

# 社会科学的因果推论与实验方法

——评 *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*

余 莎 耿 曙\*

Teele, D. Ed. (2014). *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven & London: Yale University Press. PP280.

## 一、引言：关于实验方法的大辩论

在过去三十年，实验方法的崛起为社科研究掀起了一场方法论革命。在很长一段时间内，社科研究都是基于现实世界的观察性数据来建立与检验因果关系。然而，这一研究途径（Routine）的局限很大：现实中的社会现象，要么多个因素同时变化，要么很难发生变化，使得基于经验观察的因果推论（Causal Inference）不太可靠；更为重要的是，经验观察难以排除不可观测或未知的干扰因素，使得自我选择、遗漏变量、反向因果等问题困扰因果推论（Przeworski, 2007）。对此种种困境，由于实验方法可以操纵研究关注的变量、随机分配研究对象，成为解决问题的最佳选择。<sup>②</sup>尽管在早期，出于实践或者伦理上的考虑，学界对于实验的应用不抱乐观，例如，著名的政治学者指出，“实验方法是最接

---

\* 余莎，上海财经大学公共经济与管理学院，博士候选人；通讯作者：耿曙（skeng0731@163.com），浙江大学公共管理学院社会学系，研究员。

② 实验是指研究者通过系统地干预数据生成过程（Data Generating Process）来搜集数据的过程。因此，实验研究与观察性研究（Observational Research）最大的区别就在于，研究者并不干预后者的数据生成过程，而是通过观察已经存在的数据来描述或解释现象。根据此定义，自然实验、政策实验、计算机模拟、思想实验等均不能看成真正意义上的实验。对此的详细讨论，请参考莫顿（Rebecca B. Morton）和威廉斯（Kenneth C. Williams）（Morton & Williams, 2010: 42 - 57）的相关论述。

近科学解释的理想方法，但不幸的是，由于实际上的或伦理上的局限，很少应用于政治科学”（Lijphart, 1971: 683 - 684）。但时至今日，实验方法凭借上述优势，逐渐成为理解社会现象的利器。不仅实验方法早成主流之一，其他如案例比较、定量统计也都是遵照实验的逻辑，来规划研究方案（Gerring & McDermott, 2007; Angrist & Pischke, 2009），实验方法俨然已成为社会科学因果推论的“黄金准则”（Gold Standard）（Holland, 1986）。

相较严格抽样的统计方法，实验方法尽管有内在效度（Internal Validity）的优势，但其外在效度（External Validity）却仍受质疑。例如，以学生或便利样本为实验对象能否代表研究总体，实验情景是否真实决策、能否模拟现实等（Falk & Heckman, 2009; 谢宇, 2012）。对此，实验学者常遵照“平行法则”（Parallelism Precept）进行重复实验（Smith, 1982），或尝试用随机抽样的方法招募实验对象（调查实验/网络实验）等手段来解决问题。但更有吸引力的方法是实地实验<sup>①</sup>——即在现实情境中对真实对象进行随机控制实验，因而它既能取得受控实验的内在效度的优势，同时又避免了外在效度的难题（Harrison & List, 2004; 罗俊等, 2015）。于是，在支持者看来，实地实验的内外兼修，是社科研究的万灵药（Panacea）。然而，它也还是未能幸免于批评：实地实验难以解决与未知干扰因素有关的异质性问题，实地实验的实施过程并不一定能保证完全的随机化，它还受到实际的或伦理的约束。当然，这些问题在不同的研究中表现不一、程度不一，实验学者也是想尽各种办法来回应与解决。对于实验方法是否为黄金准则，双方也僵持不下、对此展开论战。

《实地实验和批评：对社科实验的应用与误用的探讨 Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences》（简称《实地实验和批评》）就是这场方法论争辩的重要成果。针对“实验方法是否为社会科学因果推论的黄金准则”这一问题，社会科学各领域的诸多学者们展开了激烈的讨论：实验研究是否以及为何优于观察性研究？实验研究、实地实验进行因果推论面临哪些方法论的、实践的，与伦理的问题？理论与实验方法之间关系是什么？更为有趣的是，这场辩论并非仅见于笔端，而是发生在耶鲁大学的校园里。本书编者蒂尔（Dawn Teele）当时还是耶鲁大学政治学系的博士一年级学生，她于2009年10月组织了这场讨论，邀请这些学者就实验方法与实地实验进行现场辩论，吸引了数百师生赶来聆听这场切磋。正如蒂尔

