



中国教育科研参考

2020年第17期
总第(483)期

中国高等教育学会编

2020年09月15日

目 录

- 关于教育中的实证与经验研究..... 阎光才 (02)
- 扎根理论在中国教育研究中的运用探索..... 陈向明 (05)
- 中国教育实证研究中的定量方法: 五年应用述评..... 吕 晶 (13)
- 在教育研究中运用量化研究方法的问题与反思..... 蔡红红 (29)

编者的话: 教育是与实践结合最紧密、与每个人都息息相关的事业。我国举办着世界规模最大的教育, 各级各类学校组织的复杂化、结构的多样化、水平的差异化以及人民群众教育诉求的个性化都在不断增强。回溯教育科学的发展历程, 其每一次重大突破, 都离不开研究方法的发展。“工欲善其事, 必先利其器”, 要提高教育科研水平就必须掌握合适的实证研究方法。目前学界将定量研究、质性研究和混合研究列为实证研究的三种主要形式, 强调教育实证研究是基于证据解答教育问题的一种研究范式。实证研究是一个不断丰富、不断创新的方法链条, 实验研究、准实验研究、大数据分析的方法体系都属于实证研究的范畴。本刊以“实证研究方法”为选题, 集中选编若干文章, 供读者参阅。

主编: 王小梅 本期执行主编: 王者鹤 责任编辑: 李 璐
地址: 北京市海淀区学院路35号世宁大厦二层中国高等教育学会《中国高教研究》编辑部
邮编: 100191 电话: (010) 82289239
电子信箱: gaoyanbianjibu@163.com
网址: www.cahe.edu.cn (中国高等教育学会——学术动态栏目)

关于教育中的实证与经验研究

阎光才

自20世纪90年代以来，在教育研究领域，围绕量化与质性的方法与范式之争一直是热点议题。而最近一段时期以来，随着大数据热的持续升温，在“数据如金矿和宝藏”“谁掌握了数据谁就掌控了世界”等极尽夸饰的氛围中，重实证或重证据研究、推动教育研究范式全面转型的说法又开始广为盛行。然而，仔细辨析日常中人们所不假思索甚至已经习惯了的“实证”说法，不难发现其中歧义纷沓，甚至不乏一些误区。在此，对于现实中各种说法是否严谨不敢妄议，而是仅仅根据自己研究体会尝试展开一些浅显的分析，其目的并非以正视听，恰恰相反，希望在目前氛围中能够倡导一种方法意义上的多元宽容，如此或可更利于我们教育研究领域的异彩纷呈。

一、究竟如何理解实证研究

在西方学术界，真正标榜自己所从事的是实证研究（positive research）的学者很少。为何如此，这要从实证研究的源头实证主义（positivism）谈起。实证主义最早的提出者是孔德，在19世纪自然科学已经成为知识和文化领域主宰的氛围下，孔德提出了人类的智力与精神进入他所谓超越传统神学、形而上的实证阶段。严格而言，孔德的实证论是一种关于社会知识的哲学主张，而不是方法。他认为对社会研究只有追随自然科学的规范与路径，以对人类感官可感受的现象观察为出发点，才能获得确定性、具有预见性的知识。继孔德之后，奥地利的维也纳学派基于逻辑实证主义立场，提出了更为激进的一系列主张：所有合法的知识都源于经验，科学是唯一正确的方法，辨别知识是否有意义就在于它是否具有经验意义上的可证实性。但凡不能通过经验证实的如终极的精神和价值，统统都被视为形而上学而应该予以拒斥。无论是孔德的实证主义还是逻辑实证主义，从立场角度而言，他们的共同点都在于排斥传统形而上层面的思辨研究，而逻辑实证主义则不独于此，在强调基于经验事实、把一切知识都还原为统一的物理科学的同时，也把社会科学研究中始终无法撇除的价值乃至价值事实逐出了合法性知识范畴，如此激进主

张难免遭人诟病。由是，在现实之中，西方极少有人甘愿冒领这个多少有些极端的标签，把自己的研究归为“实证”，而更趋向于认同一种“经验研究”（empirical research）的说法。当然，即使有人愿意认领，要做到也委实不易。因为如果循名责实，在社会科学领域恐怕除了少数如行为主义研究以外，几乎大多研究都难以符合这种近乎苛刻的条件与要求。也许是由于中西语境差异，我国学术界称自己为经验研究者少之又少，这大概就与质性研究者不愿用“定性”说法的情形相似，因为一不小心就会被误解为工作总结性质的浅表研究。故而，我们学术界中素常意义上所言的实证，其实大都不严格，甚至其规范程度未必能及西方的经验研究。那么究竟实证研究与经验研究之间是一种什么关系？在我看来，两者的共同之处都在于强调所有研究都起于经验，所不同之处在于对研究的预期和对经验的理解存在差异。严格的实证研究所推崇的是自然科学研究路向，它强调主体与客体间严格的二分，即研究者与研究对象间的对象化关系，通过基于远距离观察（即价值与情感无涉的旁观者），以综合归纳的方式寻求共性特征，即现象意义上的共相或“是”。在此基础之上，通过建立理论假设，并演绎推断出不同变量间可能存在的因果关联，以外在客观的经验事实来验证假设，进而揭示共相背后的机理也就是回答“何以为是”的问题。实证研究的最终目的是形成一种通则即普遍意义上的理论解释。如同自然科学研究结论一样，它具有确定性、可检验性、可重复性和可预见性。研究过程与结论具有脱域性而不是局限于特定的情境，因而结论不仅不受情境的限制，而且也与研究者的立场、情感和价值偏好无关。

二、关于经验研究一解

由此，不难发现，实证研究其实是一种非常苛刻的研究取向，它更多代表一种理想化的研究范式诉求和立场主张。而在现实之中，鉴于研究者本身以及研究对象的错综复杂性，这种如同以物理学或生物学来研究社会的方式注定是臆想。即使存在某些共识性结论，它充其量也就具有统

计学意义而不具有图像论一一对应的确定性。正因为如此，即便是秉持实证立场的人们也往往尽量回避实证说法，而给予更为弹性的提法——经验研究。因此，经验研究实际上与实证研究并非对立，而是存在一种包容关系，即实证研究其实只是经验研究谱系中的一种极端取向。经验研究的对立面是纯粹思辨，与实证研究一样，它主张研究始于经验，但是，在此所谓的经验未必一定是外在的可观察到的事件和事实。在我看来，诸如人们的主观感受、直觉、个体阅历、集体历史与文化体验、直接与间接经验等等，都不妨可以纳入经验的范畴。相较于排斥性的实证研究，经验研究更具有包容性和开放性。它可能也追求自然科学研究所推崇的理论简洁和精致、有指向通则的偏好，但也不拒绝基于各种各类外在与内在经验解读所形成的片段性认识，甚至它可不以建构理论或提供理论解释为目的。在现实中，众多非理论取向的经验研究往往是严格实证研究过程不可缺少的环节，甚至是严格实证研究问题与理论生成之源。除此之外，诸如带有诠释取向的民族志、口述史和叙事研究，偏重于编码、归纳和理论抽样的扎根理论研究，注重现实问题解决的行动研究，关注效果比对的实验、准实验研究与循证研究，如此等等，也大致上可以把它们纳入广义上的经验研究范畴。后文将对此做进一步展开论述。

三、关于实证取向经验研究理解所存在的误区

鉴于社会科学研究对象（人、事）的特殊性，以自然科学研究范式来苛责和规范社会科学研究不过是种奢望。但是，无论在哪些领域，寻求确定性、稳定性与秩序性，以此来驾驭和控制周遭世界与环境变幻莫测的变化，这似乎是人类与生俱来的求知偏好。实证取向的经验研究便多少体现了人类所具有的这种堂吉诃德式骑士精神。在此，尝试对该取向研究在现实中可能存在的一些误区略加以辨析。

第一，实证取向的经验研究并不排斥理论思辨，相反，它是做好研究所必备的素养。实证取向的经验研究，旨在通过经验验证的路径揭示现象背后的因果关联机制。这种因果关联并非源自人的直观经验观察，而是研究者合乎理性与逻辑的主观推断。故而，所有实证取向经验研究的理

论假设（变量关系）形成过程本身，就是一个理论思辨的过程，也是一个由概念抽象化到具体可操作化的演绎过程。

第二，实证取向经验研究的“中立”与“客观”，大多时候不过是一种迷思。通常而言，实证取向经验研究始于问题，求证过程来自观察，但研究问题的选择与观察过程却往往与个人的价值、信念、信仰与立场判断存在关联。即使在研究过程中，例如测量工具的设计也往往难逃价值污染。调查过程中问项如何表述，采用负面还是正面语义，应答者对题项的理解与当下心境与心态，都难免影响结论的客观性。

第三，实证取向的经验研究未必一定是定量研究。可测量是实证取向的偏好所在，但这并不意味着但凡定量就是实证，各种定量的描述性研究、趋势分析以及目前所流行的大数据分析，所关注的多为对现象的特征描述以及不同类项间可能存在的相关性，这些研究的目的并不在于建立因果关联的理论解释；同样，实证取向也未必一定是定量研究，基于田野观察与实验设计，以定性证据对有关现象内因展开推理并形成合乎逻辑的理论解释，也可归为实证取向的经验研究。

第四，目前广为盛行的循证研究，也未必是实证取向的研究。循证的确关注经验证据，但是这里的证据更倾向于验证干预的效果而不是对内在机理的求证。前者意味着同样的效果可以通过不同手段来实现，而后者则执着于各种表象背后的统一归因。现实之中，很多自称为实证性的实验或准实验研究，如果没有关联到内在机理与机制的探究，是循证而非实证取向的研究。但是，循证有益于实证取向的研究，可以为后者的内在机理探究提供索引与启发。

第五，实证取向的经验研究与诠释取向的质性研究并非存在对立关系。传统上，实证与质性取向间的对立主要表现为本体论与认识论意义上的分歧，在经过长期的双峰并峙、二水分流状态之后，如今人们越来越趋向于认同在方法意义上各有取舍的共融。实证取向不再回避其所关注的事实本身的人为建构性，它在强调对人的外显行为以及集体事件观察的同时，也赋予其关于人的动机、意图、情感和价值乃至生命意义的理解，因此，诠释自然也是实证取向无可回避的选择；同样，质性研究虽青睐特定情境中一切都可视为

“文本”的深描和意义解读，无心外在效度，但它也未必排斥实证工具如定量方法的使用。甚至至于，它也不局限于研究的外在效度，而接纳不同情境差异中的研究结论的“家族相似性”。总之，至少在方法取向上，诠释也罢解释也好，并非水与火的关系，而是存在互补、互证甚至融合的可能性。

四、倡导多种方法混合与共融的经验研究

为与传统的思辨和经验研究（即有关工作总结与体会性研究）划清界限，也是为了与西方研究话语以及范式接轨，近年来教育研究中的方法以及方法论热持续升温，方法似乎已经成为提升研究水平的不二法门。因为历来缺乏实证与量化传统，20世纪90年代后我们在根本不同轨的时段，却将错就错地搭上了西方学术界关于质性与量化、现代主义与后现代主义之争的末班车。于是，在量化研究不过还是一个稻草人的语境中，质性研究在我国也一度成为告别传统的显学。进入20世纪之后，在西方后现代主义式微、带有实用取向的循证研究开始抬头，尤其大数据分析风行的环境中，我们方才发现，西方学术界中始终占据主流的实证取向（特别是量化取向）经验研究历来就是我们的短板。由是，倡导教育的实证研究又成为一种风尚。不同取向不同风尚纷纷亮相，你方唱罢我登场，有意无意之中难免相互贬抑，都希望表演独角戏。针对这种乱局与倾向，结合上述分析，在此尝试提出如下认识。

首先，倡导教育研究要面向现实与实践，开展广义上的经验研究。如此，我们不会再纠结于严格实证与非实证、定量与定性、宏大叙述与日常生活叙事、共时性与历时性、普遍性与特殊性等各种取向间的内在紧张，而是以开放包容的态度赋予各种“证据”及其内涵理解以合法性，认清不同方法的合理性与局限性。

其次，教育研究需要重视方法，但是，方法与逻辑一样是作为研究者的素养，而不是探究未知的万能钥匙。寄望于以方法及其规范，对日常教育现象甚至其中的一个片段的揭示或解读达到芝麻开门之功效，其实是方法（工具）至上主义的妄念。须知：在方法与问题之间，永远是问题为王。而问题的生成、分解以及假设形成和演绎推断过程，却需要研究者具备必要的敏锐的洞察力和扎实的理论素养。故而，以采用何种方法、

取用和挖掘哪些素材，如何去审慎地求证，求证结果是否因为方法偏颇而存在局限性，这更多取决于研究者对现实的洞察力、问题意识与理论准备，方法本身不是做好研究的独门秘笈。

第三，方法有如十八般兵器，但拥有兵器却未必代表能够掌握十八般武艺。研究过程中究竟是采用定量还是定性，使用什么工具，取用什么材料作为证据，是各种数据、影像、田野观察记录、口述记录还是历史档案等等，这都取决于研究者基于研究问题和情境的心领神会，以及对不同方法的灵活取用。因此，围绕问题的研究并非是一种方法简单的生搬硬套，而是根据需要进行不同方法的混合与共融。

第四，研究需要逻辑但更讲求证据，即在逻辑与事实之间，永远是事实为王。然而，如何理解事实？不同于自然科学，社会科学研究所面对的现象与观察的事实，其客观表象背后的内在机理却往往带有主观性。在教育研究中，人的外在行为表现是客观的，但它往往源于不同的主观意图与动机。教育研究归根结底是人及人的境遇的研究，抛开人所共同置身的场景及其历史与文化，研究者是否具备一种现实、本土和价值的关怀姑且不论，其所谓的客观结论究竟有多大的可信度，理论解释是否有深度，都很值得商榷。

总之，在如今一个越来越讲求方法特别是关注实证取向的时代，我们有必要清醒地认识到：方法本身不应是置于人的洞察力与创造力之上的樊篱与窠臼，而恰恰是被人们富有智慧与想象力掌控的策略与艺术。执着于本体论与认识论上的范式和立场之争，拘泥于刻板程序与规范工具主义取向，排斥各种方法间的互补性，并不会使得研究成果更可靠，更遑论它本应具有的生动、品味和厚重。这正如米尔斯在《社会学的想象力》中所言：“一流的社会分析学家会避免僵化的程序，尽力发挥并运用社会学的想象力，想象力非常关键，不受方法与技巧的制约……，对一流的社会学家来说，方法与理论都不是独立的王国，方法是针对一定问题的方法，理论是针对一定现象的理论。”诚哉斯言！

（阎光才，华东师范大学高等教育研究所教授，上海 200062）

（原文刊载于《中国高教研究》2016年第1期）

扎根理论在中国教育研究中的运用探索

陈向明

扎根理论 (grounded theory) 是由两位美国学者格拉斯 (Barney Glaser) 和施特劳斯 (Anselm Strauss) 在1967年出版的合著《扎根理论的发现》(The Discovery of Grounded Theory) 中首次提出的。此后, 在近半个世纪的发展过程中, 两位学者以及他们的学生对这种研究路径 (approach) 做了一些修改。然而, 这种研究路径到底有什么独特之处? 是在什么背景下提出的? 具体如何操作? 能够为我国当前的教育研究提供什么洞见? 在我国独特的学术环境中面临什么挑战, 需要什么本土化调适? 这是本研究试图回答的问题。

之所以探讨上述问题, 主要出于如下两方面的考虑。其一, 这种研究路径在社会科学研究界 (包括教育研究界) 具有很大的影响力。在质性研究的各种路径中, 从历史发展的角度看, 除了民族志之外, 扎根理论是最早且最具影响力的研究路径。其他路径 (如现象学、传记研究、叙事探究、话语分析等) 都是晚近兴起的, 而且 (除了现象学) 都可以将扎根理论作为一种分析资料的方法。其二, 我国的一些教育学者 (比如我) 在使用扎根理论路径开展研究以及教授研究方法课程时, 发现存在不少局限性。例如, 扎根理论的主旨是生成 (generate) 理论, 因此完全不讨论提出研究问题和设计概念框架, 也很少讨论收集资料的具体技术, 更不提及研究者身份、研究关系及其对研究的影响等伦理道德问题。而这些问题是我国教育研究界和研究方法课程非常重视的内容, 具有重要的社会文化意涵。因此, 需要针对我国教育研究的现状和要求, 对扎根理论的适切性进行探索。

