direct replication where researchers attempt to recreate the conditions believed to be sufficient for obtaining a previously observed finding. Whilst replication is thought to be easier in laboratory studies than in less structured contexts where control of interventions and related conditions is more difficult, it is discussed and undertaken, sometimes controversially, in many social science disciplines (Freese 2007; Nosek and Lakens 2014).

Replication and the Practice of Public Management Research

To assess the contemporary practice of replication in the public management literature, we performed a *Google Scholar* search for use of the term replication (search term replicat*). This search was conducted in the spring of 2016. A *Google Scholar* search was performed because search engines such as *Web of Science* or *Scopus* search only titles and abstracts whereas *Google Scholar* searches the entire article.[1] The search focused on *Journal of Public Administration Research and Theory* (JPART), *Public Administration Review* (PAR) and *Public Management Review* (PMR), and uncovered 261 mentions of the word replication. The findings are presented in Table 1.

[Table 1 about here]

Table 1 indicates that across the three journals, the most frequent occurrence of the word replication was as a descriptor (e.g., verb, noun or adjective). For example, authors would discuss the challenges of replicating the real world context in laboratory experiments (Bozeman and Scott 1992), or the ways in which policies could be replicated to share learning and best practice (Ma 2014). This accounted for nearly half of all identified cases.

[1] A review by Walker, Lee and James (2016) using the Web of Science identified only 10 articles in the *Journal of Public Administration Research and Theory*. Of course, a limitation of this approach is that replication may be undertaken without using the term replication, but searching for the term at least allows the inclusion of studies that are self-aware of what they are doing and that relate to the broader literature explicitly about replication's role and importance in the scientific process.

Replications, in the sense of seeking to retest the findings of other studies or employing very similar research designs or methods, were implemented by public administration scholars in 15% and 20% of the articles, respectively. The replication of others' research was more likely to be found in JPART and PAR than in PMR. Examples of studies include Van Ryzin's (2006) replication of his prior work on expectancy disconfirmation theory using a different method and sample, Bellé's (2015) replication of studies of motivational crowding effects, and Rhu's (2013) replication and extension of O'Toole and Meier's study on the consequences of contracting.

The replication of methods was distributed a little more evenly across the three journals. When authors sought to replicate methods from other studies, they typically focused on issues of measurement and replicated measurement scales and questions in the search for valid research instruments (e.g. Fernandez et al. 2015). Authors using case study methodology drew upon the arguments of Yin (2003) on the merits of multiple case studies as a form of replication to build external validity across contexts (Piening 2011; Zhu 2014).

Given the limitations of generalizability in most social science research studies, some authors used the word replication when discussing the need to establish external validity (e.g., Avellaneda 2013; Ma 2013). The last category of studies that used the word replication offered to make their datasets publicly available so that others could perform replications with them. Studies offering data for the purposes of reproducing the original study and using the word replication when discussing it were often associated with Kenneth J. Meier and several of his associates, who have long advocated making datasets publicly available to enable the reanalysis of data.

A Typology for Replication

Scholars can approach replication in different ways. Above we noted the varied use of the term in three leading public management journals that focused on explicit attempts to

replicate prior studies or partial elements of those studies, notably research designs and methods. However, to help make replications more successful, a systematic framework is needed. Such as framework should more clearly articulate the different forms of replication, discuss the roles of methods, and provide standards for evaluating studies and their evidence. In a sense, literal or exact replications are not possible because a prior study cannot be reproduced in every detail. At a minimum, time has passed and subjects have changed (Rosenthal 1991). Researchers who adopt a historical, idiographic method, that views research findings as the unique outcome of complex interactions between contexts and the researchers studying them, tend to think this issue is fundamental to much research activity. However, they extend this critique more broadly and emphasize how most forms of research are difficult to replicate. Yet mainstream social science recognizes that there are considerable benefits from replication of different kinds. Replication will likely become more important, particularly as the use of experimental methods grows in the discipline, because such methods involve very clear research procedures that others should be able to reproduce. Thus, to develop an improved understanding of replication studies in public management research, we suggest that the Tsang and Kwan (1999) typology is a useful starting point.

Tsang and Kwan (1999) identify several different types of replication studies by contrasting several study elements: these include studies using the 'same measurement and analysis' or 'different measurement and/or analysis'; and studies using the 'same dataset', 'same population' or 'different population'. This framework results in six different forms of replication, as shown in Table 2. Whilst Tsang and Kwan's terminology suggests that different populations consist of different participants (for example, different groups of public managers in different service areas or jurisdictions), their ideas also apply to the additional dimension of context (for example, the same public managers operating in different policy domains or institutional settings).