<sup>①</sup> 本文将“Field Experiment”译成“实地实验”而非一般所谓“田野实验”，一是为了将其区别于社会学与人类学研究中的田野调查，其次也是因为这种方法强调的是在实地场景、现实生活中进行实验，译成“实地”比“田野”更为准确。

将此形容为“任何社科学子都不容错过”的盛会，书评作者在读罢论文集后，仍然仿佛置身会场，思潮澎湃，于是不揣简陋，将此书介绍给师友，增长见闻，切磋方法，此其一。其二，实验方法越来越受到国内社科学者的关注与应用<sup>①</sup>，未来更有迅速增长之势，因此本书中的辩论将有助于学界掌握实验方法、认识实验优劣。但毫无疑问，这场辩论远未结束，留下的问题可能多于解决的问题。下文将首先介绍辩论双方的观点，然后在评论的基础上提出还需进一步讨论的诸多问题。

## 二、辩论主题：实验是否因果推论的黄金准则

全书共收录了9篇论文，围绕着“实验方法是否社科研究因果推论的黄金准则”展开辩论。在全书的开篇第一章中，格伯（Alan S. Gerber）等（Gerber et al. : 9-31）<sup>②</sup>（后文简称GGK）支持这一论点，并在此基础上提出了备受争议的主张：相较于观察性研究，实验研究更能有效控制“未知干扰因素”而得到无偏估计，因此建议社科学者摒弃观察性研究，将所有的资源与精力用于实验研究。之后接续的章节都是或直接或间接地与GGK对话。然而由于学者们对于“实验方法是因果推论的黄金准则”的看法不一，且每位作者支持或者反对的理由不止一个，本文在此不按照章节顺序一一展示。取而代之，本文将这场辩论过程重新梳理为以下两个方面，以便更清楚地呈现这场辩论的观点交锋过程。

### （一）研究方法层次检讨“实验是否黄金准则”

用观察性数据进行因果推论，其根本难题在于无法控制“不可观测”或者“未知”的干扰因素（Unobservable Confounders），对此，实验的随机分配则可克服。这是GGK为代表的实验支持者所秉持的坚定信念。<sup>③</sup>在此基础上，GGK提出观察性研究完全不可信，应摒弃之。这一观点被斯托克斯（Suan C. Stokes）

---

① 代表作可参考《清华大学学报（哲学社会科学版）》2016年第4期的几篇实验研究，以及马得勇与兰晓航（2016）、孟天广等（2015）、余莎等（2015）。

② 此处省略年份，表示本书中该作者所著的相应章节。下文也做类似处理。

③ 不过书中并没有对其详细阐述，读者可以参见普泽沃斯基（Adam Przeworski）（Przeworski, 2007）、德鲁克曼（James N. Druckman）等（Druckman et al., 2011）的介绍。简单来说，随机分配可以保证研究关注的自变量与其他控制变量、不可观测的干扰因素都不相关，因此，即使遗漏了未知干扰也不会影响自变量的估计。

评价为“极端怀疑主义”(Radical Skepticism)(Stokes: 34 - 42)。她认为这一观点太过武断,因为观察性的研究者并非对于研究现象完全未知,潜在的干扰也并非无穷无尽,他们对研究相关理论的掌握、对数据生成过程的了解都可以帮助其判断潜在的干扰因素,并通过对象匹配(Matching)、平衡检验(Balance Test)、增加控制变量等多种手段来处理这些干扰。