本研究首先介绍扎根理论的定义、提出的背景、基本操作步骤和方法; 然后, 通过一个课堂教学实例, 管窥扎根理论研究运用于我国教育研究的一种可能样态; 最后, 针对这个教学案例, 探讨扎根理论对于当前我国教育研究的意义、限度以及本土化创新的可能性。

一、什么是扎根理论

扎根理论是一种研究路径, 而不是一种实体的“理论”。其要义可以被总结为: 研究的目的是生成理论, 而理论必须来自经验资料 (empirical data); 研究是一个针对现象系统地收集和分析资料, 从资料中发现、发展和检验理论的过程; 研究结果是对现实的理论呈现; 通过系统的资料收集和分析程序而被发现的理论被称为扎根理论。扎根理论研究者喜欢分析胜过描述, 喜欢新鲜的概念类属 (category, 即在一个更抽象层次上组合起来的概念群) 胜过预先设定的观点, 喜欢系统聚焦的、连续收集的资料胜过大量同时收集的资料。

(一) 扎根理论提出的背景

扎根理论的提出, 源于两位创始人对当时美国社会学界三种倾向的批评。其一, 大部分学者照搬学术大师的“宏大理论” (grand theory), 使用逻辑演绎的方法, 通过经验研究对其进行验证或局部修正, 结果导致“理论资本主义”的形成 (即极少数学术大师垄断了理论的生产, 而大部分学者沦为验证学术理论的“无产阶级”)。其二, 绝大部分民族志研究局限于对现象进行“印象式”描述, 缺乏足够的理论提升, 导致研究结果稀松、平淡, 难以与学术界其他理论对话。其三, 一些研究虽然提出了理论, 但只是粗略地对经验资料进行理论分析, 没有介绍理论生成的方法、步骤和过程。因此, 扎根理论路径希望在理论与资料的鸿沟之间架起桥梁, 在经验资料、已有文献和研究者个人经验的基础上, 通过系统地收集资料和分析资料生成扎根理论。

为了将理论与经验事实联系起来, 扎根理论路径只生成“中层理论”, 即位于有关日常生活的较小的“工作假设” (working hypothesis) 与无所不包的“宏大理论”之间的理论。扎根理论路径生成的中层理论包括两类: “实质理论”与“形式理论”。前者涉及一个实质领域, 即对某一特定情境下某一现象的研究所获得的理论,

如病人护理、种族关系、专业教育、少年犯罪、研究机构等。后者跨实质领域，即对许多不同类型的情境下某一现象的研究所获得的理论，如污名化、越轨行为、正式组织、社会化、权威和权力、奖励制度、社会变迁等。在格拉斯看来，扎根理论路径归根结底要做的是：呈现行动者在处理某个问题时的行为变异（variation），找到各种行为模式（pattern），并将这些模式用理论的形式表达出来。

（二）扎根理论的实操要求和步骤

扎根理论研究的实践有如下要求：①资料收集、资料分析和理论生成同时进行；②从经验资料而不是从预想的、逻辑演绎的假设中形成代码（code）和类属；③保持理论敏感性，对事件与概念以及概念与概念的关系不断提问、比较；④为了理论生成的目的进行抽样，抽样的单位是概念，而不是人、地点或事件，不必按照人口学变量进行抽样；⑤类属需要达到理论性饱和，即属性基本齐全，再收集资料会出现重复；⑥研究伊始尽量不要带有预设，在形成独立的结论之前，不要阅读本实质领域的文献；⑦通过编码、写备忘录和画图表来完善类属和类属之间的关系，形成初步假设，最后进行理论整合。

扎根理论中的“编码”指的是，通过将事件与事件、事件与概念、概念与概念之间进行连续比较，对资料进行概念化，以形成类属及其属性（property，即类属的性质和特点）。由于研究旨趣不同，格拉斯与施特劳斯的编码方式有所不同：前者只有两级，即开放编码和选择编码；后者有三级，即开放编码、轴心编码、选择编码。施特劳斯的质性研究背景使他以资料分析过程为重，他认为诠释资料本身才是研究的主要目标。而格拉斯认为，施特劳斯的分析过于破碎、散漫，而且预设了编码类型。定量研究的背景使他致力于发展形式理论，追求理论表述的简洁、应用范围的广泛。

在开放编码中，研究者需要悬置预设，凭着理论敏感性开放地挑选资料，将资料掰开、检视、比较、概念化、类属化。代码可以是被研究者使用的鲜活的本土概念（in vivo code），也可以是能够表达资料内容的学术概念，即可以结合文化主位与文化客位的视角。施特劳斯还提出

了一些比较资料的方法，如丢铜板（flip-flop），即想象一个概念的极端反例以刺激思考；极远比较（far-out comparison），即找一个与本领域完全不相关的领域进行比较，然后再回到资料获得新的感悟；摇红旗（waving the red flag），即始终对字词和片语保持怀疑，进行深入检视。

施特劳斯的轴心编码是在开放编码的基础上，形成类属、属性和维度（dimension，即属性在一个连续统，如强度、频率上的定位），发展并检验各类属之间的关系。在此阶段，施特劳斯还提出了一个编码范式模式（coding paradigm model），将不同类属按照事情发展的通常顺序联结起来：（A）因果关系——（B）现象——（C）情境——（D）中介条件——（E）行动/互动策略——（F）结果。

选择编码即在所有类属中选择一个核心类属（core category），其他类属则成为支援类属（subsidiary category），然后通过一个整合图式（integrating scheme）或故事线（story line），将各种理论要素（类属、属性、假设）整合起来。核心类属指的是，能够将所有其他类属整合入内的中心概念，在资料中频繁出现，很容易与其他类属相联系，能够包容最大限度的变异，而且有发展为扎根理论的潜能。

除三级编码，施特劳斯还提出了一个“条件/结果矩阵”，意在区分并连结条件和结果的各种层次，如个体行动、人际互动、组织、社区、国家、国际等。这些层次表现为一个同心圆，个体行动的原始资料被概念化之后，被置于一层层相互影响的行动系统中。如此，分析就不只是停留在微观层次，而且兼顾了各类中观和宏观条件。

由于学界对扎根理论研究提出了过于技术化、过于“后实证主义”的批评，1996年施特劳斯去世之后，科尔宾对她与施特劳斯合著的《质性研究的基础》进行了三次修订。受到建构主义和后现代思潮的影响，她在2008年出版的第三版中，除了在章节开篇时保留了三级编码的概念定义，正文中不再涉及具体编码的程序和技术，取而代之的是一系列灵活、多样的备忘录和概念图。这些备忘录和概念图呈现了作者从事的一项对美国越战老兵的研究过程，包括形成概念、将

情境和过程纳入分析、整合理论要素等。这不禁让我们思考：三级编码是必需的吗？在我看来，分析的上限是无止境的，重要的是研究是否在资料的基础上提出了新颖的理论，理论是否能够有效地解释研究现象。

二、扎根理论运用于中国教育研究的实例

根据上述对扎根理论的理解，下文提供一个将其运用到中国教育研究的实例。在我开设的“教育质性研究方法”课上，每位同学都被要求独立开展一项小研究。在1999年的课上，有一位同学提出的研究问题是：人到中年为何还来北大读研。他对数位被研究者进行了连续访谈，其中一位受访者说的一段话引起了大家的注意。访谈者当时问受访者：“你到北大来学习，家里人是如何想的？”表1中左边第一列是受访者所说的那段话。现在，我使用扎根理论的方法对这段话进行一个“微分析”，以此示例扎根理论运用到我国教育研究领域有可能是一种什么样态。

上述资料具有中国社会文化急剧变迁的时代特征。首先，1998年我国高等教育大规模扩招后，地方高校对教师的学历有更加明确的要求，很多教师外出读研以提高学历水平。其次，与之前相比，中国社会变迁加剧，更多人有机会离开家乡到外地求学和工作，安土重迁的文化传统被打破；但与此同时，男主外女主内的文化习俗仍旧存在，在受过高等教育的群体中，外出学习者更多的是男性。

（一）分析过程

首先，我们根据格拉斯的建议，对资料进行一个整体性的提问：这个资料是一个关于什么的研究？回答是：这段资料揭示的是个人事业发展与家庭稳定之间的冲突关系，处于人际互动的层次；但由于个体及其家庭都处于中国社会文化变迁的大背景下，个体内心的挣扎也反映了时代的胁迫和文化传统的浸染。

然后，我们使用施特劳斯的三级编码方式对资料进行分析，因为它比格拉斯的两级编码更加细致，层次更加清楚。我的学生都是初学者，将编码步骤和过程尽可能清晰地呈现出来，对他们更有帮助。在开放编码中，我们尽量悬置前设，保持思维开放。在对原始资料进行初步概念化后，形成了如下代码（见表1第二列）：家庭负担

重，爱人工作忙，小孩小，老人需要照顾，家庭责任分担变化，全部靠妻子承担，妻子用哭表达歉疚和担心，丈夫产生歉疚感，主观与客观统一，家国同构，夫妻间关系伦理，个体行动时空感，文化传统影响等。

在轴心编码阶段，我们在上述代码的基础上，形成了4个类属及其属性和维度（见表1第三列）。为了便于理解，下面每个类属后面提供了一段文字说明。

类属	属性	维度
1) 家庭负担变化	妻子工作 孩子 老人 责任分担变化	忙—不忙 小—大 需要照顾—不需要照顾 部分靠她—完全靠她

在丈夫出来之前，家务由两个人分担。现在丈夫外出求学，负担全部落到妻子一个人身上。妻子是职业妇女，而且工作很忙。加上孩子很小，需要照顾。家中还有两代老人，都住在家中，老人数量多，而且身体不好，需要照顾。

2) 妻子做出反应	担心 歉疚 表达方式	大—小 高—低 被动—主动
-----------	------------------	---------------------

妻子的反应可以分成内容和形式两个方面。内容包括担心和歉疚，两者的程度都很高。担心的是丈夫外出求学，学成后地位高升，有可能将自己甩了。但如果不让丈夫出去求学，自己又感觉歉疚。然而，妻子表达担心和歉疚的方式都是被动的，用眼泪和身体辛劳来获得丈夫的同情。

3) 丈夫做出反应	歉疚 承诺 动机	低—高 高调—低调 主客结合—主客分离 家国同构—家国分离
-----------	----------------	--

丈夫的反应也包括内容和形式两个方面。内容包括歉疚，虽然他使用了“对不起”等字眼，但歉疚感似乎比较低；反应的形式包括承诺，表现出高调的姿态，引用了不少中国文化格言。此外，丈夫陈述了自己外出读研的动机，表达了主客结合和家国同构的维度。

4) 夫妻间关系伦理	社会地位 家庭责任 回报时间 性别差异 文化传统	平衡—不平衡 分担—不分担 现在—未来 口惠—眼泪 变心—坚守
------------	--------------------------------------	---

丈夫外出读研之前，双方的社会地位是平衡

表1 三级编码示例

原始资料	一级：开放编码	二级：轴心编码	三级：选择编码
<p>我爱人是从事商业服务行业的，当会计，天天很忙，孩子上幼儿园，她得天天接送。我在家时，有我接送；我出来，接送孩子就得她去。她父母身体不太好，有时还得照顾老人。我祖母半身不遂，所以，我父母还得照顾老太太，帮不了我们家的忙。这样，负担全靠她支撑了。我出来时，她也哭过。不让我出来，她觉得她对不起我；可出来了，我又觉得对不起她。她还怕我学成了，变心甩了她；我哪能呀！糟糠之妻不能忘，就凭她支持我出来这一点，以后就得好好营造这个家。我不爱说大话，为社会也好，为国家也好，那都是客观的；主观上主要还是想营造好这个家。没有家，哪有国呀！想想她为我付出这么多，这辈子也得好好爱她。俗话说，“海枯石烂，永不变心”。</p>	<p>家庭负担重 爱人工作忙 小孩小 老人需要照顾 家庭责任分担变化 全部靠妻子承担 妻子用哭表达歉疚和担心 丈夫产生歉疚感 主观与客观统一 家国同构 夫妻关系伦理 个体行动时空感 文化传统影响</p>	<p>类属 家庭负担变化 爱人工作 孩子 老人 责任分担变化 妻子做出反应 歉疚 表达方式 丈夫做出反应 歉疚 承诺 表达动机 夫妻关系伦理</p> <p>属性 忙—不忙 小—大 需要照顾—不需要照顾 部分靠她—全部靠她 高—低 高—低 被动—主动 低—高 高调—低调 主客结合—主客分离 家国同构—家国分离 平衡—不平衡 分担—不分担 现在—未来 口惠—眼泪 变心—坚守</p> <p>维度</p>	<p>核心类属：夫妻关系伦理 支援类属：1. 妻子做出反应； 2. 丈夫做出反应 情境条件：家庭负担变化 因果条件：丈夫外出读研 初步假设： 1. 当双职工家庭上有老下有 小、一方长期离家时，会造成夫妻关系紧张。 2. 平衡婚姻稳定的基础有三个：(1) 双方的情感状况； (2) 社会地位的平衡；(3) 家庭责任的分担。当(2)和(3)发生变化时，(1)也会发生变化。 3. 丈夫离家，妻子会比较担忧；而反过来并非如此，而且通常妻子出来的机会比较少。 4. 当丈夫现时无暇回报妻子时，通常会使用文化格言作为口惠，高调许诺未来报答；而妻子通常使用“弱者的武器”（眼泪、身体辛劳）来应对。</p>

的（不论他们各处于什么地位），但现在平衡被打破了。家庭责任之前是两人分担的，现在丈夫不分担了。一旦平衡打破，双方就会采取行动，在付出与回报之间重获平衡。夫妻的行动存在性别差异，女方流动的可能性比较小，担心的程度比较大。因此双方恢复平衡的方式也不相同，强势一方（男方）的策略是口惠（将来回报），弱势一方（女方）的策略则是通过眼泪示弱。另外，个体的行动有对时空的预想，表现为三维生活空间中的连续性：过去、现在、未来；互动性：个体、家庭、国家。个体与文化之间也存在互构关系，文化传统有可能约束丈夫在观念上“不变心”。

在上述轴心编码的基础上，我们形成了如下初步假设（见表1第四列）。（1）当一个双职工家庭上有老下有小，一方长期离家时，会造成夫妻关系紧张。（2）平衡婚姻稳定的基础有三个：双方的情感状态、社会地位的平衡、家庭责任的分担。当后两者发生变化时，双方的情感状态也会发生变化。（3）丈夫离家，妻子会比较担忧；而反过来并非如此，而且通常妻子出来的机会比较少。（4）当丈夫（强势一方）无法现时回报妻子时，通常会使用文化格言作为口惠，高调许诺未来报答；而妻子（弱势一方）通常使用低调“弱者的武器”（如眼泪、身体辛劳）来应对。

最后，在选择编码阶段，我们发现“夫妻间关系伦理”这个类属比较重要，能够统括其他类属。不论是“妻子的反应”还是“丈夫的反应”以及双方如何看待“家庭负担变化”这个现象，均与此有关。因此，我们将其作为核心类属，将“妻子做出反应”和“丈夫做出反应”作为支援类属。导致夫妻间行动、互动和情感变化的因果条件是“丈夫外出读研”，而“家庭负担变化”这个类属则成为了情境条件。

通过反复讨论，我们获得了对本段资料的初步结论：当婚姻中一方离家追求事业发展、家庭负担和双方社会地位的稳定状态被打破时，双方会采取不同行动，在情感上重新获得平衡。换言之，当处于社会变迁中的个体重新安排自己的事业定位时，会给家庭稳定带来冲击，导致夫妻各自采取不同行动，重新协商婚姻稳定的基础。虽然这个结论与研究问题“人到中年为何还来北大

读研”关系似乎不够紧密，但通过仔细分析可以看出，受访者其实是在间接回答研究问题：外出读研不仅有提升自己专业地位和水平的诉求，同时也在迂回地为自己的家庭谋求福利。

（二）后续研究思路

由于上面提供的只是一个教学案例，不是一个完整的研究项目，因此没有引用其他访谈资料、观察记录和实物分析，尚未与已有文献形成对话，也还没有根据连续比较的要求进行理论性抽样。如果我们希望针对上述初步结论进一步探究，可以继续询问如下一些问题。

第一类有关家庭的问题：（1）什么情况下会感觉“对得起”或“对不起”对方？感情稳定的基础是什么？（2）什么样的责任分担比较合适？（3）双方的社会地位如何才能平衡？（4）三者之间的关系是什么？