[Table 2 about here]

The first two types of replication are relevant to research integrity, and especially the transparency and openness debate (Nosek 2015; Nosek et al. 2015). The first type is called 'checking of analysis'. This type uses the same dataset as the original study, and the same measurement and analysis. This emphasizes the need for authors to make their data available to others, as noted above. Note that some journals in allied disciplines, such as economics, political science and psychology, require datasets to be made publically available at the review or publication stage. Second, when different measurement or analysis is applied to the same dataset, the replication is referred to as a 'reanalysis of data'. This approach may be taken because newer or more sophisticated analytical tools have been developed since the original study was conducted. This approach allows the validity of evidence to be reassessed using new methods. Nevertheless, these first two approaches focus on procedural issues and do not address concerns about external validity.

Third, an exact replication involves the same research procedures, measurement and analysis using the same population and context, but with a different sample of participants from that population. This could, for example, involve using a different sample of public managers from the larger population of public managers. This approach differs from the checking of analysis approach because the validity of the prior study's findings is examined using the new sample. If the replication reveals similar results, it reinforces the findings reported in the original study. (The challenges of conducting an exact replication are noted above.)

The fourth type of replication is labelled a conceptual extension. This type of study involves different measurement and analysis compared to the original study, while drawing a sample from the same population. A conceptual extension is a more sophisticated replication because it introduces new procedures and explores additional ramifications of the original

finding(s). As such, a single conceptual extension replication is not able to verify prior findings comprehensively because null findings may arise from the alternative measurement and analysis. The conduct of conceptual replications may, therefore, require a number of replications using similar and dissimilar measurement and analysis. Otherwise, the results will not be clear and the state of knowledge will not be advanced.

The last two types of replication are performed on different populations from the original study, but they differ by using either the same or different measurement and/or analysis. The fifth type, empirical generalizations, uses the same research design, measures and analysis, and assesses whether the original findings hold up in different populations. The sixth type, generalizations and extension, uses a different population and does not implement the original measurement and analysis. Rather, it seeks to broaden or extend the original finding's berth. However, if the results are different from those reported in the original study, discrepancies could be due to the altered research design or changes in the population. Interpretation of the replication results is thus more complex. Nonetheless, generalizations and extensions represent an important step in testing the findings of a public management study in one context and applying them to another. Findings from a study with both a different population and context with variation to the research design can help refine theory and strengthen external validity. When broadly defined in this way, replications are somewhat indistinguishable from mainstream social science research.

The articles included in this *Public Management Review* special issue undertake replications and implement experimental research designs. We map these articles onto the Tsang and Kwan framework in Table 3 to illustrate the scope and breadth of experimental replications that public management researchers are undertaking. Note that all four types of replication shown in Table 3 are being conducted. Also note that only two studies used the same population while seven varied either populations or populations and measurement. This

suggests that there is a strong tendency for public management scholars to forego exact replications and attempt to extend the original study's findings or test new theory. One challenge of this approach is that when the replication moves away from the original study's methodology, including its population and measurement, it becomes more difficult to interpret the results in relation to the original study.

[Table 3 about here]

In this special issue, an exact replication was conducted as part of an overall study in which Grimmelikhuijsen and Porumbescu (2016) also conducted a conceptual replication to extend knowledge on expectancy disconfirmation theory. Implementing exact replications in public management research is challenging because it can be difficult to obtain funding and access to target populations. Using the same population of US citizens, two articles conduct conceptual replications and corroborate the findings by applying different measurement tools or instruments. Van Ryzin, Riccucci and Li (2016) extend their prior study of representative bureaucracy by adapting their measures from the policy arenas of policing and recycling to study emergency preparedness. Grimmelikhuijsen and Porumbescu (2016) seek to extend Van Ryzin's (2013) study on expectancy disconfirmation theory by varying the manipulations to determine whether less subtle treatments could change the original findings.