如果上述回应是观察性研究者的“防守性”的辩护,那么斯托克斯还对实验方法发起了一轮“进攻性”的批评(Stokes: 42 - 52):实验方法并不能完全排除不可观测的干扰因素,因为当研究对象存在异质性时,他们对于实验处理的反应也存在异质性,如果是未知干扰导致的异质性,就会产生“异质性实验处理效应”问题<sup>①</sup>。换言之,随机分配确实可以保证实验处理与不可观测的干扰因素不相关,但若实验处理与不可观测的干扰因素存在交互作用,那么即便通过随机分配,也不能保证交互效应的外生性,得到“异质性实验处理效应”的无偏估计。这是全书对实验方法的批评焦点之一,大多数批评者都指出了这一问题。

用实验研究批评观察性研究的逻辑反击实验研究,这成为观察性研究者的“武装利器”。对于异质性实验处理效应问题,实验学者提出了重复实验、区块设计、因子设计等各种方法尝试探索处理效应的分布,但本质上并没有太大帮助,因为根据斯托克斯的批评,这些异质性永远不能被观察到。然而,班纳吉(Abhijit V. Banerjee)和迪弗洛(Esther Duflo)的回应却将这一辩论推向了新的方向(Banerjee & Duflo: 90, 105 - 109):异质性实验处理效应问题并非实验方法独有,观察性研究中也存在。实际上,不仅实验学者这么认为,部分观察性研究者也支持这一看法。如迪顿(Angus Deaton)在批评发展经济学领域的工具变量研究时,也暗示观察性研究同样不能幸免于“异质性”问题(Deaton: 144 - 149, 166)。今井(Kosuke Imai)等更是将其明确指出,纳入分析框架(Imai et al.: 196 - 227)。可见,由未知干扰产生的异质性实验处理效应并不会削弱实验方法在因果推论中的领先优势。

---

<sup>①</sup> 单位同质性(Unit Homogeneity)是因果推论的基本前提假设之一(King et al., 1994),即对每个研究对象而言,解释变量对于被解释变量的期望效应是相同的。与之相反,异质性表示解释变量对于被解释变量的期望效应因个体而异。在实验中,一般假设所有实验对象(包括处理组与对照组)对于实验处理的期望反应是相同的,因此可以通过估计平均处理效应(Average Treatment Effect)来得到期望效应;若不同,则存在异质性实验处理效应(Heterogeneous Treatment Effect)(Gerber & Green, 2012)。

## （二）研究设计层次检讨“实验是否黄金准则”

上述的讨论，都是基于“理想实验”（Ideal Experiment）的假设，存在完美的随机分配。然而在实践中，随机分配可能并不完美。实验室中进行随机分配相对容易：研究者明确知道、也可以要求所需的随机化过程。实地实验却复杂得多（Barrett & Carter: 60 - 71; Teele: 118 - 135）。首先，实地实验通常需要多方合作，研究者希望的完美随机分配，不一定符合实地实验资助者的需求；其次，实地实验的执行者在施行实验干预时，有动机偏离随机分配，尤其当涉及资源分配时；最后，研究对象是否同意参与随机分配的实验。除此之外，阻碍随机分配的还有伦理问题，例如，当实验可能带来负面效应或实验涉及利益分配时，随机分配是否公正，是否告知实验对象等等。

更进一步地，就算有完美的随机分配，也不等于实现了实验的内在效度，因为随机分配并不代表着实验对象“接受了”分配，受到了实验处理的影响。例如，被分配到处理组的某些实验对象没有接受干预，或是被分配到对照组的某些实验对象反而接受了干预，更有甚者，实验对象可能退出实验。这两个问题分别被称作“规则抗拒”（Noncompliance）、样本耗损（Attrition）。迪顿（Deaton: 167 - 175）指出，这些问题不容忽视，因为二者并不是外生的，与实验对象对实验结果的预期有关，而预期又会显著地影响实验结果。此时，简单地比较处理组与对照组已不能回答研究者关注的问题。