第二类有关个人发展的问题：（1）个体是如何思考自己的发展的？其心目中个人、家、国、社会之间的关系是怎样的？（2）这些关系如何影响到个体对自己个人发展的筹划？（3）个体是如何行动的？行动的条件是什么？（4）行动的结果如何？哪些是意图中的，哪些是非意图的？

第三类有关个人发展与家庭的关系的问题：（1）在被研究者看来，“家”是什么意思？是核心家庭、大家庭，还是国家的一部分？与个人发展是什么关系？（2）个体有关个人发展的行动对家庭有什么影响？（3）个体是如何解释这些影响的？（4）这些影响与哪些历史和社会结构问题相关联？（个体与社会存在互构关系，社会变迁条件成熟度、高校晋升新要求等为个体行动提供了必要性和可能性，但同时个体是否也有选择的可能性？）

上述问题的提出与我们发展理论的需要有关，即我们希望发展出什么样的理论来解释被研究者的行动策略。如果我们将这位丈夫的行动放入一个变异范围内考虑，可以将其命名为“主动适应型”（主动选择离家到中国顶尖大学读研），其他可能的类型有“被动适应型”（在老家附近找一所一般的大学读研）、“主动放弃型”（下海经商）、“被动放弃型”（甘愿做一个低职称的教师）。这四种类型可以形成一个矩阵表，由行为和态度两个维度构成（如表2）。每

表2 面临个人事业与家庭稳定冲突的丈夫的行为类型

态度 \ 行为	行为	
	主动	被动
适应	主动适应型	被动适应型
放弃	主动放弃型	被动放弃型

一种类型都需要采取理论性抽样，进一步收集资料，使每个类属中的属性达到理论性饱和。

具体而言，针对上述第二类和第三类问题，我们可以选择与这位丈夫不同情况的丈夫们进行访谈，如在家附近的大学读研、继续与妻子分担家务的教师，或不参与高校职称晋升竞争、甘愿当一个低职称的教师，或干脆放弃在高校工作的机会而选择下海经商，等等。针对情境条件，我们还可以结合第一类问题，选择与上述家庭情况不同的夫妻进行比较，如妻子没有工作或者工作不忙，孩子已经上大学或者寄宿、不需要妻子照顾，家中没有需要照顾的老人或者有保姆帮忙；等等。

收集了更丰富的资料后，还需要对上述初步结论进行修订和检验，增加理论的密实度（density，即类属和类属下属性的数量充足）和内部整合度（即理论要素之间的逻辑关系紧密，没有缝隙）。当然，如此获得的理论还只是实质理论，即只能解释教育这一个领域的相关现象；如果希望生成形式理论，还需要在其他相关领域进行理论性抽样，以获得更加概括的、跨领域的理论解释。

三、扎根理论对我国教育研究的意义及本土化调适

从上面的案例可以看到，扎根理论对我们从事教育研究有一些积极的作用。但与此同时，它也存在一些不足，需要根据我国教育研究的实际，进行本土化调适。

（一）扎根理论对我国教育研究的积极作用

首先，扎根理论有利于为我国的教育研究赋权，使我们更多地看到研究现象的本土特色，让研究变得更加有趣，更加富有创造性。目前，我国的教育研究受到西方“宏大理论”的严重“殖民”，绝大部分研究者都在使用西方的概念框架，缺乏自己本土的理论视角。如果我们回归中国的教育实践，面对经验资料提炼本土理论，或许能够获得更贴近中国教育实际的研究成果。当然，吊诡的是，扎根理论本身来自西方，使用它在某种意义上也是一种“殖民”。这种路径主要

使用的是分析性思维，将资料进行系统的“切割”和“拼接”，有可能在某种程度上改造中国人所擅长的整体直观、朦胧诗意的思维方式。不过，由于它只是一种研究路径（并不是实体理论），而且提倡开放地面对资料，从资料中生成理论，因此，与实体的西方理论相比，它更容易让我们看到资料中的本土特色。

例如，在研究“人到中年为何还来北大读研”这个问题时，如果我们套用西方社会学家布迪厄的理论，也许可以说这位丈夫到北大读研是为了积累更多的符号资本和社会资本，进而获得更多的经济资本，甚至为孩子的未来奠定一定的文化资本。而如果我们从经济学的交换理论出发，也可以将夫妻之间的付出和回报看作一种双方利益上的交易。然而，当我们贴近原始资料进行细致的“微分析”时会看到，所有人的行动都带有明显的中国社会文化传统中重伦理道德的印迹。例如，丈夫离家感觉“对不起”妻子，而妻子如果不让丈夫出来又感觉“对不起”丈夫。双方相互的“对不起”诉诸的是人的“良心”，表达的是一种关系伦理，一种相互依存、相互负责的存在状态。这种状态也在丈夫心目中“家”与“国”的同构关系中表现出来，“国家”中的“家”是“国”之为“国”的基础——“没有家，哪有国呀！”如果我们进一步探询这位丈夫以及其他不同类型丈夫们对“国一家”的理解，也许还可以了解中国人（特别是中年男人）面对“忠孝难两全”困境时的行动策略及其背后的文化信念。

其次，扎根理论研究通过一整套相对明晰、可操作的技术、方法和步骤，在经验资料与理论建构之间架起了一座桥梁。特别是“匹配”（fit）这一衡量通过扎根理论路径而生成的理论的标准（即理论与原始资料相匹配），对于我国教育研究的初学者尤其有用。在我国，大部分教育研究者多使用思辨的方法开展研究，他们通常不介绍自己的研究方法，也很少提供原始资料作为支持自己观点的论据。这很容易使年轻学者感到茫然，不知如何向学术前辈学习。扎根理论相对系统的研究步骤和技术，应该能够更有效地帮助初学者掌握一些容易“上手”的思路和工具。

沿用上例，我通过整体提问、三级编码、编

码范式模式以及条件/结果矩阵等工具，对一段访谈资料进行了“微分析”，得出一些基本的概念类属、属性、维度、初步假设和结论以及下一步理论抽样的标准。我希望这个相对详细的分析过程能够告诉读者，这项有关丈夫读研决策的小研究，其部分结论是如何做出来的，结论有什么原始资料作为论据，在解释丈夫和妻子的行动和互动策略上有何新意。例如，在这段话中，丈夫6次使用了“得”这个字，其意思是“不得不”或“必须”“应该”。这表明，这个中年男人在兼顾个人发展和家庭责任的两难时，表现出一种无奈、纠结的心态。

再次，扎根理论路径不仅能够服务于学术研究者，而且能够为实践者改变现状提供思路。在这里，衡量通过扎根理论路径而生成的理论的另一个标准“有用”（work），具有很强的适切性（这来自扎根理论的理论基础之一实用主义）。所谓“有用”指的是，理论可以解释被研究者在处理相关问题时行为的各种模式，能够提供有新意的、促进人们理解的甚至能够促使人们改进工作和生存状态的解释。因此，“有用”的标准也包含了相关性（relevance）、可理解、可推论（generalization，具有概括性）、可控制等要素。由于此类研究生成的理论属于中层理论，与经验事实密切相关，实践者很容易在其中看到经验资料的形态，继而依此改变自己的社会规则和行为方式，也能够预测这些改变可能带来的后果。如果需要，他们还可以对资料进行重新编码，以形成能够解释自己情况的新理论（即满足扎根理论“可修改”的标准），找到新的行动可能性，以改进自己的处境。

在上例中，如果丈夫意识到自己与妻子在社会分工上确实存在性别不平等，他可能会适当调整自己的行动策略，更经常地给妻子打电话，或更频繁地回家料理家务。同理，如果妻子通过阅读上面的分析后意识到，自己的“弱势”地位和“弱者的武器”的使用不仅仅是自己个人的问题，而且具有时代特征和文化遗产，也许她会变得更加坚强，主动与丈夫协商维护家庭稳定的策略，而不仅仅被动地表达自己的担心和歉疚。

（二）扎根理论的限度及本土化调适

扎根理论只是质性研究中的一个路径，它有

自己的优势，但也有一些不足。特别是被运用到中国教育情境时，需要我们做出本土化的调适。

首先，由于扎根理论要求研究者带着尽量开放的头脑进行实地研究，因此很少讨论研究问题的提出、重要概念的定义和概念框架的设计等问题。然而，在我国教育研究界，无论是研究生的论文开题，还是研究者向上级主管部门或基金会申请课题，都需要提供详细的研究计划。而且，初学者如果没有一个相对稳定的研究问题，会感到无所适从，缺乏必要的参照系。

为此，在上例中，我在资料收集之前，专门研讨了研究设计的各个环节，而“人到中年为何还来北大读研”这个问题就是在这样的讨论中生成的。研究主问题提出之后，同学们将其分解成几个子问题，如：“被研究者来北大读研时的背景（国家、家庭、个人）是怎样的？”“被研究者当时是如何做决定的？重要他人（妻子、父母、同事）产生了什么影响？”“来北大后被研究者对这个选择的感受如何（是否有变化）？”然后，大家对研究问题中的重要概念进行了定义，如“人”“中年”“北大”“读研”“背景”“决定”“重要他人”“感受”等，每个概念都提供了规范性定义和操作性定义。此后，大家根据重要概念之间的关系，通过画图表的方式形成了概念框架。我发现，这么做并没有妨碍学生们保持开放的头脑进入现场，因为“开放的”（open）脑袋并不等于“空洞的”（empty）脑袋。只要他们能够暂时悬置之前的前设，反而具有了更加敏锐的眼光和更加缜密的思维，即扎根理论所说的“理论敏感性”。

其次，由于扎根理论主要关注理论的生成，很少讨论研究者身份和研究关系及其对研究的影响以及诸如自愿、回报、保密等伦理问题。虽然这种路径（与质性研究其他路径一样）也要求研究者经常写备忘录，但（与质性研究其他路径相比）更关注理论概念的生成，备忘录中很少讨论研究关系和研究伦理问题。而在当代中国文化语境下，这些问题是非常重要的问题，需要认真对待。中国人很看重人际关系，如果研究双方没有建立起信任关系，研究者很难收集到需要的资料，更不可能为对方打开心扉。相对西方社会，中国社会有更加清晰的等级观念，这也使研究者

需要对自己的身份以及自己与被研究者之间的权力关系有清醒的认识。在教育研究中，特别是当成人面对儿童、上级面对下级时，研究关系对研究的影响非常大。

例如，在上例中，研究者与被研究者是同学，相互之间是熟人关系，这不可避免地会影响到研究的过程和结果。在课堂上有同学提出，本段资料中丈夫显得如此“高调”，有可能由于彼此是熟人，丈夫担心作为访谈者的同学（今后有可能）认识自己的妻子或者两人共同的朋友，因此有意“拔高”自己，使用很多文化格言，以彰显自己作为丈夫的“高大形象”。假设两个人互不认识，受访者是否会更加放松，其表达是否更自然、坦率，贴近日常生活中的“丈夫”形象？

再次，也许是收集资料、分析资料与理论生成同时进行的缘故，扎根理论虽然对资料的类型提出了要求（范围广泛、内容丰富有差异等），但对资料收集的具体技术（如观察、访谈、实物分析等）很少讨论。这也导致它常常被误认为只是一种分析资料的方法。而我自己研究和教学中发现，收集资料的方法本身会严重影响到资料的质量。正如扎根理论的理论基础之一“符合互动论”所指出的，“现实”是由行动者的行动和互动建构的，研究者收集资料的方式会对“什么是资料”以及“收集什么资料”之类的问题产生形塑作用。

例如，在上述研究过程中，很多学生由于没有掌握收集资料的方法，在访谈时不会倾听和追问，在观察时看不到行动中的机制和模式，导致很多资料停留于表面，分析时难以形成清晰、有结构的结论。上述案例中的研究者也分享了自己访谈这位丈夫时的困惑，例如，如何将他从“冠冕堂皇”的文化格言拉回到日常平实的话语层次，如何让他具体举例说明自己的心情，等等。为此，我的课程专门增加了观察、深度访谈、焦点团体和实物收集等环节，并做了很多提高倾听、停顿、追问、重组能力的游戏。在整理资料时，为了更加整体地理解资料（避免过早使用三级编码肢解资料），我们还增加了填“接触摘要单”（每次收集资料后对资料进行整体性描述和提问）、画“矩阵表”（在至少两个维度将资料

入格，形成基本类型）、画“思维导图”（将资料中的重要概念或主题之间的关系用图式直观地表达出来）。如此整理过的资料变得更加丰富、更加有结构，为后面的编码提供了更加坚实的资料基础。

最后，需要特别注意的是，扎根理论诞生于美国。受西方分析性思维的影响，西方研究者使用这种路径生成的理论，其表征方式大多是命题式（proposition）或讨论阐发式（running discussion）。相比之下，中国的学者（特别是教育研究者和实践者）由于更善于捉住“灵光乍现”和“福至心灵”，对“理论”的理解似乎更加丰富，扎根出来的理论有可能具有其他表征形式，如格言、谚语、成语、俚语、歇后语、八股文、诗词歌赋等。在我们对教师实践性知识的研究中，也发现了一些有新意的表征方式，如意象、隐喻、行动公式、身体动作、叙事等。

在上例中，我们也能看到一些新颖的表达方式，如丈夫对时间和空间的定位和想象（过去“我在家时”——现在“我出来了”——未来“一辈子”），意象和类比的杂糅（“海枯石烂永不离心”），文化格言的使用（“糟糠之妻不能忘”），主观和客观的嵌套（主观为家，客观为国）。因此，中国的教育学者有可能拓展扎根理论路径所生成的理论的呈现方式，将命题式陈述和理论性讨论延伸到更加多样的、另类的、富有诗意的表征方式。

总之，相比其他社会科学研究领域，教育学面对的是教育现象，而教育现象具有情境性、独特性、不确定性、价值性等特征。与扎根理论兴起的西方社会学领域相比，我国的教育学有很多特殊的问题需要探讨。因此，我们需要根据我国教育研究的实际，创造适合自己的研究路径和方法论。这意味着我们需要有选择地借鉴和创生其他研究策略和方法，以弥补扎根理论的偏颇与不足。

（陈向明，北京大学教育学院教授，北京100871）

（原文刊载于《北京大学教育评论》2015年第1期）

中国教育实证研究中的定量方法：五年应用述评

吕 晶

一、引言

回溯教育科学的发展历程，其每一次重大突破，都离不开研究方法的发展。教育调查方法的进步，使教育调查受到重视并逐步走向规范；教育测量方法的进步，使各种客观标准化测量迅速兴起；教育统计方法的进步，使教育研究向量化、科学化跨进了一大步。“工欲善其事，必先利其器”，要提高教育科研水平，就必须掌握合适的研究方法。定量和质性研究方法是教育科学研究中最主要的两大类研究方法。其中，定量研究方法主要服务于教育实证研究中的定量研究或混合研究，是对实验性数据、观察数据和挖掘数据进行统计分析，对理论假设进行检验的过程中所应用的一系列数学、统计学方法的统称。建立在实证数据基础上的定量研究方法具有客观性、准确性和可预测性，为教育研究走向科学化奠定了基础，也越来越受到国内外教育研究者的重视。

定量研究方法被越来越广泛地应用于教育实证研究中是符合社会科学的一般发展规律的。社会科学是研究并阐述各种社会现象的特征及其发展规律的科学。而探索社会现象的特征及其发展规律又需要依靠反映社会现象量的概念抽象和反映社会现象质的概念抽象。相比于质的概念抽象，量的概念抽象更直观、客观和具体，且往往为质的概念抽象提供科学基础。换句话说，在没有量的事实依据情况下所进行的质的研究往往不会被人们当作科学论断所接受。虽然不是所有的社会现象都可以定量地描述，但总体上，相较于质性描述，定量描述能更客观、科学地反映社会现象的特征和发展规律。由于量的概念抽象需要依靠定量方法来实现，现代社会科学正在完成一个由以质性研究方法为主向定量研究方法为主且质性、定量研究方法相结合的过渡，社会科学发展的一般规律也是随着研究的深入，其量化水平越来越高（佟庆伟，2004）。

