

实证社会科学

(第五卷)

主编 钟杨

社会科学研究的实验、复制与中国学者的使命
吴建南

第三方公共服务绩效评估可靠吗——广东省的案例研究

马亮 杨媛 刘柳

公共和私人组织绩效感知：来自调查实验的证据

Ulrik Hvidman Simon Calmar Andersen 刘遥 吴建南 译

组织科学中的可重复性检验与理论发展：以批判实在论为视角

Eric W.K. Tsang Kai Man Kwan 谢恩 译

社会科学定量研究的复制标准——为什么社会学不这样做呢

Jeremy Freese 廖福崇 王思琦 朱春奎 译

公共管理研究中的复制、实验和知识

Richard M. Walker Oliver James Gene A. Brewer 阎波 程齐佳徵 韩东伶 周鼎元 译

启动效应、复制，与最难的科学

Joseph Cesario 孟凡蓉 李思涵 译

复制、复制

Gary King 赵娟 孟天广 译

Social Science Research



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS



实证社会科学

Social Science Research

(第五卷)

钟 杨 主编



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

图书在版编目(CIP)数据

实证社会科学. 第五卷 / 钟杨主编. —上海: 上海交通大学出版社, 2018

ISBN 978-7-313-19990-4

I. ①实… II. ①钟… III. ①社会科学-文集 IV. ①C53

中国版本图书馆 CIP 数据核字(2018) 第 190465 号

实证社会科学(第五卷)

主 编: 钟 杨	地 址: 上海市番禺路 951 号
出版发行: 上海交通大学出版社	电 话: 021-64071208
邮政编码: 200030	
出 版 人: 谈 毅	
印 刷: 江苏凤凰数码印务有限公司	经 销: 全国新华书店
开 本: 787mm×1092mm 1/16	印 张: 11.25
字 数: 186 千字	
版 次: 2018 年 9 月第 1 版	印 次: 2018 年 9 月第 1 次印刷
书 号: ISBN 978-7-313-19990-4/C	
定 价: 42.00 元	

版权所有 侵权必究
告 读 者: 如发现本书有印装质量问题请与印刷厂质量科联系
联系电话: 025-83657309

实证社会科学

Social Science Research

主办单位：上海交通大学国际与公共事务学院

主 编：钟 杨

副 主 编：吴建南 樊 博

编委会成员：(按姓氏笔画排列)

边燕杰(西安交通大学)

杨开峰(中国人民大学)

吴建南(上海交通大学)

陆 铭(上海交通大学)

陈 捷(上海交通大学)

庞 珣(清华大学)

钟 杨(上海交通大学)

唐世平(复旦大学)

敬义嘉(复旦大学)

蓝志勇(University of Arizona)

李连江(香港中文大学)

肖唐镖(南京大学)

何艳玲(中山大学)

陈映芳(上海交通大学)

邱泽奇(北京大学)

赵鼎新(University of Chicago)

唐文方(University of Iowa)

阎学通(清华大学)

谢 宇(Princeton University)

樊 博(上海交通大学)

编辑部成员：

钟 杨 吴建南 樊 博 陈映芳 刘帮成

陈永国 黄琪轩 陈慧荣 陈 拯 魏英杰

杜江勤 韩广华 杨 姗 邓玉玲

目 录

专题研讨	/ 1
社会科学研究:实验、复制与中国学者的使命	吴建南 / 3
第三方公共服务绩效评估可靠吗?——广东省的案例研究	马 亮 杨 媛 刘 柳 / 8
公共和私人组织绩效感知:来自调查实验的证据	Ulrik Hvidman Simon Calmar Andersen 刘 遥 吴建南[译] / 27
组织科学中的可重复性检验与理论发展:以批判实在论为视角	Eric W.K. Tsang Kai Man Kwan 谢恩[译] / 47
社会科学定量研究的复制标准——为什么社会学不这样做呢	Jeremy Freese 廖福崇 王思琦 朱春奎[译] / 81
公共管理研究中的复制、实验和知识	Richard M. Walker Oliver James Gene A. Brewer 阎 波 程齐佳微 韩东伶 周鼎元[译] / 97
启动效应,复制,与最难的科学	Joseph Cesario 孟凡蓉 李思涵[译] / 113
复制,复制	Gary King 赵 娟 孟天广[译] / 130
书评	/ 147
权势变动、议题领域竞争性与国际组织变革	——评《世界秩序再谈判:国际关系中的制度变迁》 陈 拯 / 149
数据库介绍	/ 159
亚洲晴雨表调查数据库介绍	翟一达 / 161
投稿须知	/ 169