可见，实验实施过程能否保证随机分配、实验对象是否按照干预行动，都会影响到实验的因果推论。但针对这两个问题，实验学者的回应相对较弱，仅有防守性的辩护。GGK 在文中提及这两个问题会削弱实地实验的内在效度，增加实验研究的偏误，只能通过改进实验设计或统计方法来解决问题。班纳吉和迪弗洛（Banerjee & Duflo: 98 - 102）列举了发展经济学的实地实验设计，来展示研究者们尽力保证随机分配的各种安排：例如改变随机分配的层级单位、提供激励或者监督机制、在更易随机分配的环境进行实验等。但很难判断这些做法是否有效。对于第二个问题，实验学者受到观察性研究者的启发，通常采用工具变量法来解决问题——用“实验分配”做“实验处理”的工具变量。然而，这一做法也并没有让批评者满意，因为工具变量依然不能解决上述的异质性问题，而且工具变量的有效条件往往不易满足（Deaton: 144 - 160; Barrett & Carter: 68 - 69）。

### 三、评论与反思：从内、外在效度看实验方法的因果推论优势

由上可知,《实地实验和批评》为社科的读者们描绘了一幅激烈争辩的画面:一方面,实验学者在极力维护实验方法作为因果推论的黄金准则所具有的独特优势——随机分配消除未知干扰;而另一方面,观察性研究者却大力批评、甚至嘲笑这一优势是建立在乌托邦的假设之上——完美分配以及成功干预。如此争锋相对、鞭辟向里地呈现实验方法的优势与劣势,这在学界是非常罕见的,也是本书最吸引读者、最具有贡献的地方。

从书中讨论来看,对于实验方法的批评多于实验学者的回应。实验之所以被称作因果推论的黄金准则,是因为实验可以控制不可观测的未知干扰因素,得到清晰的因果关系。换言之,与观察性研究相比,实验方法更容易实现研究的“内在效度”,这是实验支持者的共识,但在本书的辩论中却受到了观察性研究者的挑战与质疑。更不用说,观察性的研究者还有批评实验方法的另一利器:实验研究的“外在效度”有限,甚至连实验学者也都默认(如GGK)。到此,这场辩论的结果似乎已见分晓。然而,若有更多篇幅,将辩论往前推进一步,我们可以发现这一判断也比较武断。若是将观察性研究与实验研究置于同等的位置来检视上述辩论所涉的各种方法论问题,无论是从“内在效度”还是“外在效度”来看,实验方法都仍具有相对优势,本节接下来将对其进行详细论证,以补充和丰富这场辩论。

#### (一) 确保内在效度:两种方法孰优孰劣?

根据上述分析,对于实验或观察的支持,在很大程度上取决于研究者对于“是否存在不可观测的未知干扰以及其是否产生异质性实验处理效应”的认识。这就产生了一个困境:如果研究者对于两者都未可知,那么任何社科研究对于现象的解释都是徒劳,因为我们不能得到任何完全可靠的因果推论。但在适当的假设前提下,社科研究的因果推论是可能的:(1) 如果不存在不可观测的未知干扰,实验研究与观察性研究的推论效果相同;(2) 如果存在不可观测的未知干扰,但不存在异质性实验处理效应,实验方法优于观察性研究;(3) 但如果这些未知干扰会导致异质性实验处理效应,那么实验方法也不能解决根本问题,即便研究者此时放弃估计异质性实验处理效应,只关注实验处理的平均效应,实验与观察同样都存在困难。其实,GGK、斯托克斯、迪顿以及班纳吉和迪弗洛(GGK、Stokes、Deaton以及Banerjee & Duflo)等的分歧也是在于对此选

择了不同的立场。然而，从总体来看，只要未知干扰不是产生异质性效应的来源，实验方法都略胜一筹。

理想的实验可以实现研究的内在效度，那么不理想或不完美的实验呢？面对批评者对于随机分配可行性、实验对象是否接受了处理等质疑时，支持实验的学者竭力证明他们努力做到最好：在预算与现实约束下的最优的实验设计、最精妙的实证策略。但这些努力通常被批评者（无意或有意地）忽略。然而，此处的问题关键却是，不完美的实验与有偏误的观察，谁更能实现因果推论？虽然这本质上是一个实证性的问题，但我们也可以在理论上稍加分析：首先，观察性研究有类似“随机分配”的过程，如匹配，但不能完全控制未知干扰（Imai et al. : 206 - 214），相比起来，通过推敲实验过程中各参与者偏离随机分配的激励，研究者比较容易找出误差来源进而矫正估计结果；其次，与观察性研究相比，实验研究用“实验分配”做“实验处理”的工具变量更容易满足工具变量的成立条件；最后，尽管实验不完美，但它可以为研究者展示数据的生成与收集过程，而观察性研究的数据生成与收集过程对于研究者来说却通常是未知的，后者就更难保证不出现模型设定错误、数据搜集过程的偏误（如测量误差、数据造假）等问题，因此从数据来源的透明度来看，不完美的实验也更有利于因果推论。