Empirical generalizations, which involve using a different population but the same measurement, are performed in four studies. Gaardboe, Filtenborg and Sigsgaard-Rasmussen (2016) implement a research design that is similar to Grimmelikhuijsen and Porumbescu (2016). The authors take 'baseline' tests of expectancy disconfirmation theory prior to generalizing and extending the study. However, given that they are undertaking their study in Denmark rather than the USA, the first experiment they report is an empirical generalization. Jilke et al. (2016) and Lee, Moon and Kim (2016) use empirical generalizations to test the robustness of research findings from one context to another. Jilke et al. (2016) also highlight

the importance of measurement equivalence to ensure that the concepts under investigation hold in the different context prior to drawing conclusions, and offer guidance on this important issue.

Generalizations and extensions are also undertaken on four occasions. Gaardboe, Filtenborg and Sigsgaard-Rasmussen's (2016) Danish study of expectancy disconfirmation theory also includes different policy arenas. George et al. (2016) replicate Nielsen and Baekgaard's (2015) study of performance information use among politicians in a different context, Belgium rather than Denmark, and extend the study to include strategic goals. They are able to qualitatively replicate the findings of the original study but cannot find empirical evidence to support the extension. Kaufmann and Tummers (2016) replicate Tummers et al.'s (2015) US red tape study in the Netherlands. They also substitute student subjects for citizens (an issue that we will return to below). Finally, Olsen (2016) tackles a perennial question in public policy research – inaction – and extends it to public management among Danish citizens.

Towards A Protocol for Replication in Public Management

This special issue of *Public Management Review* demonstrates that there is an appetite for undertaking replication studies in the public management scholarly community. Table 3 described a variety of approaches to replication, with most types at least changing the study population. To assist with the development of replication studies in public management, a protocol is outlined to assist researchers who implement replications. The protocol is built from Tsang and Kwan's six types of replications. We suggest that replication performs different functions so it is important to be somewhat flexible and suggest practices that are consistent with the type of replication being undertaken. Table 4 lists the six replication approaches and proposes a number of questions that scholars should pose in relation to subjects, data, methods, analysis, and whether the researchers conducting the replication are

the same as those that undertook the original study. Below, we provide a step-by-step guide to the key questions and challenges for each type of replication.

[Table 4 about here]

1. Checking of measures and analysis

Checking of measures and analysis is likely to be conducted by different researchers because it is largely an audit function for the purpose of ensuring research integrity. Reanalysis of data may also be done by the same researchers. This raises an important question about who should conduct the replication, and this concern is also relevant to other types of replication. The key question is: should the replication be done by the same researchers who completed the original study, or should different, independent researchers undertake the task? An argument in favour of the latter is that independent researchers may have less emotional attachment to the findings and less reluctance to disconfirm them, so independent researchers can be more impartial. An argument against independence is that involving the original researchers can be beneficial because they know how to replicate the study and interpret the findings. There is, for example, considerable tacit knowledge in research that cannot be codified easily. However, given the need for research findings to be objective and independent of who conducts the research, a more independent approach to replication is generally preferable, although there is likely to be variation on a case-by-case basis.

Checking of measures and analysis is a straightforward task so long as the original researchers keep careful and accurate records of their subjects, data, methods and analytical techniques. Researchers implementing the replication must be able to understand and implement the original study, and to recognize why discrepancies might arise. Another issue is the pre-registration of experimental designs, which is increasingly seen as a way to ensure fidelity to the original research intentions (Nosek and Lakens 2014). With a replication, such statements of purpose are easier to formulate because the research builds explicitly on

previous designs, including their interventions, contexts and measures.

2. Reanalysis of data

Reanalysis of data requires transparency with data and analytical techniques provided by the original authors. Researchers implementing the reanalysis of data have to be theoretically guided by the possibility of new measurement, if the dataset permits. Similarly, if new analytical tools are to be applied, the researchers should provide a detailed rationale and explain why different analytical techniques might lead to different results.

3. Exact replication

Exact replication similarly requires detailed information on the prior study's execution. Researchers must make decisions about the new subjects. However, given that the subjects are drawn from the original population, the main concern is with the sampling approach.

4. Conceptual replication

Conceptual extensions, as with empirical generalizations, require researchers to be very clear on the purpose of the study and the hypotheses they are testing: in this case, theory, and refining measurements to see if the findings are upheld in different contexts. Conceptual extensions require access to the same population; thus, sampling procedures need to be considered and justified. Differences in measurement, as shown in some of the articles included this special issue, can range from alternative contexts, including different countries or policy arenas, or can extend to improvements in measurement.