CONTENTS

Special Topic

Social Science Research: Experiment, Replication and the Mission of Chinese Scholars	Wu Jiannan / 3
Are Public Service Performance Evaluations Reliable? Evidence from Guangdong Province	Ma Liang; Yang Yuan; Liu Liu / 8
Perceptions of Public and Private Performance: Evidence from a Survey Experiment	Ulrik Hvidman, Simon Calmar Andersen; Liu Yao, Wu Jiannan / 27
Replication and Theory Development in Organizational Science: A Critical Realist Perspective	Eric W.K. Tsang, Kai Man Kwan; Xie En / 47
Replication Standards for Quantitative Social Science: Why not Sociology?	Jeremy Freese; Liao Fuchong, Wang Siqu, Zhu Chunkui / 81
Replication, Experiments and Knowledge in Public Management Research	Richard M. Walker, Oliver James, Gene A. Brewer; Yan Bo, Cheng Qijiahui, Han Dongling, Zhou Dingyuan / 97
Priming, Replication, and the Hardest Science	Joseph Cesario; Meng Fanrong, Li Sihan / 113

Replication, Replication	Gary King; Zhao Juan, Meng Tianguang / 130
--------------------------	--

Book Review

Power Transition, Policy Areas' Propensity for Competition, and Reforms of International Organizations: Review on <i>Renegotiating the World Order: Institutional Change in International Relations</i>	Chen Zheng / 149
---	------------------

Introduction of Social Science Data

Asian Barometer Survey	Zhai Yida / 161
------------------------	-----------------

专题研讨

社会科学研究：实验、复制与中国学者的使命

吴建南*

科学研究旨在发现规律并予以证明。尽管所发现的规律 $Y = F(X)$ 并予以证明在自然界和人类社会有着不同的特点,但按照唯物主义的观点,如世界是物质的、物质是运动的、运动是有规律的、而规律是可以为人们所认识的,我们总可以不断努力发现特定规律并且证明,尽管这个过程在不同历史时期可能需要不同的时间、不同的方法、不同的数据来源等。既然要证明,那就意味着每个研究都需要考虑是否可以被复制。

改革开放 40 年来,中国社会科学得以迅猛发展。各个学科的国际顶级刊物也出现越来越多的中国学者的论文,国际学术会议也越来越多在中国举行。毫无疑问,国际社会科学研究的中国声音正在积极唱响!就国内发表而言,据不完全统计,代表中国社会科学研究较高水平的 CSSCI 来源文献数量在 2009 年已近百万,来源期刊数超过 500 本,涵盖社会学科的不同领域。然而,对于特定研究是否可以复制的关注却亟待提升。

近期的文献梳理表明,已有关注主要包括:①对科研成果出版物质量的担忧,指出现有期刊论文不可被重复验证的问题(吴小康,2014;胡传鹏等,2016;黄小茹,2017);②研究方法的探讨,比如对单一实验效度的质疑(杜晓新,2002),提出实验方法的复制性和可控性优势(周星、林清胜,2004;孙慧荣,2007);③少数几项复制实验(蔡建红,1999;任衍具等,2007;廖飞等,2010)。涉及学科以心理学为主,其余还包括经济学、社会学、管理学等。陈云松和吴晓刚(2012)以定量分析技术的视角,初次较为系统地倡导了社会科学中的“复制性研究”。

笔者本科和硕士期间从事的是电子工程学科的离子源研究,在导师指导下,那时我们做了一个又一个实验,总结出特定类型离子源的最佳参数并制造

* 吴建南,上海交通大学国际与公共事务学院教授、中国城市治理研究院研究员。

出来提交给用户。很明显,基于大量重复与非重复实验的研究使我们产品性能稳定,而我们的实验内容与结果也发表在美国物理学会的期刊上与同行分享(尽管某些关键参数作为技术诀窍我们并未分享)。1999年,笔者在Syracuse大学做博士后,所接触到的第一个研究就是实验研究,2008年,笔者和Richard Walker教授多次讨论公共管理领域的实验研究,2010年我们也尝试做了社会认同、公共服务动机与合作的实验,但在回应评阅人意见、修改论文的过程中,经历了若干波折,曾经一度考虑以真实的公务员为对象重做当时的实验。

2014年5月,Richard Walker教授成立了香港城市大学公共管理证据实验室(Public Management Evidence Lab),作为该实验室国际理事会理事,我参加其年会并参与其中的一些工作。得益于实验研究良好的“可被重复操作性”,以Richard Walker教授为代表的一批学者选择了实验研究作为复制的载体,并在复制实验研究中取得了重要进展,如*Public Management Review*期刊在2017年第9期上专门刊载了8篇复制实验研究,2017年开始组织香港、首尔、上海、哥本哈根等十余个城市的学者进行复制实验研究,并于2018年1月在首尔召开了研讨会,交流初步研究结果。