定量研究方法在20世纪30年代被引入中国教育领域后，逐渐引起重视。尤其是2015年华东师范大学教育学部、北京师范大学教育学部、全国教育科

学规划领导小组办公室、光明日报教育研究中心共同在全国范围内发起教育实证研究论坛后，定量方法作为教育实证研究的主要方法学范式之一被更多地应用在中国教育实证研究中。袁振国（2017）在《实证研究是教育学走向科学的必要途径》一文中更是强调了定量研究方法、手段和技术是促进教育科学取得突破性进展的关键。然而，目前我国教育科研中使用定量研究方法的频率、水平还是远远低于美国。例如，在美国教育技术领域，结构方程模型、元分析等定量方法被广泛应用在远程教学理论建构、理论综述等方面。相对而言，我国教育技术领域多采用质性方法辅以现代教育媒体技术的研究范式去建构、探索教学理论。另一方面，由于我国很多教育科研人员没有系统地学习过定量研究方法，还没有对教育实证研究中定量方法的应用形成全面客观的认识，所以在应用定量方法时显得力不从心，甚至误用、错用。针对这些问题，刘建设（1999）、黎荷芳（2001）提出定量方法的应用要与研究问题相切合、与质性方法结合使用。有研究更是指出我国教育研究中对定量方法的重视还是不够，还有待加强（谢美华，2005，第33—34页）。Yue & Xu（2020）随后通过对自1978年改革开放以来我国教育实证研究中有关定量方法研究的分析，总结出定量研究方法在我国教育领域发展的四个主要阶段，肯定了我国教育研究中定量方法的发展进步，建议再进一步关注教育测量与评价、提高大数据的挖掘和利用率、增强研究人员的协作和资源共享。然而，这些研究述评多是对统计数据的简单描述与概括分析，针对应用中遇到的具体问题的详细分析几乎没有，更没有针对实际方法误用的纠正，无法切实帮助到具体方法的应用者。

因此，详细梳理、分析定量研究方法在我国的实际应用情况，并尝试对其不足给出较为具体的建议，对定量研究方法在教育领域的发展与成熟具有重要价值。借鉴相关研究，通过综合运用频率统计法、关系网络法、比较研究法、内容分析法，本文以11本教育综合类中文社会科学引文索引

(CSSCI) 期刊在2015—2019年间发表的应用定量方法的论文(仅应用了最基本的频次统计的排除在外)为研究对象,分析总结定量研究方法五年来在我国教育实证研究中的应用现状,并针对具体问题提出对策与建议;整理了一些常见的定量研究方法的误用情况,并针对这些误用给出正确应用建议;分析了定量方法在教育实证研究中的使用趋势。

二、研究方法

(一) 研究对象

本研究随机选取了11本影响因子高于1.00且排名前25%的教育综合类中文社会科学引文索引(CSSCI)期刊,以这些期刊在2015—2019年间发表的使用定量方法的论文为研究对象(频次统计除外)。这些期刊包括《华东师范大学学报(教育科学版)》《教育研究》《清华大学教育研究》《中国教育学刊》《北京大学教育评论》《复旦教育论坛》《教育发展研究》《教育学报》《教育科学》《教育研究与实验》和《湖南师范大学教育科学学报》(如表1所示)。选取综合类期刊是为了从总体角度比较分析教育科学各领域的情况,而排名靠前的CSSCI论文可以代表目前我国教育领域较优秀的研究成果,更有利于分析定量方法在我国教育实证研究中应用的成熟度和先进程度。

(二) 文献筛选与数据采集

1. 文献筛选。本研究先后进行三次文献筛选。第一轮筛选是在中国知网(CNKI)数据库高级检索设置的检索条件中设置“定量”“量化”“统计”“问卷”“量表”“测量”“实验数据”“观测数据”“挖掘数据”“数据分析”“样本”为关键

词、篇名、摘要和全文的共同检索词,即只要论文的关键词、篇名、摘要或全文中出现这些检索词的任何一个就会被检索出来。把发表时间设定为从2015年1月1日到2019年12月31日,初步筛选出4722篇论文。批量下载这些文章的PDF版;在第二轮筛选中,粗略地查看第一轮筛选的4722篇论文,留下有数据的文章,共余2669篇;在第三轮筛选中,再一次查看2669篇论文的方法论部分,剔除掉没有应用定量方法的和只应用了频次统计或百分比的论文,剩余1018篇论文。

2. 数据采集。在CNKI的4722篇论文中选择筛选出的1018篇论文,导出数据,数据导出类型设为“Reworks”。将导出数据上传到CiteSpace,并将“题目、作者、发文机构、关键词、期刊、年份”数据导出为.csv文件。

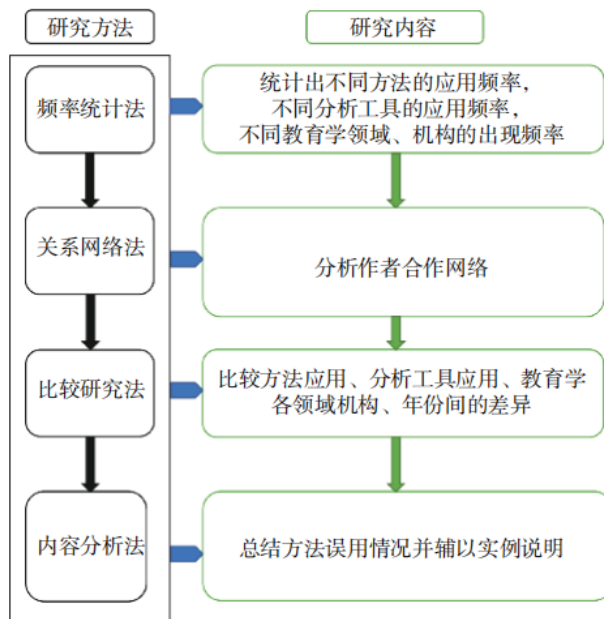


图1 文献分析流程图

表1 选取期刊简介

序号	期刊名	IF (2018)	排序
1	华东师范大学学报(教育科学版)	6.327	2/274
2	教育研究	4.487	5/274
3	清华大学教育研究	2.661	14/274
4	中国教育学刊	2.429	17/274
5	北京大学教育评论	2.379	20/274
6	复旦教育论坛	2.010	29/274
7	教育发展研究	1.893	33/274
8	教育学报	1.608	39/274
9	教育科学	1.584	41/274
10	教育研究与实验	1.360	48/274
11	湖南师范大学教育科学学报	1.187	57/274

（三）文献分析

本研究采用的具体文献分析方法如图1所示。

由于定量研究方法的种类很多且很少被列在目标文献的标题或关键词中，单靠工具软件记录不同方法的应用频率等统计方式，会出现较大的偏误。因此，本研究依次浏览1018篇目标文献的“研究方法”和“研究结果”部分，人工记录具体定量方法和分析工具的使用次数。教育学各领域和署名机构出现的频次则依赖CiteSpace对关键词和机构的计量分析。作者合作网络也通过CiteSpace分析完成。2015年到2019年各年段的发文量通过Excel进行统计。

三、定量研究方法在我国教育实证研究中的应用现状分析

根据对目标文献的分析，发现全国教育实证研究论坛的召开对重视定量研究方法的应用具有一定的推动作用，这五年来定量方法在我国教育领域的应用越来越广泛。然而，我国教育科学领域在定量研究方法的应用上取得一定进展的同时也存在着诸多问题，并呈现出以下特点。

（一）应用定量研究方法的论文数量增多，但总占比仍然较小

如图2所示，应用定量研究方法的目标文献数量逐年增加，且每年较上一年的增率都大于3%（2016—2019年增率分别为4.28%、3.59%、

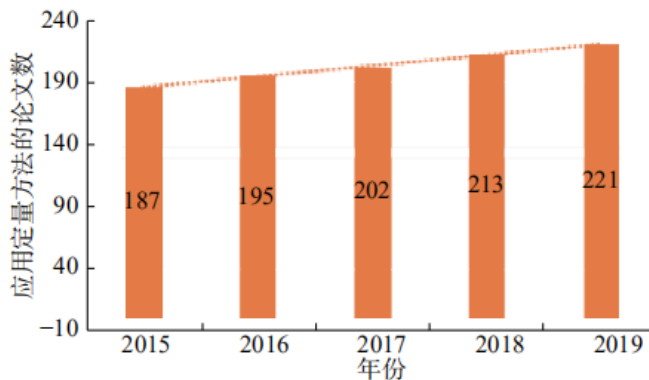


图2 应用定量方法的论文数量

5.45%、3.76%)。由于“全国首届教育实证研究论坛”的召开是在2015年10月，所以2015年的文献量可以当作实证论坛召开前的基准量。由此可见，定量方法的应用比率有一定提高，“教育实证论坛”的开展也加速了学界对定量研究范式的重视。然而，根据朱军文和马银琦（2020）《教育实证研究这五年：特征、趋势及展望》一文中的表6统计，11本目标期刊五年间的发文总量为6994篇，其中实证文献为2434篇。那么，应用定量方法的研究只占了总发文量的14.56%、占实证发文量的41.82%。作为教育科学研究的主要方法学范式之一，14.56%的总占比说明定量方法在我国教育研究者中的普及度还有待加强，并且将会有很大的提升空间。

（二）研究团队间沟通较少，学者间存在不稳定的小团体合作

图3左半边展示了完整的作者合作网络，除了

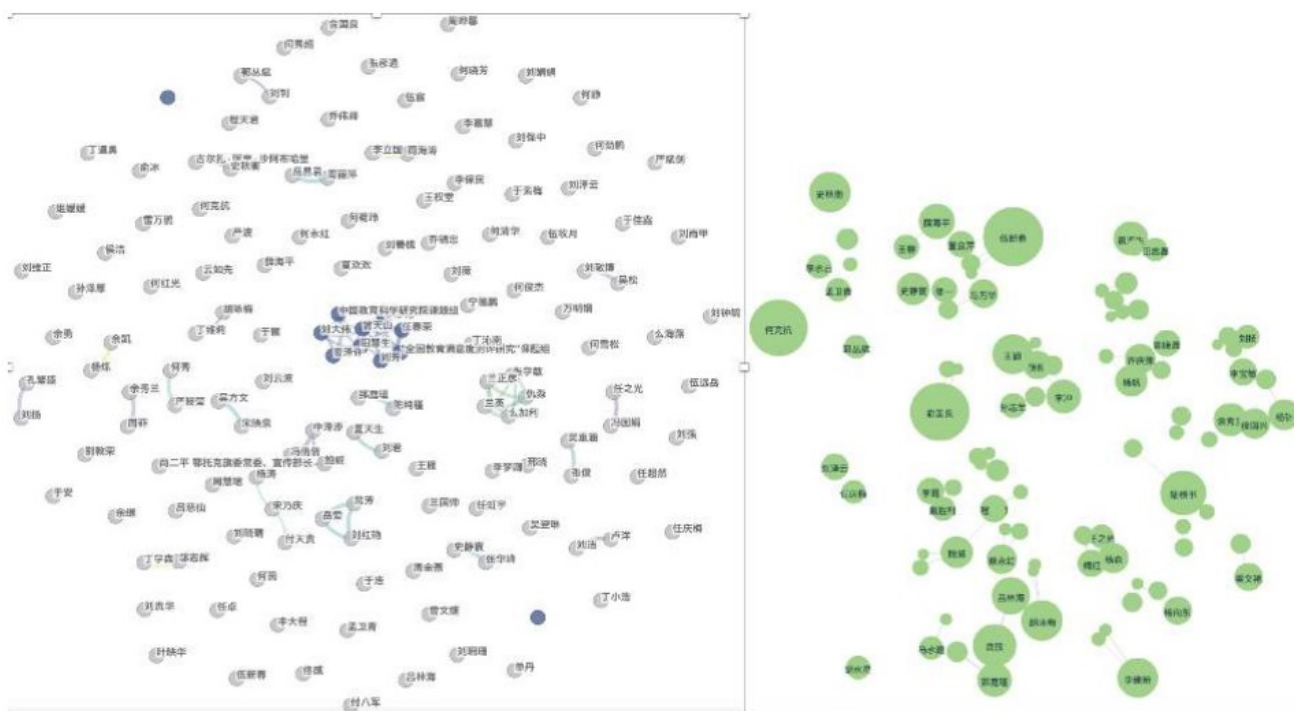


图3 核心作者合作网络图

少数团体间存在着密集的合作网络外，大多数团体网络联结密度小、沟通性和互动性差，尤似一盘散沙。图3右半边展示了随机选取的200篇论文作者的合作网络，核心作者周边会有些较小较弱的联结，但核心作者之间很少存在联结。说明核心作者倾向于和不同的人合作，但这种合作并不是长期稳定的，且强强联合的研究合作十分匮乏。由此可见，我国教育科学领域研究团队之间的沟通交流不够，知识的传播和分享比较差，长久下来将不利于促进整个领域的发展。因此，各大研究机构需要鼓励团队之间的合作，争取通过大团队作战获得更突出的成果。

（三）方法涉及面广，但新方法、新技术的引进速度慢

如表2所示，在目标论文中，最常用的统计分析方法是描述性统计、回归分析、相关分析、方差分析、差异检验等传统定量研究方法。少部分研究运用了结构方程模型（SEM）、因子分析（FA）、多层线性模型（MLM）、元分析（MA）等较为先进的定量方法。元分析结构方程模型（MASEM）、多层结构方程模型（MSEM）、多层元分析（MLMA）等在国际教育科研中运用越来越多的新方法更是没有出现在目标文献中。在辅助分析的工具软件上，绝大多数研究（94.18%）使用了统计产品与服务解决方案（SPSS）和AMOS软件；极少的研究（5.49%）用到了Mplus软件、STATA软件及统计分析系统（SAS）等功能性软件。极个别的研究

（0.33%）用到了R语言、Python语言等这些国际流行的开源工具。而SPSS与AMOS都属于易学易用但处理前沿统计能力相对较差的工具，比如AMOS由于采用最大似然估计（maximum likelihood estimation, MLE）或加权最小二乘法（WLS）或未加权最小二乘法（ULS）等估计导致处理类别数据、类别潜变量的功能不够完善。如果使用Mplus就可以较好地处理类别潜变量、类别数据，并提供稳健估计。此外，利用数据挖掘技术（0.27%）、大型调查数据库（4.13%）收集数据的研究也非常少。这在一定程度上说明了我国教育科学研究中方法涉及面广，大多数成熟的方法都有应用，但新方法、新技术的引进、推介速度慢，远远落后于美国等教育强国。研究方法的落后又制约着我国教育科学的发展。因此，各高校教育学院和教育科研院所要更多地吸收、培养专门从事定量研究的人才，开设相关课程，普及量化知识。不仅要吸纳借鉴国际上先进的技术和分析方法，更要融合其他学科的方法、技术作为辅助，在本土、本学科内结合实际应用不断改进新技术、研发新方法。

表3列出了一些国际上常用的但目标论文中较少或从未出现的定量研究方法供教育研究者参考。比如，MLMA、MASEM和MSEM这三种在国际上被认为是很有前途的分析方法就没有出现在目标文献中。其中，MLMA特别适用于总结层次结构数据，它能够检查研究之间的差异、揭示组间异质效应（van den Noortgate & Onghena, 2008）。鉴于基于层

表2 定量方法和分析工具应用情况

定量方法	论文篇数占比	分析工具	论文篇数占比
描述性统计	77.56%	SPSS	80.05%
回归分析	54.09%	AMOS	14.13%
相关分析	25.59%	Mplus	3.84%
方差分析	17.12%	STATA	0.49%
t检验	34.09%	SAS	0.85%
卡方检验	4.80%	R	0.19%
结构方程模型	9.88%	Python	0.14%
因子分析	15.05%	其他	0.31%
多层线性模型	1.90%		
元分析	0.88%		
其他	4.87%		

注：由于某些研究中使用了不止一种定量方法，故总百分比加起来超过100%。

次结构数据的已发表研究越来越多，使用MLMA变得越来越重要（van den Noortgate & Onghena, 2003）。在目标文献中，有一篇是运用MA综合不同研究的效应量来分析课外辅导对学生学业成绩的影响效应。有时，一个研究可能既报告每个学生的成绩，也会报告不同班级的整体成绩，那么采用MA就会忽略文献中可用的一部分研究数据。而MLMA则能够总结不同班级和学生个体的特定效应大小，在样本量足够的情况下，会给出更详尽的分析。MASEM结合了MA和SEM的优点，是一种强大的方法学工具