## （二）确认外在效度：两种方法孰优孰劣？

实验研究的外在效度也常被观察性的研究者质疑。所谓“外在效度”是指基于一项实验研究的结论可以推广到实验外的研究对象中、实验外的情境中。然而，一方面，从定义本身来看，任何一项研究，无论是实验的抑或观察性，要使其结论外推到研究外的对象与情境中，都存在障碍，<sup>①</sup>观察性研究本质上并不会比实验研究面临更少的问题。例如，利用观察性数据考察教育与政治立场的关系，使用某一国家的公民样本如何推广到其他国家或代表总体？再例如，几个发达国家的案例比较如何能得出普遍性的结论？另外，观察性研究为了寻

---

<sup>①</sup> 正如麦克德莫特（Rose McDermott）（McDermott, 2011）所强调，仅有一项研究是不可能得到外在效度的，无论研究范围多大多广；研究的外在效度只能通过重复检验得到。这一主张也得到了德鲁克曼和金辛迪（Cindy D. Kam）（Druckman & Kam, 2011）的支持，他们认为这比较符合波普尔式的因果论，因果关系永远不可能得到证实，经验证据是从多重检验中积累起来的，不管这些检验是否有局限性。

求内在效度还会牺牲外在效度。<sup>①</sup>

另一方面，从方法论的角度来看，如果一项研究在不同的样本或者不同的情境中得到不同的结果，那么可以认为“因果效应”具有异质性，样本的特征或者情境的特征就决定了“因果效应”如何发生变化。例如，仍然是考察教育对政治立场的影响，假设来自于民主国家的实验证据不足以揭示威权国家将会如何，那么探索教育对政治立场的因果效应（处理效应）如何随着政治体制的变化而变化就至关重要，也成为提高外在效度的重要方法。换言之，外在效度问题可以看成是考察因果效应的分布，是否具有异质性实验处理效应。如上文所述，这其实并非实验研究独有的问题，观察性研究也并不比实验研究更具优势。

除了考察因果效应，社科研究要实现外推还依赖于因果机制，因为因果机制可以帮助研究者了解研究对象或研究情境中，有哪些特征可能影响因果关系（Deaton: 177 - 178）。尽管这是用来批评实验研究只能得出“是否产生影响”而不能得出“为何产生影响”，但同样的逻辑也可用来批评观察性研究。例如，大样本的定量研究也不能直接验证因果机制。反观案例比较，尽管这种方法容易揭示因果机制，但前提是该因果机制所依赖的一连串因果效应都成立，而支持它们成立的因果机制又难以从该项研究中得出（刘骥等，2011）。总之，从因果机制来看，观察性研究也并不比实验研究更具有外在效度。

观察性的研究者批评实验的外在效度，其背后一个潜在的假设是：观察性研究的外在效度本质上就高于实验研究。但这其实是一个误解，前文已经将其澄清。那么，接下来的问题是，要提高外在效度，实验研究更有效还是观察性研究更有效？首先，要考察因果效应在不同样本、不同情境中是否都成立，这就需要重复研究，而实验更容易复制。此外，要考察因果机制，实验并不比观察性研究更差。定量研究检验机制不容易，因为需要搜集更多的“干净”数据来逐一识别因果机制；案例比较考察的因果机制都只是“说理”层次，缺乏经验的系统支持；相比之下，实验研究比较容易通过操纵因果机制链条上的那些变量来加以识别，而一旦弄清楚原因通过何种渠道来产生影响，就更容易判断一项研究外推的范围如何（Imai et al., 2011）。