5. Empirical generalization

Empirical generalizations should be relatively easy to implement since the original measurement and analysis plan remains unchanged. However, shifting to different contexts may prove challenging. Jilke et al. (2016) suggest a set procedures in public management research to ensure that measurement is the same in different places. Thus, researchers should be aware that changing context and time may fundamentally influence the research findings.

Using different subjects may in some cases be a simple process of substituting the citizens of one country for another – see Gaardboe, Filtenborg and Sigsgaard-Rasmussen (2016), Jilke et al. (2016), and Lee, Moon and Kim (2016). However, an important topic for public management replications, addressed in Kaufmann and Tummers (2016) and touched on by Walker et al. (2017), concerns the role of students as subjects – a practice common in many behavioural science studies. Public management researchers often assume that students, particularly MPA students, are training to be managers and are thus appropriate subjects for management research. Yet little systematic evidence is provided to support this assertion. Conversely, researchers might expect that studies conducted on practicing managers will produce stronger support for the hypotheses being tested, while those conducted on students will offer weaker results. This is because students may lack practical experience and public managers are the objects of the theories and hypotheses being tested. Again, there is little systematic evidence to support this assumption. Replications should thus compare findings across the various subject groups to assess their degree of similarity and appropriateness for future research.

6. Generalizations and extensions

Generalizations and extensions raise all of the questions posed above. A generalization and extension moves the replication the furthest distance from the original study. Where appropriate, replications should consider the issue of measurement equivalence raised by Jilke et al. (2016). In considering the replication's purpose, Bouwman and Grimmelikhuijsen (2015) remind us of three concerns regarding the realism of the experiment: its setting, the real-world context, and the nature of the treatment. Their review of studies in public administration that implement experimental methods concludes that the external validity of the research designs employed are quite high. They point towards the widespread use of treatments that resemble the real world, but note that subjects do not

always correspond with the people that actually experience the stimuli under investigation. Replication represents a challenging test of the external validity of a study. Undertaking an experiment in a different setting, using a different population or sample, or extending the original design are fruitful ways for researchers to test how well their findings hold up.

Conclusion

We have sought to raise questions about, and advance an agenda for, the use of replication in public management research, especially for studies implementing experimental research designs. Replication is not very widespread in public management, partly because replications are difficult to publish and faculty tenure and promotion practices often reflect a 'publish or perish' mentality. In encouraging researchers to conduct replication studies, we have suggested that Tsang and Kwan (1999) offer a useful classification scheme that could be used to guide scholars considering what type of replication to undertake. We have built on and extended the Tsang and Kwan framework by sketching out a protocol that can be used to guide replication investigations. Finally, we hope the review and framework offered here, and the replication studies published in this special issue, encourage others to undertake replication studies. It is our belief that replication is an essential part of the scientific process that can help produce a sounder knowledge base for our field.

References

Avellaneda, C. N. 2013. "Mayoral Decision-making: issues 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.

Belle, N. 2015. "Performance-Related Pay and the Crowding Out of Motivation in the Public Sector: A Randomized Field Experiment." *Public Administration Review* 75 (2): 230-241.

Blom-Hansen, J. and S Serritzlew. 2015. "Special Issue: Experiments in Public Management Research." *International Public Management Journal* 18 (2): 151-342.

Bouwman, R. and S. Grimmelikhuijsen. 2016. "Experimental Public Administration from 1992-2014: A Systematic Literature Review and Ways Forward." *International Journal of Public Sector Management* 29 (2): 110-131.

Bozeman, B. and P. Scott. 1992. Laboratory Experiments in Public Policy and Management. *Journal of Public Administration Research and Theory* 2 (3): 293-313.

Fernandez, S., W. G. Resh, T. Moldogaziev and Z. W. Oberfield. 2015. "Assessing the Past and Promise of the Federal Employee Viewpoint Survey for Public Management Research: A Research Synthesis." *Public Administration Review* 75 (3): 382-394.

Francis, G. 2012. "The Psychology of Replication and the Replication of Psychology." *Perspectives on Psychological Sciences* 7 (6): 585-594

Freese, J. 2007. "Replication Standards for Quantitative Social Sciences. Why Not Sociology?" *Sociological Methods & Research* 36 (2): 153-172.