（Lv & Maeda, 2020）。它不仅可以帮助研究者利用SEM进行理论的定量综述、增强结论的可概括性，还可以帮助研究者测试在单个研究中无法测量的复杂模型。这种方法的应用正在逐年增加（Lv & Maeda, 2020），研究人员越来越有兴趣将该方法应用于实证研究。著名期刊《研究综合方法》（Research Synthesis Methods）在2016年还特别出版了一期关于MASEM的专刊，讨论了当前MASEM遇到的问题和对未来研究的建设性意见。相似的，MSEM是一种广泛应用于多层次研究的先进的SEM技

表3 国际上常用但目标论文中较少或从未出现的定量研究方法

方法	简介	代表性分析软件	参考资料
项目反应理论 (IRT)	是一系列心理统计模型的总称，其分析结果综合直观地呈现出项目的难度、区分度等特征，主要被用来指导量表的编制和项目筛选	R、MatLab、Mplus、PARSCALE、Winsteps、MUITILOG	Handbook of Item Response Theory: Volume 1, Models (van der Linden, 2016, pp. 1-623); Volume 2, Statistical Tools (van der Linden, 2017, pp. 1-456)
聚类分析 (Cluster Analysis)	把研究对象进行归类，使同一类的个体具有高度的同质性，不同类之间具有高度的异质性的多元分析技术的总称	R、SAS、SPSS、S-Plus、MatLab	Handbook of Cluster Analysis (Hennig, Meila, Murtagh, & Rocci, 2016, pp. 1-773)
时间序列分析 (Time Series Analysis)	分析随机数据序列所遵循的统计规律的方法，用于系统描述、系统分析和预测未来趋势	R、SPSS、Eviews (计量经济学软件包)、SAS、Stata、Python、MatLab	Time Series Analysis and Its Applications with R Examples, third edition (Shumway & David, 2011, pp. 1-596)
基于Agent的建模 (ABM)	结合了博弈论、复杂系统、计算社会学、多智能体系等元素，用于模拟个体或团体的行为和交互的计算模型，以评估他们对整个系统的影响。	Python、Swarm、C++、Smalltalk	Agent-Based Models, second edition (Gilbert, 2020, pp. 1-153)
多层线性模型 (MLM)	处理层次数据或多水平数据的多元统计方法，能正确处理个体效应和组效应之间的关系。	R、Mplus、HLM、SAS、MLWIN、MIXOR	多层次模型 (卢克, 2016, pp. 1-89)
元分析 (Meta-Analysis)	用测量和统计分析技术对众多实证文献的统计结果再次统计，得到一个综合的结论。	R、CMA (综合元分析)、RevMan、Meta-Analyst、MIX	Introduction to Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009, pp.1-421)
多层元分析 (Multilevel Meta-Analysis)	由于元分析本身就是多层次的，所以多层元分析一般指三层或以上的元分析。该方法特别适用于总结层次结构数据	R、Mplus	Quantitative Synthesis of Research Evidence: Multilevel Meta-Analysis (Moevaert, 2019)
元分析结构方程模型 (MASEM)	结合了元分析和结构方程模型方法，对应用结构方程模型的实证研究结果再次统计，增强结论的可概括性；还可用于测试某些变量在单个研究中难以测量的复杂的结构方程模型	R、Mplus、SAS	Meta-Analysis: A Structural Equation Modeling Approach (Cheung, 2015, pp. 214-278)
多层结构方程模型 (Multilevel SEM)	结合了层次线性模型和结构方程模型方法，使得结构方程模型中变量之间的关系可以存在于不同层面，并在不同层面上研究变量之间的复杂关系	R、Mplus、SAS	Multilevel Structural Equation Modeling (Silva, Bosancianu, & Littvay, 2019, pp. 1-179)
社会网络分析 (Social Network Analysis)	用数学方法、图论等测量和调查社会系统中各部分的特征、相互间的关系，包括人的关系、互动和网络结构	UCINET、Pajek、NetMiner、MultiNet、StOCNET	社会网络分析，第二版 (诺克, 杨, 2012, pp. 1-201)

术 (Ryu, 2014), 用来弥补MLM和SEM的缺憾。很多方法论研究也阐述了如何将MSEM应用在不同的SEM模型中, 如验证性因子分析 (Geldhof, Preacher, & Zyphur, 2014)、测量和结构模型 (Rabe-Hesketh, Skrondal, & Zheng, 2007)、中介模型 (MacKinnon & Valente, 2014) 和调解模型 (Preacher, Zhang, & Zyphur, 2016), 并且将其实现方法编码到已有的工具软件中, 如Mplus version 5和Stata version 8, 进而引入到社会科学领域的实证研究中。随着其理论和软件包的发展, 如今的MSEM在社会科学领域的使用也越来越普遍 (Ryu, 2014)。

(四) 方法误用情况屡见不鲜

目标文献中有些研究未能根据实际情况来正确选择更合适的定量方法。如某作者收集到的是逐月的追踪数据, 在因素回归时把时间当作一个自变量来分析, 以预测某学生群体未来成绩。事实上, 使用时间序列分析可能就会更合适该研究。需要注意的是, 并不是复杂的方法就比简单的方法好, 有时复杂的方法不但费时费力, 还可能在解决旧偏误、甚至未能解决旧偏误的同时增加新的偏误。分析方法的选择是要以完成研究主题、目的为出发点, 在保证分析结果具有高信效度的基础上, 尽可能选择相对简单、易于操作的定量分析方法。

另一方面, 相当多的目标研究没有对其所用定量方法的适用条件进行考量, 而是直接将采集到的实证数据用来跑数学模型、进行统计检验等。从方法论层面讲, 所有定量研究方法都有其适用条件, 并且不同方法的适用条件存在一定差别。例如, 对多元回归而言, 当自变量之间的相关性较强时, 回

归系数的估计就会出现很大偏差、甚至无法进行。只有当回归模型的自变量之间没有密切关系时, 多元回归分析的结果才有效。而应用多元回归的几乎所有目标文献都没有完成自变量之间的独立性检验。同样地, 目标文献中应用了MLM的研究也没有对观测数据做任何独立性检验。在MLM中, 虽然数据的嵌套结构并不遵循回归的独立性假设, 观测数据的聚类可以有较强相关性, 但其也是一种基于回归的分析方法, 因而应用MLM需要对每一层次的观测数据的独立性进行检验。同样情况也普遍存在于其他定量方法的应用中, 如方差分析时缺少对观测数据分布状态的检验、修正等。

其他方面的误用导致分析结果出错的情况也同样存在, 如使用错误的估计方法, 忘记给不同样本数据做加权平均而直接用了平均数, 或选择错误的数据收集和处理方法等。再如, 在应用FA、SEM等依赖于大样本量方法的文献中, 却发现了小样本量 (小于200, 且未使用贝叶斯等方法) 的存在等。此外, 绝大多数存在缺失数据的目标文献都没有在统计分析前对缺失数据进行预处理。然而, 由于绝大多数数理运算、估计都是建立在完整数据的前提下, 对缺失数据的检测和正确处理是保证分析结果具有高可信度和有效度的前提。

由此可见, 在应用特定定量方法前, 必须在考虑其适用条件的情况下对实证数据进行检验, 并尽量避免因方法的误用而得出错误的结论。因此, 教育研究者应努力加强自身定量分析的专业素养, 在提高统计分析知识水平的基础上多与定量方法论专家交流合作, 以避免在方法设计和应用上出现漏洞、偏差。期刊还应专门邀请从事定量方法论研究

表4 各教育领域每年发文量

教育领域	2015年	2016年	2017年	2018年	2019年
教育经济	39	52	42	50	52
教育社会	32	20	19	27	22
教育心理	20	24	22	27	33
教师教育	11	4	9	9	10
职业技术教育	9	7	15	13	6
教育技术	10	17	19	22	23
高等教育	20	17	14	17	11
教育学原理	1	1	0	2	0
教育史	0	2	0	0	1
其他	45	51	62	46	63

的专家参与方法设计等方面的审稿工作，以保证期刊质量，避免错误信息对读者的误导。

（五）不同教育领域间差距大

我国教育研究的各个领域对定量方法的重视程度有所不同，应用频率、成熟度也差别较大。如表4所示，在目标文献中，教育经济学、教育社会学、教育心理学这些偏实验的、收集数据较多的交叉学科领域应用定量方法较多，在方法的掌握上也更为成熟。教师教育、职业技术教育等专业领域对定量方法的应用非常少。教育技术、高等教育等领域虽然对定量研究方法的重视程度不如教育经济学等学科，但也有一定占比，尤其是教育技术领域研究的量化程度在五年内呈现逐步上升的趋势。除此之外，在应用定量研究方法的目标文献中，几乎看不到如教育学原理、教育史等教育基础理论领域的研究。

教育学各领域间在应用定量研究方法上的这些差异在一定程度上是与各个领域的性质相关联的。教育经济学、教育社会学、教育心理学是教育学与数学、经济学、社会学、心理学交叉的学科领域。经济学、社会学和心理学的量化程度相对比较高，定量研究方法在这些学科发展的也更加成熟，导致教育学与这些学科的交叉学科领域的量化程度要比纯教育的学科领域要好很多。与此相反，教师教育、职业技术教育是教育学本身衍生出的学科领域，基于定量研究和混合研究在教育学领域的发展相较其他社会科学（如心理学、经济学）领域的发展更为缓慢，这些教育学领域在定量方法的应用上也会与其他教育学领域（如教育心理学、教育经济学）有一定差距。而教育学原理、教育史这些更偏重哲学思辨的人文领域，往往比较难实施测量和收集大量数字化数据，因而更加难以广泛地应用定量方法。

事实上，在定量方法的应用上，这种教育领域间的差距在欧美等国家也是存在的。我国教育学各领域间存在这样的差异也说明定量研究方法的实际应用中虽然存在很多问题，但也没有严重到“滥用”定量方法的地步。我们需要提高教师教育、职业技术教育领域的研究人员的量化素质，同时鼓励教育学原理、教育史等领域采用量化分析辅助哲学思辨，使相关研究更客观、科学。

（六）缺乏定量与质性方法的结合，研究深度不够

目标文献中鲜有定量与质性研究方法的综合运用，多是单一定量研究方法的应用或多元定量研究方法的嵌套应用，很少辅以质性数据的收集和严谨的质性分析。这说明我国教育研究者在方法的使用上常常是把定量和质性研究方法明确划分开的，认为有了定量方法就不需要考虑质性方法，而往往不去关心选定的研究方法对具体研究主题的贡献大小。在某种程度上也说明有些研究者甚至可能是基于研究方法而选择研究主题，而不是基于研究主题进而选择研究方法，本末倒置了研究主题和研究方法的关系。另一方面，绝大多数研究采用定量研究方法的单独应用，从侧面表明了我国教育实证研究有流于表面、研究深度不够的缺憾。

由于定量研究方法是建立在大量实证数据基础上的、针对教育现象的本质或因果关系等做出概括性分析的研究范式，其统计分析结果往往强调教育现象的本质或因果关系等在特定群体中的“普遍性”。然而，教育是一种错综复杂的人文现象（付瑛，周谊，2004），许多个体或小样本群体具有特殊性。虽然定量研究方法逻辑严谨、客观科学，但很难对大样本中的少数特征做精确测量与分析，无法进一步深入研究教育现象。相反，定性研究着重观察、描述个体或少数群体的行为（Mcintyre，2003，p. 15）、感知和交互活动，可以帮助研究者发现教育现象中蕴含的复杂规律。因此，在实际应用中，很多定量研究方法是和质性研究方法混合着使用的，即从定量角度研究教育现象的一般性规律，再从质性角度对教育现象进行深入探究。

例如，某目标研究应用多元回归分析探究了自信心、父母期待、同伴竞争等因素对学生考试成绩的影响。然而，只进行多元回归分析，研究者仅能知道自信心帮助学生提高学习成绩，却无法了解到自信心让学生对学习产生兴趣从而愿意花更多时间来学习。其实，研究者可以采用收敛并行设计补充信息量，研究过程的同一阶段分别使用定量和定性方法，对两类方法进行同等优先级排序，并在分析过程中保持两类方法独立性，然后在分析完成后混合结果、整体解释、得出结论（Creswell & Clark，2018），即在一个学期里，研究者通过调

查问卷收集到有关学生自信心、父母期待、同伴竞争等量化数据，以该学期中的某次综合考试成绩为因变量，应用多元回归分析测出这些因素对学生成绩的影响；并在同一学期中，研究者对部分学生进行焦点小组访谈，以探究自信心、父母期待、同伴竞争等因素是如何具体影响学生成绩的；最后，研究者把多元回归分析结果与对焦点群体进行的质性分析结果相融合，得到相对全面的研究结论。

这种混合应用在美国教育研究中获得了广泛的关注、认可和运用（唐涌，2015）。例如，Greene（2012）阐述了混合研究方法比仅使用纯定量或纯质性研究方法进行研究产生了更有意义的结果。Teddle & Tashakkori（2012）概述了教育研究者在使用混合研究方法前应考虑的问题。Creswell（2015）出版了有关混合研究方法的工具书。甚至还有诸如《混合方法研究杂志》（Journal of Mixed Methods Research）这种专门发表应用了混合研究法论文的期刊。而定量和质性方法的混合使用虽然被我国不少学者提议（如，高潇怡，2010；向荣，2019；田虎伟，2007；张东辉，2013；邓猛，潘剑芳，2002；张绘，2012），但是在实际教育实证研究中的应用范围还有待扩大。因此，建议我国各高校教育学院和教育科研院所以教育研究成果的质量评估为基础和出发点，在政策上鼓励教育研究的深入，从而促进混合研究方法在教育实证领域的应用。

四、定量研究方法误用的纠正建议

正确使用定量研究方法对教育实证研究至关重要，倘若误用，即便是成熟、科学的方法和技术，也会导致结论的错误。本研究通过阅读分析目标文献发现，定量研究论文中存在着诸多定量研究方法误用的现象，如忽略定量研究设计、忽略统计学方法的应用条件、盲目使用统计软件、分析方法不恰当等，影响研究结论的正确性。为改善定量方法误用的现状，下文对目标文献中出现频率较高的方法误用情况进行梳理、分析。

（一）回归分析误用的纠正建议

在教育实证研究中，回归分析这种研究因变量和自变量关系的预测性建模技术常常受到研究者青睐，不同回归建模方法在目标文献中有所涉及，如多分类logistic回归、二元logistic回归和分层回

归等。虽然很多研究者能够根据研究目标和自变量、因变量类型等选择较为合适的回归分析方法，但由于忽略回归分析的条件限制，仍然存在方法的误用。

例如，某目标研究收集了2473份关于学生每周花在学生工作与社团活动的时间、是否担任学生干部、大学学业成绩对大学学业排名影响的研究数据。在没有对原始数据进行任何检查和预处理的情况下直接进行多元线性回归分析，且未进行残差分析就得出回归模型构建较好的结论。

谬误1：没有对数据的异常值（包括缺失数据）进行任何预处理。回归分析对异常值（包括缺失数据）较为敏感，如果有异常值存在，不对异常值做任何预处理的话，可能会使估计结果产生偏差。

纠正1：在回归分析前可通过散点图、箱线图、正态图、描述统计等检验数据中是否有异常值（包括缺失数据）的存在。若存在异常值，一般要先把异常值剔除。但如果考虑到实际情况确实无法剔除异常值，则应考虑应用稳健回归建模。

谬误2：没有对自变量的共线性情况进行分析。在实际分析中，自变量之间彼此相关的现象很容易出现在线性回归中。如果一个或多个自变量和其他自变量之间显著相关，则可能存在共线性问题。虽然适度的共线性不会对回归分析造成大的问题，但是严重共线性会导致分析结果不稳定，导致本该显著的自变量不显著、本该不显著的自变量显著，甚至导致回归系数的正负估计与实际完全相反情况。

纠正2：在回归分析时，一般使用VIF值判断共线性情况。如果VIF小于5，则可判定回归模型存在轻微共线性问题，不用进行处理。如果VIF介于5和10之间，则可判定回归模型存在共线性问题，根据实际需求判断是否需要进行处理。如果VIF大于10，则可判定回归模型存在严重的共线性问题，需要进行处理。或使用容差值，即 $1/VIF$ ，判断共线性情况。如果检测出模型存在严重共线性问题，可以通过相关分析找到重要的共线性自变量并剔除，也可以应用逐步回归法让系统自动选择共线性自变量并剔除。若考虑到实际研究情况不希望把某个或某几个自变量从回归模型中剔除，则应考虑使用岭