2018年是中国极为重要的一年,这一年中国的改革开放整整40年了!在我们回顾中国改革开放40年的同时,也得回顾中国哲学社会科学的40年研究。作为中国科学学与科技政策研究会青年工作委员会(简称“青工委”)的负责人,我和来自不同学科青工委的同事们将复制研究作为2018年的重要工作之一,特别是从较容易组织实施的翻译工作开始。通过梳理我们发现,即便是国际社会科学界,不同学科对于复制研究的认识也并不相同;而且,复制研究并不局限于实验研究,很多学科都关注研究的可复制性和复制研究。

本期第二篇文章是由中国人民大学公共管理学院马亮副教授与他的学生所作。他们对两项不同来源的公共服务满意度调查数据进行了横纵比较。基于A大学和B机构4年公众满意度评价的横向比较与各自每年评价情况的纵向比较,这篇文章探讨了第三方公共服务绩效评估研究的有效性,反映了同一研究者在不同时间以及不同研究者在相同时间所从事研究的可重复性。

第三篇是刘遥博士生和我翻译的发表在*Public Administration Review*上的一篇实验研究论文,探讨的是公共组织和私人组织感知绩效的问题,分析“公共性”(Publicness)对公众组织感知绩效评价的影响。该实验已在上海、香港、首尔、哥本哈根等十余个城市复制,目前正在分析数据、进行整体和局部的

论文撰写，期待不远的将来与读者分享。

第四篇文章是谢恩教授翻译的 *Replication and theory development in organizational science: a critical realist perspective* 一文，这篇文章以批判实在论的视角，探讨了组织科学中的可复制性问题以及可复制性检验对该领域理论发展的重要作用。论文首先从批判实在论的视角回应了社会科学不可复制的问题；其次给出了复制的六个类别，阐释了可复制性检验对理论发展具有拟合与预测、证实与证伪的重要作用；论文还通过复制性研究与创新性研究的对比指出，虽然创新性研究看上去比可重复性检验更具理论魅力，而后者对于理论发展也是不可或缺的；论文最后指出复制研究将有助于建立具有更牢固基础的长效知识积累机制。

第五篇文章是朱春奎教授与他的团队翻译的 *Replication standards for quantitative social science: Why not sociology?* 探讨了定量社会学研究中的复制标准。这篇文章始于对越来越复杂的定量研究可复制性的担忧，主张社会学在文章发表时，要建立在线数据库，公开尽可能多的数据信息，以便其他研究者能检验已发表成果的可复制性。文中还比较了经济学相对社会学论文发表时信息公开方面的优势，指出应当让社会学研究者意识到公开必要信息是优秀期刊发表其投稿的必要代价。

第六篇文章是阎波副教授和他的团队翻译的 *Replication, experiments and knowledge in public management research*，探讨了公共管理领域中实验研究的复制问题。这篇文章梳理了近年来国际顶级公共管理期刊上发表的实验研究，采用测量与分析核查、数据再分析、精确型复制、概念型复制、经验一般化和推广扩展等六项分类对这些文章所进行的实验复制进行分析，指出不同类型复制具有不同研究目的，阐述了不同类型复制的方法和准则。

第七篇是孟凡蓉副教授及其学生翻译的 *Priming, replication, and the hardest science*，主要探讨了心理学中启动效应的复制问题。论文指出心理学领域已越来越多地关注研究结果的可复制性和稳健性，并以启动效应为例来分析复制的本质以及它与心理科学的关系，指出复制的重要性不是体现在它自身，而在于它代表所产生的知识能被社会验证。论文也指出失败的复制研究可能是由多种因素导致的，对复制结果的过高期望是不合理的。

第八篇是孟天广副教授与其合作者翻译了 Gary King 在 1995 年发表的 *Replication, replication*，探讨了经验政治科学中的复制问题。这篇论文首先指出复现的标准就是有足够的信息来理解、评估和信赖既有研究，以使第三方

可以不借助任何来自作者的额外信息,即复现其结果。论文还强调复现标准应当同等地适用于定量和定性研究。然而,当前的学术规范却不鼓励复制研究,据此,文章最后对教师、学生、论文写作者、研究生、作者、评审人、资助机构以及期刊和书籍编辑们提出了一些建议。

复制(replication)对现有社会科学研究体系提出了一个重大挑战,以往实证分析中的纰漏都会一个个显现出来。但这也是一个大好时机,我们可以借此机会进一步规范实证研究,增强研究结果的信度和效度,并在复制的过程中发现新理论、新知识,共同推进中国社科研究的“科学化”进程!