综合上述，从因果推论的内在效度与外在效度来看，实验研究都比观察性研究更有优势：在解决未知干扰与异质性实验处理效应问题时，实验的估计偏

---

<sup>①</sup> 匹配方法选出来的样本，并不能代表总体。这一点被班纳吉和迪弗洛（Banerjee & Duflo: 95 - 96）用来论证“观察性研究也面临内、外在效度间的权衡”。今井等（Imai et al. 216 - 224）也着重强调了匹配方法的这一缺点。

差更小，就算实验过程并不完美，其理论上也能比观察性研究掌握更多的信息，来帮助矫正偏误；在将研究结果进行外推时，观察性研究并没有固有的优势，相反，实验凭借易控制与易复制，能更好地帮助研究者通过重复研究来检验多重因果关系、探索因果机制。这一分析将全书的辩论往前推进了一步：在辩论两种方法的长短时，需要将其置于同一情境、同一问题下考虑，才有助于更全面地理解实验研究与观察性研究分别如何应对。

#### 四、结束语：两种方法的交流与结合

众所周知，社会科学的因果推论深受“鲁宾因果模型”（Rubin Causal Model）的影响，由此得出因果推论的根本问题则是：无法观察到研究对象同时处于“受到影响”与“未受影响”两种状态（Imbens & Rubin, 2015: 3 - 22）。对此，实验的随机分配可以解决，于是实验也几乎成为所有研究设计的起点，成为因果推论的黄金准则。为了反思这一传统，蒂尔发起了这场辩论，希望学界看清楚实验方法在未知干扰与异质性实验处理效应、以及不完美的实验过程等方面的局限。毫无疑问，这些批评都非常关键，也促使实验学者放慢了脚步，慎重反思这些问题可能带来的麻烦，并探索解决之道。然而，这并不代表批评者赢得了这场辩论。根据本文的分析，实验方法所面临的问题，观察性研究同样不能避免，甚至更难处理。无论从内在效度或者外在效度，实验都可能因更强控制、更多信息而有因果推论的相对优势。

到此，这场辩论远未结束。全书对于“实验是否为因果推论的黄金准则”这一问题争辩不止，对于“实验与观察，孰优孰劣”尚无最终定论。除了从上述方法论的视角来比较“实验与观察的高下”之外，还需要从研究议题、学术发展等视角来检视二者，这也是本书稍有涉及、但未深究的部分。首先，“实验与观察，孰优孰劣”或许没有一个放诸四海皆准的答案，而是取决于研究问题与研究目标，那么何时该观察、何时该实验？其次，另一种可能的情况是，对于同一个研究问题，两种方法可以形成互补：观察法可用于构建宏观、抽象的因果假说、初步考察因果关系，而实验在进一步检验微观的因果效应与因果机制具有优势。最后，随着信息技术的发展，我们对于未知（干扰）的认知能力增加，也能搜集到所有资料进行“整体”的研究，到那时，观察与实验的对立关系，是否还将继续存在？

可见，“实验”与“观察”的分歧还将长期存在，要尝试沟通两类研究，任重而道远。但是，雄辩研究方法的长短毕竟不是社科研究的终极目标，回答因

果问题才是（王天夫，2006）。从这一点来看，尝试沟通两种研究方法也非常重要，因为这关系着研究者如何选择合适的方法来建立理论、检验理论。对此，夏皮罗（Ian Shapiro）以及今井等都建议采取务实的态度，以问题为导向、理论为目标，能够回答问题的方法就是适用的方法（Shapiro：238 - 239；Imai et al.：214）。正如三十年前政治学界掀起了一场“定性 vs. 定量”之争一样，“观察 vs. 实验”之争已经成为社科研究方法论辩论的新焦点。或许，学界也需要一本沟通实验研究与观察性研究的 KKV，帮助指导研究者如何做出恰当的研究设计。<sup>①</sup>沟通的起点在于相互了解，本书中关于实验方法、实地实验的辩论历程，将能帮助研究者全面了解实验研究与观察研究各自的优劣，找到最适合的研究方法。