回归（应用较多）或Lasso回归（应用非常少）。

值得注意的是，本例是应用回归模型检测多个自变量对学业成绩的影响情况，因此研究者需要考虑是否处理共线性问题。倘若某个回归模型只是用来预测，那么只要模型拟合指标好，共线性问题通常不会影响预测结果，也不必处理。

谬误3：没有进行残差分析。由于回归模型中的预测值和观测值的差异是随机且不可预测的，那么回归残差（即真实误差的估计）也应该是随机且不可预测的。如果在残差中发现有可预测的信息，则说明回归模型中缺少了某些可预测信息。由此可见，残差分析是回归分析中非常重要的部分。若残差的正态性、独立性及方差齐性假设不能满足，说明回归模型的构建差，回归估计结果不准确。

纠正3：回归分析过程中保留残差项。然后，对残差的正态性、独立性及方差齐性依次进行检验。残差的正态性可使用正态图等进行检测，若残差整体上满足正态性，说明模型构建好，若残差正态性差，说明模型构建差，需要重新构建回归模型。残差的独立性可使用德宾—沃森（D—W）自相关性检验，若D—W值介于1.7到2.3之间（接近2），说明残差独立，模型构建好，若D—W值小于1.7或大于2.3（明显偏离2），说明残差自相关，模型构建差，需要重新构建回归模型。残差的方差齐性可通过分别与自变量或因变量作散点图进行检测，若散点没有规律性，说明方差齐性，模型构建好，若散点有明显规律性，说明方差异质，模型构建差，需要重新构建回归模型。

（二）t检验误用的纠正建议

在目标文献中，常用独立样本t检验或配对t检验来比较两组数据间的差异（几乎很少出现应用了单样本t检验的研究）。即便大多数研究能够根据比较样本等具体情况适当选择独立样本t检验（两比较组样本量不同）或配对t检验（两比较组样本量相同），但不少研究在t检验过程中仍存在问题，影响分析结果的可靠性。

例如，某目标研究采用t检验（未说明t检验的类型）比较不同类型学校支持对抑郁大学生心理健康干预的差异。在该研究中，实验组有男生24人，女生26人，实验组干预前的总体均分是189.72（SD=58.07），实验组干预后的总体均分是165.18

（SD=47.73），且不存在干预后对同一被试得分的重复测量。该研究用t检验对比了实验组干预前和干预后的总体得分情况。

谬误1：没有交代清楚所应用t检验的类型。不同类型t检验过程是不同的。如独立样本t检验是以t分布为基础，用于检验两个独立样本总体均值是否相等，即是否成立；而配对t检验用于检验两个关联样本的总体差值均值是否为0，即是否成立。因此，两种检验方式得出的结果也会有些许差别。

纠正1：在本例中，实验组干预前和干预后的样本都是实验组原样本（自身配对），实际上两个样本的个体是两两配对的，只有50个独立个体。只有采用配对t检验，把每组配对当作一个单位进行统计检验，才能确保结果的可靠。但是原文中未交代清楚具体t检验的使用情况，降低了分析结果的信度。

例如，某目标研究采用独立样本t检验比较大学生的直系亲属中是否有人吸毒对贝克抑郁量表（BDI）得分的影响的差异。在该研究中，直系亲属中有人吸毒的样本量为10，BDI得分均值为5.20（SD=4.39）；直系亲属中无人吸毒的样本量为2778，BDI得分均值为5.83（SD=6.31）。除了平均数、标准差外，在独立样本t检验前，研究者未对两组数据做其他统计学分析。

谬误1：忽略了独立样本t检验对样本正态性的要求。在本例中，直系亲属中有人吸毒的样本量为10，对独立样本t检验而言，若是小样本，则要求样本的总体必须服从或近似正态分布。由于本例中直系亲属中有人吸毒的样本量过小，仅靠提供的均值与标准差无法判断该样本的总体是否服从或近似正态分布（t检验对正态性稳健）。

纠正1：鉴于本例中的样本量较小，在进行独立样本t检验前先要检验样本的正态性。如果数据不呈正态或近似正态分布，可以使用非参数检验。或可以先通过算法转换把原始数据转换，使之服从正态或近似正态分布后再进行独立样本t检验。如果原始数据转换后仍不能呈正态或近似正态分布，再使用非参数检验替换独立样本t检验，来分析数据。但数据转换可能引起BDI得分和对直系亲属中是否有人吸毒关系的估计偏误，因此不建议使用数据转换法。

谬误2：忽略了独立样本t检验对方差齐性的要求。独立样本t检验要求两组样本的总体方差齐性。在本例中，直系亲属中有人吸毒的BDI得分均值为5.20（SD=4.39），直系亲属中没人吸毒的BDI得分均值为5.83（SD=6.31）。两样本的标准差有一定差别，可能出现方差不齐现象。

纠正2：在本例中，两样本的总体可能出现方差不齐现象，因而需要对两样本数据进行方差齐性检验。如果方差不齐，则采用Wilcoxon秩和检验、或近似t检验方法替代独立样本t检验，进行分析。

（三）方差分析误用的纠正建议

三组或三组以上样本的差异性检验需要用到方差分析这种重要的处理多元信息的分析方法。但在部分目标文献中，不同的方差分析方法却被混淆误用，致使所采用的方差分析模型与研究设计不匹配，得出的结论出现较大偏误。

例如，某目标研究采用多个单因素方差分析分别比较我国不同区域幼儿园物质条件、师幼互动、健康与安全及幼儿发展的差异。在该研究中，各地区个案数及分析结果如表5所示。研究者对东中西部地区的数据进行了方差齐性检验、事后检验，但未做正态性检验。

谬误1：未检验方差分析的适用条件。在本例中，东部样本量10个、中部5个、西部7个，虽然是三个独立样本且方差齐性，但样本量过小，各种数据很可能不服从正态或近似正态分布（方差分析对正态性稳健），可能不满足方差分析的适用条件。

纠正1：在进行方差分析前，应先对各个样本数据进行正态性检验。但由于每组样本量过小，很难直接对每组因变量进行正态性检测，研究者可以

选择合并检验因变量残差的正态性。如果残差不呈正态或近似正态分布，可以转换数据使其残差服从正态或近似正态分布，或使用非参数检验替换单因素方差分析。

谬误2：方差分析方法选择有误。多个因变量可能是相关的，可能更适合应用多元方差分析。在本例中，幼儿园物质条件、师幼互动、健康与安全及幼儿发展间很可能存在相关关系，本例更适合使用多元方差分析。

纠正2：类似本例情况，鉴于方法的简便性，大多数目标研究都应用了单因素方差分析。但是，如果因变量间存在相关，使用单因素方差分析会忽略因变量之间的相关关系，导致犯I型错误的概率变大、检验效率低，即拒绝了实际上成立的、正确的假设。尤其当各个因变量的单因素方差分析结果不一致时，难以对分析结果下总体结论。因此，本例应先检查各因变量（幼儿园物质条件、师幼互动、健康与安全及幼儿发展）之间是否具有显著的相关性，如果相关，推荐采用多元方差分析，如果不相关，则可采用单因素方差分析。

例如，某目标研究采用重复测量方差分析探索学生从小学到大学元认知控制的准确性和一致性是否随着年级升高而逐渐提高。在该研究中，研究者采用2（时间：前、后测）×6（学段：小学一、三、五年级，初中、高中、大学）混合实验设计。样本量为345名不同学段的学生，时间和学段为自变量，重读选择为因变量。在进行重复测量方差分析前未对数据进行任何检验。

谬误1：没有进行一般方差分析的条件检验，更没有检验协方差矩阵的球对称性。重复测量方差分析除了要满足一般方差分析的条件

（即独立性、正态性、方差齐性）外，还需要满足协方差矩阵球对称性。如果协方差矩阵球对称性得不到满足，会增大犯I型错误的概率。

纠正1：在确定数据满足独立性、正态性和方差齐性后，用Mauchly方法检验协方差矩阵的球对称性。如果协方差矩阵的球对称性得不到满足，则需要对与时间有关的F统计量的自由度进行调整，以降低犯I型错误的概

表5 不同区域幼儿园的比较分析结果

变量	地区	样本量	F值	p值
物质条件	东部	10	2.376	0.120
	中部	5		
	西部	7		
师幼互动	东部	10	0.180	0.837
	中部	5		
	西部	7		
健康与安全	东部	10	2.970	0.075
	中部	5		
	西部	7		
幼儿发展	东部	10	0.573	0.573
	中部	5		
	西部	7		

率。一般采用 Greenhouse-Geisser (G-G) 法、Huynh-Feldt (H-F) 法和 Lower-Bound (L-B 下界) 法对自由度进行调整。最后根据原 F 值和调整后的自由度, 判断时间或学段的主效应、时间和学段的交互效应是否显著。

(四) 卡方检验误用的纠正建议

卡方检验在目标文献中的应用也十分广泛, 其使用频率仅次于回归分析、t 检验和方差分析、结构方程模型、因子分析, 主要用于分类资料间的比较。虽然对分类变量的统计常常需要用到卡方检验, 但卡方检验在对分类变量的频数分布分析上并不是万能的, 因此目标文献中也或多或少地出现了误用、乱用卡方检验的情况。

例如, 某目标研究采用卡方检验比较分析不同办园体制下幼儿园的师资素质。其分析结果如表 6 所示。研究者得出“不同办园体制普惠性幼儿园在教师学历、教学年限、职称上均呈现出显著差异, 且普惠性他办园的师资水平在多个维度上优于教办园和普惠性民办园”的结论。

谬误 1: 误用卡方检验分析数据。在本例中, 研究者想要比较不同办园体制普惠性幼儿园在不同师资水平维度上是否有差异, 并判断出孰优孰劣。

然而, 运用卡方检验仅仅能够回答不同办园体制普惠性幼儿园在不同维度师资水平“分布”上的差异是否在统计学上呈现显著状态, 无法判断出“他办园的师资水平在多个维度上优于教办园和普惠性民办园”。

纠正 1: 由于本例中的因变量是等级变量, 因此可以使用非参数检验, 如秩和检验或 Ridit 分析, 来比较不同办园体制普惠性幼儿园在师资水平的不同维度上是否有差异和孰优孰劣。

例如, 某目标研究采用卡方检验考察不同性别和民族大一新生的抑郁状态等级构成的分布是否有差别。其分析结果如表 7 所示。

谬误 1: 忽视卡方检验对样本量的要求。卡方检验的一个重要应用条件是理论频数不能太小。在 $R \times C$ 列联表的卡方检验中, 要求在没有格子的理论频数小于 1 的条件下, 理论频数小于 5 的格子数不超过总格子数的 15。而在本例中, 每个 2×4 列联表中都有两个格子的理论频数小于 5。

纠正 1: 每个 2×4 列联表中都有两个格子的理论频数小于 5 且大于 1, 应该使用 Fisher 确切概率法替换卡方检验。

例如, 某目标研究采用卡方检验探究在就业和

表 6 不同办园体制幼儿园教师资质的比较分析结果

变量	教办园		他办园		普惠性民办园		卡方值	p 值	
	N	%	N	%	N	%			
学历	本科以下	15	28.8	18	36.0	33	62.3	13.296	0.001
	本科及以上	37	71.2	33	64.0	30	37.7		
教学年限	小于 3 年	13	25	8	15.7	25	47.2	29.945	0.000
	3 到 5 年	11	21.2	9	17.6	7	13.2		
	6 到 10 年	13	25	8	15.7	6	11.3		
	大于 10 年	15	28.8	26	51	15	28.3		
职称	无	7	13.5	10	19.6	26	49.1	45.299	0.000
	三级	9	17.3	3	5.9	13	24.5		
	二级	16	30.8	9	17.6	7	13.2		
	一级	17	32.7	16	31.4	5	9.4		
	高级	3	5.7	13	25.5	2	3.8		

表 7 大一新生抑郁情况

		无抑郁或极轻微 N (%)	轻度抑郁 N (%)	中度抑郁 N (%)	重度抑郁 N (%)	卡方值
性别	男	421 (60.93)	206 (29.81)	45 (6.51)	19 (2.75)	20.12***
	女	1074 (51.22)	763 (36.39)	189 (9.01)	71 (3.38)	
民族	汉族	1365 (53.87)	872 (34.41)	211 (8.33)	86 (3.39)	3.77
	少数民族	130 (51.18)	97 (38.19)	23 (9.06)	4 (1.57)	

表8 两种情境下不同调节定向被试的决策偏好差异

	就业N (%)	创业N (%)	χ^2
促进定向	16 (55.17)	13 (44.83)	1.00
预防定向	23 (79.31)	6 (20.69)	33.64***
χ^2	4.30*	8.73**	
	熟悉创业项目N (%)	陌生创业项目N (%)	χ^2
促进定向	18 (62.07)	11 (37.93)	5.76*
预防定向	25 (86.21)	4 (13.79)	51.84***
χ^2	3.89*	11.08***	

创业两种情境下，不同调节定向被试决策偏好的差异。其分析结果如表8所示。进一步对数据卡方检验后得出，预防定向被试的就业选择显著高于促进定向被试的就业选择 ($\chi^2=4.30, p<0.05$)，促进定向被试的创业选择显著高于预防定向被试的创业选择 ($\chi^2=8.73, p<0.01$)，促进定向被试在就业和创业选择中不存在显著差异 ($\chi^2=1.00, p<0.05$)，预防定向被试的就业选择显著高于其他创业选择 ($\chi^2=33.64, p<0.001$)。

谬误1：滥用卡方检验或卡方分布分析数据。不论是用卡方对拟合度进行检验，还是对变量的独立性（即分布的差异）进行检验，都不适用于本例。在本例中，研究者试图比较同一决策情境下不同调节定向的优劣，无法应用卡方检验得出。更严重的问题是，一般情况下，仅靠两组实际频数（理论频数）算不出卡方检验值。因此，本例中的研究者具体应用了什么方法得出的卡方值还有待商榷。

纠正1：根据本例的数据情况，只能得出同一决策情境下不同调节定向的数理频数孰高孰低，无法在统计检验层面上得出有关差异的显著性结论。

（五）探索性因子分析(EFA)误用的纠正建议

探索性因子分析（EFA）主要用来探索观测数

据的基本结构，多应用在量表开发的过程中。EFA在目标文献中的应用也有一定的占比（约5%），且越来越成熟。然而，研究者误用EFA的情况却时有发生，导致其研究结果缺乏可信度、甚至错误。

例如，某目标研究采用EFA对《博士生学术经历量表》进行探索性因子分析，以提炼出构念维度。在本例中，测量工具《博士生学术经历量表》采用4级计分方式，共收集到有效问卷188份，应用SPSS软件分析数据。在探索性因子分析过程中，研究者“使用‘最大方差法’进行正交旋转，项目选取标准为因子负荷大于0.45，因子提取标准为特征值大于1。最终在删除5道题后得到一个包含5个因子的稳定因子结构（KMO=0.82，累积解释变量为72.1%）”。其分析结果如表9所示，该研究并未做出有关该探索性因子分析过程的其他描述。

谬误1：忽略EFA的适用条件。EFA很重要的一个应用前提是要求观测变量之间存在模式化关系，且不存在多重共线性问题。这就需要研究者在分析前，首先获得所有变量的相关矩阵（或协方差矩阵），以检验是否有变量缺少模式化关系，或是否存在问题项导致多重共线性问题，并通过Bartlett球形检验确定变量间的模式化关系，及KMO测度确定是否适合应用EFA。但在本例中，

研究者并没有进行这些分析，而是直接报告EFA分析后的KMO值（KMO测度应当在EFA分析前进行）。

纠正1：在正式分析前，先算出所有变量的相关矩阵（或协方差矩阵）。如果存在变量具有大量低相关系数（ $-0.30 < r < 0.30$ ），则表示这些变量缺少模型化关系，需要进行删除处理。如果相关矩阵中存在 $r > 0.90$ 或 $r < -0.90$ 的情况，说明数据可能存在

表9 因子提取结果

题目	课堂讨论	项目参与	前沿涉猎	跨学科学习	导师交流
1	0.89				
2	0.89				
3	0.86				
4		0.84			
5		0.84			
6		0.75			
7			0.80		
8			0.79		
9			0.73		
10				0.83	
11				0.76	
12				0.74	
13					0.79
14					0.75
15					0.61

共线性问题，需要使用Haitovsky检测是否存在多重共线性问题。如果存在多重共线性问题，则需要确定导致多重共线性问题的变量并将其删除。然后，通过Bartlett球形检验来确定变量（剩余变量）间确实存在模式化关系。最后，通过KMO测度进一步确定EFA是否可应用于本例。如果KMO值大于0.50，则表示EFA可适用于本例。如果KMO值小于0.50，则意味着数据无法产生显著而可靠的因素，EFA不适用于本例。