Gaardboe, F., A. Filtenborg and J. Sigsgaard-Rasmussen. 2016. "An Experimental Test of the Expectancy Disconfirmation Theory of Citizen Satisfaction." *Public Management Review.*

George, B., S. Desmidt, P. A. Nielsen, and M. Baekgaard. 2016. "Rational Planning and

Politicians' Preferences for Spending and Reform: Replication and Extension of a Survey Experiment." *Public Management Review.*

Grimmelikhuijsen, S. 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

Grimmelikhuijsen S. and G. Porumbescu. 2016. "Reconsidering the Expectancy-Disconfirmation Model: Three Experimental Replications." *Public Management Review.*

Jakobsen, M. and S. C. Andersen. 2013. "Coproduction and Equity in Public Services." *Public Administration Review,* 73 (5): 704-713.

James, O. 2011. "Performance Measures and Democracy: Information Effects on Citizens in Field and Laboratory Experiments." *Journal of Public Administration Research and Theory*, 21 (2): 399-418.

James, O., S. Jilke, G.G. Van Ryzin 2017. "Causal Inference and the Design and Analysis of Experiments" in *Experiments in Public Management Research: Challenges and Contributions* edited by James, O. , S. Jilke, G.G. Van Ryzin. Cambridge: Cambridge University Press.

James O. and C. Petersen 2017. "International Rankings of Government Performance and Source Credibility for Citizens" Public Management Review.

Jilke, S., N. Petrovsky, B. Meuleman and O. James. 2016. "Measurement Equivalence in Replications of Experiments: When and Why It Matters and Guidance on How to Determine Equivalence." *Public Management Review.*

Kaufman, W. and M.K. Feeney. 2014. "Beyond the Rules: The Effect of Outcome Favourability on Red Tape Perceptions." *Public Administration* 92 (1): 178-191.

Kaufmann, W. and L. Tummers. 2016. "The Negative Effect of Red Tape on Procedural

Satisfaction." Public Management Review.

Lee, M.J., M.J. Moon and J. Kim. 2016. "Insights from Experiments with Signaling Games: Rational Incremental Decision-making." *Public Management Review*

Lewin, K. 1943.. "Defining the "Field at a Given Time"." *Psychological Review* 50 (2): 292–310.

Ma, L. 2013. "The Diffusion of Government Microblogging." *Public Management Review* 15 (2): 288-309.

Ma, L. 2014. "Does Super-Department Reform Improve Public Service Performance in China." *Public Management Review* 18 (3) 369-391.

Morrell, K and J.W. Lucas. 2012.. "The Replication Problem and its Implications for Policy Studies." *Critical Policy Studies* 6 (2): 182-200

Nielsen, P.A., and M. Baekgaard. 2015. "Performance Information, Blame Avoidance, and Politicians' Attitudes to Spending and Reform: Evidence from an Experiment." *Journal of Public Administration Research and Theory* 25 (2): 545–569. doi:10.1093/jopart/mut051.

Nosek, B.A. 2015. "Estimating the reproducibility of psychological research." *Science,* 349 (6251): doi: 10.1126/science.aac4716.

Nosek, B.A. and Lakens, Daniel 2014. "Registered reports. A method to increase the credibility of published results." *Social Psychology* 45 (3): 137-141

Nosek, B.A., G. Alter, G.C. Banks, D. Borsboom, D., S. D. Bowman, S. J. Breckler, ... and M. Contestabile. 2015. "Promoting an Open Research Culture: Author Guidelines for Journals Could Help to Promote Transparency, Openness, and Reproducibility." *Science*, 348 (6242): 1422-1425

Olsen, A. 2016. "Responding to Problems: Actions are Rewarded Independently of the Outcome." *Public Management Review*

Piening, E.P. 2011. “Insights into the Process of Dynamics of Innovation Implementation.” *Public Management Review* 13 (1): 127-157.

Rho, E. 2013. “Contracting Revisited: Determinants and Consequences of Contracting Out for Public Education Services.” *Public Administration Review* 73 (2): 327-337.

Rosenthal, R. 1991. “Replication in Behavioral Research.” In J. W. Neuliep (Ed.), *Replication Research in the Social Sciences*: 1-30. Newbury Park, CA: Sage.

Shadish, W.R., Cook, T.D. and Campbell, D.T. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton Mifflin: Boston, MA

Tsang, E.W.K. and K-M. Kwan. 1999. “Replication and Theory Development in Organizational Science: A Critical Realist Perspective.” *Academy of Management Review* 24 (4): 759-780

Tummers, L., U. Weske, R. Bouwman, and S. Grimmelikhuijsen. 2016. “The Impact of Red Tape on Citizen Satisfaction: An Experimental Study.” *International Public Management Journal* 19 (3): 320–341.