### 参考文献

- 耿曙、陈玮(2013). 比较政治的案例研究：反思几项方法论上的迷思. 社会科学, 5: 20 - 29.
- 刘骥、张玲、陈子恪(2011). 社会科学为什么要找因果机制——一种打开黑箱、强调能动的方法论尝试. 公共行政评论, 4: 50 - 84.
- 罗俊、汪丁丁、叶航、陈叶烽(2015). 走向真实世界的实验经济学——田野实验研究综述. 经济学(季刊), 14(3): 853 - 884.
- 马得勇、兰晓航(2016). 精英框架对大学生有影响吗——以实验为基础的实证分析. 清华大学学报(哲学社会科学版), 3: 160 - 171.
- 孟天广、杨平、苏政(2015). 转型中国的公民意见与地方财政决策——基于对地方政府的调查实验, 公共管理学报, 3: 57 - 68.
- 王天夫(2006). 社会研究中的因果分析. 社会学研究, 4: 132 - 156.
- 谢宇(2012). 社会学方法与定量研究(第二版). 北京: 社会科学文献出版社.
- 余莎、耿曙、孔晏(2015). 如何有效征税：来自纳税遵从实验的启发. 公共行政评论, 3: 151 - 175.
- Angrist, J. & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

<sup>①</sup> 在社科领域，KKV 是 Gary King、Robert Keohane、Sidney Verba（1994）所著 *Designing Social Inquiry: Scientific Inference in Qualitative Research* 一书的代名词。该书以沟通定性与量量的分歧为目的，将定性研究与定量研究整合到统一的因果推论框架中，它一经面世就受到广泛关注，并持续产生影响，成为社科研究设计的“圣经”。有关该书以及研究方法的“定性 vs. 定量”之争的讨论，详见《公共行政评论》于 2015 年第 4 期耿曙组稿的研究专栏，以及耿曙、陈玮（2013）。

- Druckman, J. , Green, D. , Kuklinski, J. & Lupia, A. (2011). Experiments: An Introduction to Core Concepts. In Druckman, J. , Green, D. , Kuklinski, J. & Lupia, A. Eds. *Cambridge Handbook of Experimental Political Science*. Cambridge & New York: Cambridge University Press.
- Druckman, J. & Kam, C. (2011). Students as Experimental Participations: A Defense of the “Narrow Data Base”. In Druckman, J. , Green, D. , Kuklinski, J. & Lupia, A. Eds. *Cambridge Handbook of Experimental Political Science*. Cambridge & New York: Cambridge University Press.
- Falk, A. & Heckman, J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science*, 326(5952):535 – 538.
- Gerber, A. & Green, D. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York & London: W. W. Norton & Company.
- Gerring, J. & McDermott, R. (2007). An Experimental Template for Case Study Research. *American Journal of Political Science*, 51(3):688 – 701.
- Harrison, G. & List, J. (2004). Field Experiments. *Journal of Economic Literature*, 42(4):1009 – 1055.
- Holland, P. (1986). Statistical and Causal Inference. *Journal of the American Statistical Association*, 81(396):945 – 960.
- Imai, K. , Keele, L. , Tingley, D. & Yamamoto, T. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanism from Experimental and Observational Studies. *American Political Science Review*, 105(4):765 – 789.
- Imbens, G. W. & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Science: An Introduction*. New York: Cambridge University Press.
- King, G. , Keohane, R. & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Lijphart, A. (1971). Comparative Politics and the Comparative Method. *American Political Science Review*, 65:682 – 693.
- McDermott, R. (2011). Internal and External Validity. In Druckman, J. , Green, D. , Kuklinski, J. & Lupia, A. Eds. *Cambridge Handbook of Experimental Political Science*. Cambridge & New York: Cambridge University Press.
- Morton, R. B. & Williams, K. C. (2010). *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge University Press.
- Przeworski, A. (2007). Is the Science of Comparative Politics Possible?. In Boix, C. & Stokes, S. Eds. *The Oxford Handbook of Comparative Politics*. New York: Oxford University Press.
- Smith, V. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review*, 72(5):923 – 955.
- Teele, D. (2014). *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven & London: Yale University Press.