谬误2：忽略因子提取方法的选择。在SPSS中有七种因子提取方法供选择，包括“主成份法”“未加权最小二乘法（ULS）”“广义最小二乘法（GLS）”“最大似然法（ML）”“主轴因子提取法”“ α 因子提取法”和“映像因子提取法”，且这些因子提取方法提供的结果稍有不同。需要注意的是，“主成份法”不是一种探索性因素分析方法，而是一种减少拒绝测量误差维度的方法，一般用于主成分分析（PCA）中，不应用于EFA分析中。研究者需要根据研究目的、数据的基本情况及对采用拟合优度指数的兴趣，选择合适的因子提取方法。在本例中，研究者没有交代使用了具体哪种方法来提取因子。

纠正2：在该例中，测量工具《博士生学术经历量表》采用4级计分方式，属于类别变量（5点或以上的李克特量表的变量才可当作等距变量处理）。ULS或加权最小二乘均值和方差调整（WLSMV）都可用在基于类别变量的EFA分析中，且WLSMV是更优的选择（Muthén, du Toit, & Spisic, 1997）。由于SPSS中没有WLSMV方法的设置，因此，在SPSS中，对于类别变量，建议使用ULS方法。如果研究者应用Mplus作分析，则最好使用WLSMV方法。

谬误3：忽略选择正交旋转的前提条件。正交旋转的前提条件是因子间不相关，而斜交旋转则允许因子间存在相关。在本例中，研究者并未给出因子之间的相关关系，因而无法判断因子之间是否存在相关。如果“项目参与”与“前沿涉猎”“跨学科学习”“导师交流”之间存在相关，那么进行正交旋转会忽视这些因子间的相关，对分析结果造成影响。

纠正3：首先检查一种常用的斜交旋转方法产

生的解。斜交旋转后，会得到三个基本结果，即因子模式矩阵、因子结构矩阵和因子相关性矩阵。如果因子相关性矩阵中因子间的相关较低，那么可以进行最大方差法正交旋转。如果发现其中一些因子相互关联，那么应当采用斜交旋转。需要注意的是，当采用斜交旋转后，若发现因子间的相关较高（ $r > 0.60$ ），那么因子模式矩阵和因子结构矩阵可能会有显著差异。在这种情况下，建议报告两种矩阵结果，或者明确说明报告的载荷是因子模式系数还是因子结构系数。

另外，在本例中还存在着其他缺陷。本例没有交代是否具有或如何处理极端值（包括缺失数据）。也没有交代清楚删除5个观测变量的原因，是因子载荷小于0.45或是有显著交叉载荷的存在，应明确说明的信息没有完整提供，会降低研究结果的说服力。

（六）结构方程模型（SEM）误用的纠正建议

结构方程模型，包括验证性因子分析（CFA），是一种基于变量的协方差矩阵来分析变量间关系的多元统计方法。近五年在目标文献中的应用逐年增多，2015年占比约8.5%到2019年占比约12.3%。但是，部分研究者由于对SEM的基本原理理解不够透彻而误用了该统计方法，导致结论出错。

例如，某目标研究采用SEM分析教师满意度、教育期望、政府保障和学校支持之间的关系。在该研究中，样本量为8150（并未对数据进行检验），测量工具采用7点量表，估计方法选用极大似然（ML）估计，分析软件是AMOS17.0，初始模型建立如图4。

初始分析得到27个题目的路径系数均显著。总体来看，模型拟合度较好，但是部分拟合指数未达基本要求。根据模型修改建议，研究者通过增加[e12~e13][e15~e16][e16~e17][e6~e7]这些M.I.值大于20的残差路径，最终得到拟合度更好的模型。模型修正前后拟合度情况如表10所示。

谬误1：忽略多元正态分布的前提条件。为获取准确的参数估计值和稳定可靠的分析结果，SEM要求样本量足够大且观测变量应当服从多元正态分布。如果数据违背多元正态分布假设，那么采用ML估计方法就会增大犯I型错误的概率，更可能拒绝实际上建构良好的模型。在本例中，虽然样本量足

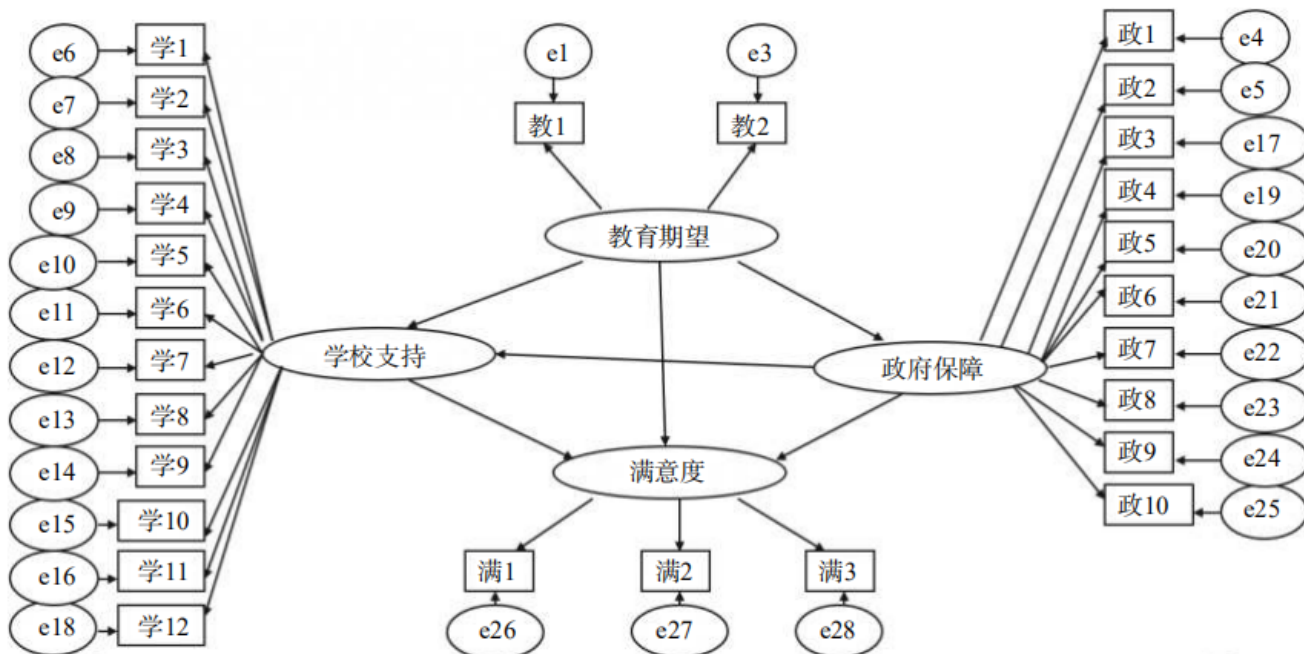


图4 初始模型

表 10 模型修正前后拟合度情况

	χ^2	df	χ^2/df	RMSEA	NFI	CFI	GFI	AGFI	AIC
初始模型	11798	246	47.959	0.067	0.914	0.915	0.871	0.843	11906.01
修正模型	8999.2	243	35.502	0.061	0.937	0.939	0.906	0.884	8707.477

够大，但仍不能保证所有观测变量服从多元正态分布，研究者未作数据分布检验是不恰当的。

纠正1：在本例中，7点量表数据可以当作是连续数据处理。在进行SEM分析前，应首先对数据进行多元正态性分布检验，诸如R语言、SAS软件、Python语言、STATA软件均提供了多元正态性分布检验。如果数据不符合多元正态性可使用任意分布估计方法、或用S-B调整法调整基于正态理论的模型检验统计量和参数标准误、或通过数据转换使数据分布正态化。但是数据转换后再分析得到的因子载荷不再是原观测变量的数据载荷，对估计参数的解释应当按照新的测度进行。

谬误2：模型存在局部过饱和问题。模型的局部过饱和现象是指模型局部的自由度为负值。当模型局部存在过饱和现象时，模型局部的参数估计随意性大且不可信，却往往能在模型的总体拟合指数上得出不错的结果。而在本例中，潜变量“教育期望”下只有两个观测变量，其局部自由度为-1（3-4=-1），存在模型局部过饱和问题。

纠正2：为了保证SEM分析结果的高可靠性，研究者需要对原始模型进行修改，即增加潜变量“教育期望”下面的观测变量数目，使其不少于3个。

谬误3：修正模型存在较严重的测量误差相互关联问题。在本例中，作者通过增加[e12~e13][e15~e16][e16~e17][e6~e7]这些M. I. 值大于20的残差路径，得到拟合度更好的模型。但是，这种允许测量误差在事后关联的理由是不合理的。出现误差相关联的情况实际上是由于有误差相关联的潜变量中仍然存在潜在的分离，即观测变量还有遗漏，或相关的观测变量存在问题。况且[e16~e17]是两个潜变量“学校支持”和“政府保障”下观测变量测量误差的关联，更是不能出现在模型中的。因此，出现较严重的测量误差相互关联的情况说明初始模型的结构存在问题，初始模型成立理论不充分，需要增加或减少观测变量。

纠正3：[e12~e13][e15~e16][e6~e7]这些测量误差存在相互关联说明初始的假设模型中遗漏了观测变量，研究者可以根据理论分离这些观测变量中的部分变量（即补充遗漏观测变量），或者分别删掉e12和e13、e15和e16、e6和e7中因子载荷较小的那个观测变量。[e16~e17]属于不同潜变量中的观测变量的误差相关联，研究者应当根据理论重新检查初始模型，决定是删掉目标观测变量或是修正相关观测变量的题目。依据修正指数，在该例

中，“学11”这个观测变量可能存在较大问题。不论如何修正初始模型，建立新的模型后，建议研究者用新的独立数据测试新模型。

需要注意的是，SEM只是一种检验理论的工具，通常在没有充分理论支持的情况下，不建议研究者轻易地遵照修正指数调整初始模型。此外，因为该例的样本量足够大，而卡方值容易受样本量的影响，即当样本量足够大时卡方值也会较大，所以卡方值在本例中对判断模型的拟合优度没有太大的意义。

五、定量研究方法在我国教育实证研究中的应用展望

由于定量研究方法起源于自然科学领域，那么其在教育实证研究中的应用除了受到教育学本身发展的影响，也会受到数学、计算机科学等学科发展的影响，将会呈现出跨学科、多元化的发展趋势。以下就从教育科学的发展角度，结合自然科学、技术的发展趋势及国内诸多学者的分析，对定量研究方法在我国教育实证研究领域中的应用进行初步的展望。

（一）从五年发展特点看趋势

1. 定量研究方法将更广泛地被应用。从对五年来定量方法的应用特点的分析可知，定量方法的应用率虽然是在逐年稳步地增加，但是总占比还是很低，未来还有很大的上升空间。因此，随着“教育实证论坛”的继续召开，有关定量方法的普及会继续扩大，必然会有更多教育研究会使用定量方法。

2. 机构间的合作将更频繁和深入。从对五年来作者合作网络的分析可知，我国教育科学领域研究团队之间的合作交流还有待提高。由于本身团队间的合作基数小、合作网络较疏散，所以只要各大研究机构鼓励、重视团队间的合作，未来学者间合作的比例会有较大上升。并且随着研究深入，核心作者间的合作也会逐渐增多，形式也会趋于长久、稳定，团队间的合作网络也会更牢固。

3. 定量研究方法的应用将更成熟。从对五年来定量研究方法应用特点的分析可知，新方法、新技术的引进速度慢，且方法误用屡见不鲜。而“教育实证论坛”的召开和中国教育实证研究“这五年专刊”的举动意味着我国各高校教育学院和教育科研院所已经开始重视普及量化知识，未来也会更加重

视吸收、培养专门从事定量研究的人才。学者们一旦对定量方法的应用重视起来，新方法、新技术的引进和研发速度也会增快，方法的误用也会有所改善，其应用也将会更成熟。

（二）从教育科学的发展看趋势

1. 定量研究方法趋向跨学科交叉应用。教育现象本身涵盖的内容会涉及经济、生理、心理等多个学科领域，如有条件的现金转移支付承诺可以改善我国农村贫困学生高中完成情况（易红梅，何婧，张林秀，2019）。教育科学研究要有效地揭示教育现象的本质、解决教育实践中的问题，就需要在跨学科的基础上进行研究与合作。因此，针对教育现象的跨学科研究是教育科学发展与进步的必然趋势。而定量方法在各个学科领域中又都是相通的，它的应用是属于交叉学科范畴。只要研究主题和数据符合要求，一种定量方法可以应用在不同学科中，如回归分析可以应用在研究心理现象上，也可以应用在研究影响经济效应的因素上。在美国，不论自然科学界还是社会科学界对定量研究方法的应用已经相当普遍，并且定量研究方法已经成为跨学科研究的重要工具和媒介。由于我国教育科学领域目前正处在提升研究深度和广度、鼓励学科融合、吸收交叉人才的阶段，那么随着跨学科研究的逐渐增多，定量方法也会被更加广泛地应用在跨学科和交叉学科的教育实证研究中。

2. 定量研究方法趋向多元化、混合化应用。随着教育科学的发展，单一的定量方法很难处理较复杂的教育研究问题，也不容易得到准确客观的研究结论。为了更好地研究教育问题，研究者往往需要多元化应用定量研究方法，而计算机科学技术、统计学、心理学等学科的飞速进步又为定量研究方法的多元化应用提供了工具、方法论和研究范式。如李宪印等人（2019）在研究大学生创新行为的构成因素时综合运用了项目分析、探索性因子分析、验证性因子分析、信度检验和方差分析。更进一步，为适应社会科学发展的需要、深入具体地分析问题，将质性和定量方法相结合的混合研究也逐渐在国际学术界流行起来。比如Lv et al.（2020）在研究正念对抑郁的干预效果的研究中就混合了元分析和叙述性综述两类方法。“全国首届教育实证研究论坛”的召开标志着我国教育科学将进入一个新

的历史发展阶段，教育研究将更加关注实证，也更趋向于科学、规范和多元化，教育研究方法的应用也将会更加多元且有混合化趋向。

（三）从自然科学、技术的发展看趋势

1. 定量研究方法趋向大样本化、大数据化应用。云计算、互联网和物联网技术的发展使得社会信息化的程度不断加深，研究者可以利用网络挖掘、大型调查数据库等方式获得大量数据。大样本的使用，既能提高定量分析结果的准确度，又能减少重复工作、节约科研成本，已经被越来越多的学者所青睐，也成为定量研究方法应用的一大趋势。另一方面，随着大规模并行处理数据库、数据挖掘、云计算平台、分布式文件系统等技术的出现，使得通过对全样本进行定量分析从而得到相对理想的分析结果成为可能。同时，我国又将大数据的研究和发展视为新的国家战略之一（何哲，2015），使得教育大数据的应用成为必然趋势。因此，定量方法会越来越地被用来辅助大数据技术分析教育大数据。需要注意的是，虽然大数据弥补了传统数据的很多缺陷，但是也有其局限性。如由大数据方法获得的总体本身往往存在偏差，会导致分析产生系统性偏误等，因此大数据分析也不能完全取代抽样分析。

2. 定量研究方法趋向“AI+教育”的应用。伴随着人工智能（AI）的不断发展，其在教育领域的应用也愈加广泛，如伴读机器人走入课堂、智能阅卷开始出现等。在2017年国务院印发的《新一代人工智能发展规划》中明确提出，“实施全民智能教育项目”（国务院，2017），全方位地实现“AI+教育”成为教育领域的理想和目标（吴站杰，秦健，2003）。而将定量方法应用到AI领域不仅能推

动“AI+教育”的发展，还能促进新的定量方法的产生。如研究者可以运用深度学习设计并训练出多款辅助教学机器人，将其分给具有相同学习基础和条件的学生使用，然后收集相关数据，通过定量分析找出辅助学习效果最佳的机器人并投入教育市场。再如，分布式计算和AI技术的发展使得基于Agent的建模方法应运而生，并越来越多地应用在社会科学研究领域，通过模拟个体或团体的行为和交互来评估其对整个系统的影响。目前，我国的“AI+教育”尚处在发展的初始阶段（余胜泉，2018，第106页），在政策和市场的影响下，有关“AI+教育”的研究会迅速增多，定量方法被广泛应用到“AI+教育”研究中也是必然趋势。