对于中国学者而言,在改革开放 40 年之际,开始关注复制研究是一个重要的开始。这首先意味着中国学者应该直面中国社会发展存在的诸多问题,基于现实的观察、抽象,形成基于中国实践的概念、命题与假设,在中国的这个最具变化的时代,以现代科学技术的各种方法手段,对中国的 $Y=F(X)$ 进行证明;在此过程中,要特别关注中国地域辽阔、差异较大的效应,在发现和验证中国理论的基础上,进一步考虑谋划超越国界的比较研究,探索中国研究成果的可复制性,形成中国特色社会科学理论!

参考文献

- 蔡建红. 中国人印象形成中核心品质研究方法初探[J]. 江西师范大学学报(哲学社会科学版), 1999(1):43-47.
- 陈云松, 吴晓刚. 走向开源的社会学 定量分析中的复制性研究[J]. 社会, 2012, 32(3):1-23.
- 杜晓新. 单一被试实验研究中的效度问题[J]. 中国特殊教育, 2002(3): 21-24.
- 胡传鹏, 王非, 过继成思, 等. 心理学研究中的可重复性问题:从危机到契机[J]. 心理科学进展, 2016, 24(9):1504-1518.
- 黄小茹. 科研成果不可验证性问题——发现机制的失效及可能的对策[J]. 科学学研究, 2017, 35(7):961-966.
- 廖飞, 茅宁, 刘文娟. 不确定下的管理者期权思维:一个基于复制实验的探讨[J]. 现代管理科学, 2010(1):90-93.
- 任衍具, 禰宇明, 傅小兰. 编码时间对序列视觉信息整合的影响[J]. 心理科学, 2007, 30(5):1077-1080.
- 孙慧荣. 实验经济学的研究方法与应用[J]. 商场现代化, 2007(27):393-393.

- 吴小康. 可复制性与经验研究的透明化[J]. 经济学动态, 2014(10):121 - 129.
- 周星, 林清胜. 交叉学科研究的典范:实验经济学和行为经济学[J]. 经济学家, 2004(3):85 - 90.

第三方公共服务绩效评估可靠吗？ ——广东省的案例研究*

马 亮 杨 媛 刘 柳**

摘 要:公民满意是政府绩效评估的核心问题之一,公共服务绩效则是政府绩效的重要组成部分。目前已有较多公共服务绩效评估项目增加了公民满意度调查,但是对这些调查结果的有效性却少有研究。本文以广东省为例,对 A 大学和 B 机构的公共服务满意度调查数据进行横向对比和纵向分析,试图揭示其评估的有效性。总体上看,两个机构的公民满意度调查结果有较大的相似性,并且二者逐年接近,说明公民满意度调查可以较为有效地反映公共服务绩效。研究发现,A 大学评估结果的纵向稳定性较差,而 B 机构的纵向结果相对稳定。基于对两个项目有效性的探究,本文提出了改进公民满意度调查的政策建议。

关键词:公民满意度;公共服务绩效评估;有效性;政府绩效

一、引言

公共服务绩效是指公共服务提供的数量和质量,它是政府绩效的重要组成部分。政府的公共属性、目标的多元性、运作的非交易性导致绩效评估的复杂性,但是政府绩效评估的核心问题是指向“公民满意”和“实现政府责任”(郑方辉,2013)。因此,公民调查往往被视为公共服务绩效评估的重要一环。因为公民作为公共服务的接受者,对政府公共服务的供给情况有更加直观的感受。

* 本文受国家自然科学基金面上项目“第三方评估如何提升组织绩效:中国地方政府的实证研究”(批准号 71774164)的资助。

** 马亮,中国人民大学公共管理学院副教授;杨媛,中国人民大学公共管理学院硕士研究生;刘柳,中国人民大学公共管理学院硕士研究生。

受(陈振明, 刘祺, 蔡辉明, 2009), 通过问卷调查获得的公民满意度, 就成为衡量政府公共服务供给能力的重要标准之一。然而在现有研究中, 学者大多从公民参与政府绩效评估的必要性和积极意义, 以及完善政府绩效评估中公民参与机制的角度出发(郑方辉, 雷比璐, 2007; 郑方辉, 王琲, 2008), 而没有考虑公民评估是否能准确反映政府绩效这一理论预设(倪星, 李佳源, 2010)。已有的公共服务绩效评估框架, 大多是将公民满意度和客观绩效指标结合起来进行测量(陈振明, 刘祺, 蔡辉明, 2009)。基于简明实用和提高回答率的需要, 公民调查问卷大多设计简单, 一般通过李克特量表直接测量公民的满意度。这种过于空泛的满意度测量, 并不能有效获取公民对城市公共服务的感知水平(Brown and Coulter, 1983)。加上街头拦访这一调查形式的抽样误差, 使公民主观调查是否能够准确测量公共服务绩效, 成为一个值得研究的问题。