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.

Van Ryzin, G.G. 2013. “An Experimental Test of the Expectancy-Disconfirmation Theory of Citizen Satisfaction.” *Journal of Policy Analysis and Management* 32 (3): 597-614.

Van Ryzin, G.G. N. Riccucci and H. Li. 2016. “Representative Bureaucracy and its Symbolic Effect on Citizens: A Conceptual Replication.” *Public Management Review*

Van Witteloostuijn, A. 2016. “What Happened to Popperian Falsification? Publishing Neutral and Negative Findings: Moving away from Biased Publication Practices.” *Cross Cultural & Strategic Management* 23 (3): 481-508.

Walker, R.M., G.A. Brewer, B. Bozeman, M.J. Moon, and J. Wu. 2013. “An Experimental

Assessment of Public Ownership and Performance: Comparing perceptions in East Asia and the United States." *Public Management Review* 15 (8): 1208-228.

Walker, R.M., Lee, M.J. and O. James 2017. "Replications of Experimental Research: Implications for the Study of Public Management" in *Experiments in Public Management Research: Challenges and Contributions* edited by James, O. , S. Jilke, G.G. Van Ryzin. Cambridge, Cambridge University Press.

Yin, R.K. 2003. *Case Study Research: Design and Methods* (3rd edn). Thousand Oaks, CA: Sage.

Zhu, X-F. 2014. "Mandate Versus Championship: Vertical Government Intervention and Diffusion of Innovation in Public Services in Authoritarian China." *Public Management Review* 16 (1): 117-139.

Table 1: Uses of the word replication in three PA journals, column percentages in parenthesis

	Journal of Public Administration Research and Theory	Public Administration Review	Public Management Review	Total
Replication of others' research	16 (17.4%)	17 (15.7%)	5 (8.2%)	38 (14.6%)
Replication of methods	20 (21.7%)	16 (14.8%)	15 (24.6%)	51 (19.5%)
Replication required to establish generalization	13 (14.1%)	14 (13.0%)	9 (14.8%)	36 (13.8%)
Data available for replication	4 (4.3%)	8 (7.4%)	1 (1.6%)	13 (5.0%)
Use of word 'replication'	39 (42.4%)	53 (49.1%)	31 (50.8%)	123 (47.1%)
Total	92	108	61	261

Table 2: Different types of replication

	Same measurement and analysis	Different measurement and analysis
Same dataset	Checking of analysis	Reanalysis of data
Same population	Exact replication	Conceptual replication
Different population	Empirical generalization	Generalization and extension

From: Tsang and Kwan (1999)

Table 3: Types of replications presented in the PMR special issue

Exact replication	*Conceptual replication*
Grimmelikhuijsen and Porumbescu (2016) Exp 1	Grimmelikhuijsen and Porumbescu (2016) Exp 2 & 3 Van Ryzin, Riccucci and Li (2016)
Empirical generalization	*Generalization and extension*
Gaardboe, Filtenborg and Sigsgaard-Rasmussen (2016) Exp 1 Jilke et al. (2016) Lee, Moon and Kim (2016)	Gaardboe, Filtenborg and Sigsgaard-Rasmussen (2016) Exp 2 & 3 George et al. (2016) Kaufmann and Tummers (2016) Olsen (2016)

Table 4: Replication protocol

	Researchers involved	Population/Subjects	Methods	Analysis
1. Checking measures and analysis	Different researchers or same as original study?	No change, use of original subjects	No change, use of original methods	Re-implementation of prior analysis
2. Reanalysis of data	Different researchers or same as original study?	No change, use of original subjects	No change, use of original methods	No change, use of original methods
3. Exact replication	Different researchers or same as original study?	Same population but different subjects: what subject to choose and why?	No change, use of original methods	No change, use of original methods
4. Conceptual replication	Different researchers or same as original study?	Same population but different subjects: what subject to choose and why?	Different methods: what methods and why?	New analysis: what analytical techniques and why?
5. Empirical generalization	Different researchers or same as original study?	Different subjects: what subjects to choose and why?	No change, use of original methods	No change, use of original methods
6. Generalization and extension	Different researchers or same as original study?	Different subjects: what subjects to choose and why?	Different methods: what methods and why?	New analysis: what analytical techniques and why?