尽管这个问题至关重要, 但是我国学界对这个问题的回答寥寥无几。我们总结西方学者研究后发现, 公民满意度这一主观绩效评估指标并不总是稳定地反映政府部门的实际工作成效。Brown 和 Coulter 对美国警务服务公民满意度的研究发现, 公民的主观评估结果会受到种族、年龄、政治观点、自身经历以及对警务服务的期望等因素影响, 与警务工作的实际成效完全没有相关关系(Brown and Coulter, 1983)。Parks 虽然对 Brown 及 Stipak 的研究结论表示质疑, 但是在他新构建的模型中, 仍然发现客观绩效指标和主观满意度之间没有直接的相关性(Parks, 1984)。

在公共服务满意度测评中, 问卷的设计、样本的代表性、数据的处理方法等都会影响评估结果的准确性。为了检验公民主观测量的有效性, 一种方法是将主观评估结果与客观测量绩效进行相关性分析。但是, 这种方法会面临两个问题, 一是通过客观指标测量的绩效并不一定准确, 它会受到指标体系的影响, 也会与客观数据来源关系密切。我国现行的城市公共服务绩效评估大多采用政府档案或统计年鉴中的数据, 而这些数据本身的有效性就面临质疑。二是在进行数据分析时采用的模型也会影响分析结果, 而通行的线性模型可能有误。有学者指出, 随着公共服务质量的提高, 其带给服务使用者的边际效率是递减的。因此, 民众感知到的公共服务绩效和真实的公共服务水平并不一定呈现线性相关关系(倪星, 李佳源, 2010)。

不同于上一种方法, 另一种方法是将多个机构实施的公民调查进行相关性分析, 看同一城市同一项的公共服务测量是否一致。这种方法的优势在于,

公民调查所测量的内容在维度和性质上一致,因此更具有可比性。本文就使用该方法对公共服务绩效测评的有效性进行研究,并以广东省为例开展实证分析。

改革开放以来,广东省一直是我国经济和政治改革的试验田,在政府绩效评价方面也走在全国前列。2007年A大学首创独立第三方评价政府模式,对广东省所有的市县进行政府绩效评价,成为中国“第一个吃螃蟹”的机构。其独特之处在于,每年都会实施大规模的公民调查,来测量广东省居民对公共服务的满意度。与此同时,2008年起B机构每年也会发布针对广东省市县的公共服务绩效评价结果,涵盖服务型政府建设、供电专项、环境专项等多项内容。这些评价结果已经被广泛应用在广东省政府的各项决策中(郑方辉,毕紫薇,2009; Gao, 2012),但对这些评估的有效性却缺少研究。基于此,我们选择A大学的“幸福广东”和B机构的服务型政府调查展开比较,考察对广东省各市的两项评价结果是否一致。在评价模式上,它们均属于公众主观评价模式;在评价对象上,它们都涉及广东省21个地级市和若干县;在评价内容上,它们都是调查了经济、社会、生态等各个方面的政府服务情况;在评价结果上,都是以公民满意度为输出。因此,两个评价具有可比性,可以对其加以比较。

在以下部分,我们首先介绍本文使用的数据和方法,其次报告本文的主要发现,最后对数据分析展现的结果进行分析和讨论,并评价广东两项公民调查的有效性。

二、两个案例比较

A大学政府绩效评价中心和B机构省情调研中心分别于2007年和2008年开始对广东省下辖的地级市及部分县进行绩效评价,每年公开其评价结果。两项评价都包含了公共服务的公民满意度调查,但两个机构每年公布的评估结果却不尽相同。因此,有必要对两项评估的信度和效度进行检验。

(一) A大学政府绩效评价

A大学自2007年开始每年对广东省21个市和121个县进行绩效评价,至今已持续10年。作为全国第一个独立的第三方政府绩效评估,A大学的政府整体绩效评价吸引了全国各级政府的关注。每年发布的红皮书不仅会报告广东各市县的绩效排名,还会推出“幸福广东”指数,衡量公民对政府公共服务