六、本研究的局限性

由于受到研究对象、研究方法、研究工具的制约，本研究存在一定的局限性。一方面，本研究的11本期刊并不能全完无偏的代表所有的中文期刊，定量研究方法在某些非核心期刊的应用情况可能更糟，或者在一些专业性较强的期刊的应用情况可能更好。另一方面，受专业性和评价体系的影响，某些学者会更倾向于把文章发表在专业性期刊上，或者一部分中国学者会把质量较高的文章发表在SSCI期刊上。这样有些方法的误用情况就可能会被忽略；同样，有些应用了新方法、新技术的文章也会被忽略。因此，后续研究者可以尝试把研究样本扩大到这些期刊范围，进行更加深入的分析。

（吕 晶，华东师范大学高峰博士后，上海200241）

（原文刊载于《华东师范大学学报（教科版）》2020年第9期）

在教育研究中运用量化研究方法的问题与反思

蔡红红

社会科学的量化实证取向源自19世纪中叶以孔德为代表的实证主义，最初主要受自然科学领域的启发。20世纪二三十年代，美国社会学界一方面为了吸引学生，满足学生对实用性、工具性知识的需求，另一方面为迎合政府和财团法人对现实问题调查研究的资助偏好，推行“实用为本”的改革。与此同时，受自然科学领域“客观、科学”的数学化研究方法和沃森（Watson J B）行为主义（Behaviorism）的影响，统计方法逐渐在社会学研究中取得主导地位。由此，美国社会科学研究走向量化与反规范性论述的“科学”方向。而教育研究以社会科学的量化研究传统为依据，20世纪初以来，美国的主流教育研究也往往带有定量性质。

一、教育研究中的量化研究方法

教育研究方法是人们在研究教育问题时所采取的步骤、手段和方法的总称，它是决定教育研究质量的关键因素。近年来，我国教育研究领域出现了一些推崇实证研究方法的声音，以统计方法为主的量化研究是其重要组成部分，它通过实验、调查、测验、结构观察及已有的数量化资料，对教育现象进行客观分析，并将所得结果作相应的统计推断，使研究结论具有普遍适应性。研究方法决定教育学的科学性质。

尽管伦德伯格（Lundberg G A）意识到社会现象与自然现象有所差异，如人有态度、动机、意向等，但他依旧坚定地认为，问题的关键在于如何处置这些差异。他认为，诸如渴望、希望、恐惧、价值、目的、意图等的心灵状态均可被客观地观察和记录，即“自然科学”化。对于社会科学的发展而言，研究方法的“自然科学化”成为其“科学性质”的重要体现，同时加强了社会科学自身的合法性地位。在崇尚实证研究的研究者意识里，通过将教育现象或研究对象的态度、看法与意见等转化为量化的客观材料与数据进行研究，至少抽离了人的个体性和主观色彩，避免陷入“自说自话”的窠臼，得出的结论更客观、可靠，具有较高的可信度。这无疑也更符合科学研究中价值无涉的标准。此外，在日常生活中，

统计方法在经济消费取向、文化观念和政治意向等方面的运用屡见不鲜，人们也已经习以为常。因此，运用量化方法进行教育研究，有着广泛的社会现实基础。但运用此方法研究教育问题的适切性和具体操作过程中存在的问题着实引人深思。笔者将以教育研究中最常见、运用最频繁的量化研究方式——通过对研究对象进行问卷调查，并将所得的资料与数据进行统计分析，进而验证或推断结论，这一研究形式为例，阐述其在具体操作过程中存在的争议与问题。

二、在教育研究中运用量化研究方法面临的问题

马克思曾说：“一种科学只有在成功地运用数学时，才算达到真正完善的地步。”“数学化”的研究方式一度促进了社会科学的发展，但运用该方法进行研究，不可避免地带有鲜明的精确化和机械的“测量”印记。教育现象是一种复杂、独特、富有价值色彩的社会现象，其运行、发展与变化都有与其他现象不同的特点和规律：教育研究的对象，是具有自组织能力的人及其在教育过程中的多层次、多维度的关系，非任何自然现象和生物现象所能比拟。然而，当前对“数学化”、以统计分析为重点的量化研究方法的推崇容易使人们忽略了该方法用于探究特定教育问题的适切性及方法本身的缺陷。不假思索地套用自然科学精细化的数学模式，很可能导致研究结论与教育现实并不相符，不但无法揭示客观的教育规律，也贬损了教育研究的科学性。在教育研究中运用量化研究方法面临着如下诸多问题。

（一）量化研究存在简单的“数学化”取向

物理测量数据的每个数字都有其确定的位置与意义。在教育研究中，量化研究者经常将不可观察与直接测量的被研究者主观态度或意见等概念操作化为测量量表，并赋予相应数字以程度高低之含义。以李克特五分态度量表为例，“1至5”这之间的5个整数分别代表被研究者“非常不同意—非常同意”的态度。这些数字没有绝对零点和单位，两个相邻数值之间的数字如4.5，它介于同意与非常同意之间，但在被调查者的认知

中，这个数值是未被赋予意义的，且无法说明被调查者赞成4.6即代表同意程度高于4.5。因此，研究者擅自赋予除五分量表中五个整数之外的数值以内涵，其与被研究者本人的初衷可能不符。拜尔考（Berka K）指出，在大多数情况下，心理或社会测量是一种较精致的分类，或者说，它是一种理念的或概念的方法，具有相对参考性，与实际意义上的“测量”还存在一定差距。

社会学家邓肯（Duncan O D）也注意到了社会学研究中态度量表可能面临的基本问题，他认为，在“非常不同意—非常同意”的数轴上，每个人都有他独特的位置，只有他自己才可能恰当地指出这个位置在哪里。人们对数轴上每个态度的判断标准是有差异的。如有人对事物的态度比较宽容，有人则比较严格，前者“非常同意”的选择标准可能相当于后者“同意”的判定尺度。在分析时，研究者一律将在李克特五分量表中回答“非常同意”的人赋值5分，并直接将其作为等距且等比的物理测量数据进行计算，如此便能将他们作为类别或等级变量时获得更多信息，且能用更高级的统计分析方法来处理数据，以显示分析的复杂性与研究的深刻性。这样的做法体现了当前教育领域的量化研究对精确的追求，带有简单化、易操作倾向，却忽略了问题的真正焦点：能否如实呈现被研究者的态度与想法。

（二）量化研究具有表层化倾向

基于问卷调查的量化研究，其价值与深刻性时常遭到质疑。弗莱克斯纳（Flexner A）指出，调查问卷在花了大量的时间和精力之后被整理成各种图表，但最后得出的结论却要么是凭调查开始前的常识就能直接判断的，要么是最终得不到可靠证据支持的。他对问卷调查进行了强烈批判，称其只是一种廉价、方便和快速获取信息资料或非信息资料的方法，问卷填答的随意性也令其不具备科学特征。他强调，不管多么巧妙地收集资料与信息，报告和检查等都不构成研究。尽管许多研究者指出，量化研究是通过将数据和资料的统计分析以探求各个研究变量背后的因果关系。但当前教育领域许多的“量化文章”，甚至相当部分教育学领域的学位论文，更类似于调研报告，而不是科学研究文本。

在教育研究中，量化研究往往只提供了一种呈现教育现象的方式，它能向我们展示“实然”

的状态，却无法指引我们“应然”的方向。如我们可以通过问卷调查，知晓大学当前及以往的所作所为，以及这如何影响社会；或根据调查，揭示政府、教师或学生对大学行为的态度与期待。但这都不能回答“大学应该扮演什么样的角色”这个问题。无论量化研究结果如何，我们都可以合理发问：这是大学应该做的吗？在收集了所有的调研数据，并进行统计分析后，这个问题仍然存在。无怪乎有研究者指出，量化其实只是一种近乎没做什么决定的决定。

与此同时，量化的数字将人们的认知限缩在表层化的维度内，量化研究者时常容易忘记在教育研究中这些数字代表的概念与内涵，陷入纯粹计算数值间各种关系的狂热之中。对于这一现象，孔德曾指出：在实证体系下，基于盲目联系的本能作用，我们热衷于为同时存在或相继出现的现象建立联系，但对外部世界的合理探索和考察却证明，世界的连带关系比我们所设想或希望的松弛得多，许多这类关联在现实生活中纯属虚妄。另一方面，量化研究者根据数据或统计结果“看图表说话”，这一模式促使其习惯于将重心置于数据或数量关系的解释，忽视了对教育现象自身更深层次内涵的关切，也不利于个人创新思维的发展。正如弗莱克斯纳所说，无休无止的计算绝不会产生理论、原理或思想。工具实证主义倾向导致量化研究者选择性地关注当前研究工具所能解决的范围内的的问题，之所以使用量化，其实是因为看到了该问题可运用与自然科学类似的测量指标进行测算与分析。在教育研究领域，这表现为部分定量研究者越来越倾向于研究能够直接用测量量表（特别是西方广泛使用的、较成熟的量表）进行定量分析的教育问题。就如卡西尔（Cassirer E）所说的方法决定论，方法本身决定了“科学性”，因而也决定了“事实性”和“真理性”。在当前社会科学崇尚实证研究的环境下，这体现了一种片面追求量化，研究方法决定研究问题的不良倾向。

（三）量化研究的信度难以保证

数量化资料的可靠性是量化研究具备一定信度的基础。通过问卷调查获取调研数据的方式，因其高效、便捷和易量化等特点，在社会学、心理学和教育学等社会科学领域被广泛使用。在教育研究中，问卷类型主要以自填式调查问卷为

主，这意味着被研究者能否根据自身的现实情况，如实、准确地填答问题是数据真实可靠与否的关键。但是，有些学者指出，中国是一个低信任度的国家，中国人缺乏以共同的信任和制度保障为基础的对一般人的“普通信任”。在问卷填答过程中，低信任度的人际关系可能导致被调查者不愿意填写问卷；不完全相信调查的匿名性，戒备心较强，未依据自己的真实情况或态度回答问题；对一些通过行政手段收发的问卷，鉴于行政压力与社会称许性的影响，被研究者在填答问卷时倾向于美化组织或自身形象。有研究进一步指出，即使在匿名回答的基础上，采用投射、情境故事等技术性手段，中国被试仍对问卷调查保持较高的防御性。这都导致问卷调查所回收的数据可能并不反映被研究者的真实情况与想法。

除了天然存在的人际信任问题，教育研究对象的特质与调查工具的设计也深刻影响着调研数据的信度。在教育研究中，调研对象经常包含多维度的人的态度、情绪与意见等具有强烈主观特质的概念尺度。人的多变性、不同的问卷填答情境，往往也意味着这些主观概念的测量结果可能并不稳定。例如，同一位被调查者填答两次相同的问卷可能会得到不同的结果；是否有人监督、问卷发放者的身份、以及调查是否涉及自身利益等，这些因素也显著影响着被调查者的问卷填答情况。而我们无法通过观察等客观手段对问卷的信度进行监测，以确保收集到他们真实的想法。此外，问卷设计的合理性也极大影响了可靠数据的回收。一些研究者希望通过一次问卷调查尽可能多地收集信息，在设计问卷时，未考虑问卷篇幅的科学性，也未设身处地地感受被调查者填答问卷时的心理状态，经常出现一份问卷包含上百甚至几百个问题的情况。通过这样的问卷收集而来的数据，其信度是存疑的。

一项量化研究的论证依据与结论建立在调研数据的分析之上，常有学者用“Garbage in, garbage out!”来形容数据品质对量化研究的重要性，即如果你收集回来的数据质量（指数据的信度、效度）很低，则不论你用何种分析方法，得出的研究结论都将无法令人信服。因此，能否收集到人们的真实信息与想法，并合理地将其转化为可信和有效的数据，是进行规范、科学的量化研究的基础，也是得到可靠结论的重要保证。

（四）量化研究者容易陷入数据的泥沼

学术发表在很大程度上象征着一位学者的科研能力与声誉，且与其职称评定、晋升、薪酬待遇等一系列现实问题有着千丝万缕的联系。在教育领域推崇实证研究的当下，逐渐有些研究者，特别是一些研究新手开始注意到基于数据（尤其是大样本数据）撰写的文章在学术发表中的优势：量化文章在部分教育期刊中的比例逐渐增加，有了数据的论证，论文可能更易发表；相比哲学思辨而言，运用基于数据的量化分析方式撰写文章的门槛较低；量化研究更关注教育领域中的现实问题，很多研究成果被采用为政府决策的依据，这为研究成果和研究者带来更高的社会关注度和认可度。这些特点使得部分研究者对数据的收集与运用趋之若鹜，也因此时常容易陷入数据的泥潭：倾向于探究能够用数据分析和回应的教育问题，对于教育领域中不可操作化的理论或伦理问题不感兴趣，也束手无策；将数据作为“原材料”，企图通过“概念-属性-变项-测量-因果关系-统计”这一套特定程序，解释其所研究的所有教育问题；只看到数值及其之间关系的测算与内涵，将教育研究对象抽离出具体的历史-文化情境，使其成为“没有历史”的人与物。

另一方面，量化研究者希望从数据出发，让数据为现实发声，但这需要建立在对数据科学严谨和实事求是的处理、分析与解释之上。然而，在测量、探索变量关系等环节，部分研究者的量化操作却并未严格遵守统计要求或学科研究规范。他们忽视数据特点，简单套用各种统计方法，极度关注统计结果的“价值性”；操纵甚至篡改原始数据，直到获得各项指标良好的模型和符合其预期假设的统计结果。这样“数据至上”，却又随意对待数据的行为，是一种典型的为了发表而研究，缺乏信仰和思想，没有精神、没有灵魂、没有价值追求的“投机行为”。

（五）量化研究者热衷于使用“高级”统计方法

由于数理概率理论提供统计学科学的基础，以致统计学连带地提供了社会学量化研究以科学的基础。依据此认知模式，数理统计得以被“正当化”为科学。在自然科学研究中，越高级的技术越有利于揭示复杂、未知的科学问题，以此类推，数理统计作为量化研究在教育领域中的分析

工具，这是否意味着，越繁复和高阶的统计方法就越有利于分析深奥的教育问题，或是将问题分析得愈深刻呢？但事实似乎并不是这样。部分教育研究虽然运用了回归分析、结构方程模型等看似“高级”的统计方法，但其要处理或说明的问题，其实使用较为简单的交叉表分析或差异分析等“普通”方法就能解决，因为基础与高阶的统计方法之间往往存在着部分相似的功能。但当前许多量化研究者却乐于“化简为繁”，倾向于使用复杂而不是简单易懂的统计方法来分析问题，以使研究过程看起来更有深度，结论更具“科学性”。然而，方法的“精进”并无助于厘清事实或揭示更深刻的道理。在教育研究中，同时运用基础或高阶的统计方法，得出的结论别无二致的情况时有发生。“给小孩一根铁锤，他会发现，任何他看到的東西都需要给它一锤。”科塞（Cosser L A）用这个有关工具的法則諷刺20世纪70年代那些热衷于使用当时被视为“先进”的结构方程模型等量化社会学家的做法与想法——以为严谨而复杂的方法即是具备了刚性科学的特质，因此可以克服理论上的薄弱。

仔细阅读教育类量化文章会发现，研究者使用“高级”统计方法，通常建立在将量表“数学化”操作的基础上，它意味着数轴上的每个数字都具有内涵，且每位选择同一数值的被研究者的态度都是一样的，但这样精细的“数学化”操作可能并不恰当。这一量化研究的客观性与教育研究的主观性冲突，笔者已在上文进行了详细阐述。在数值意义尚未明晰的前提下，“精深”的统计方法或精巧的计量模型无法增加研究的价值。另一方面，“高级”统计方法是易复制的，片面追求“高级”统计方法的技术主义倾向无助于学科知识的积累，也背离了“研究方法始终应为研究问题服务”的初衷。

三、结论与反思

在教育领域推崇实证研究的当下，定量研究由于其数学化的“科学”特性而备受关注。但教育研究的主观性与量化研究的客观性之间的冲突、量化研究的表层化倾向、调研数据的难以确信、部分研究者对“数据”和“高级”统计方法的偏执，这些都表明：将定量方法运用于教育研究领域应当是谨慎的。所以，我们应该始终明晰对教育研究怀抱着的期待。为了把社会学推进以

物理学为典范的“科学”殿堂而一味向自然科学的认知模式倾斜，将只是一种东施效颦的作为，更是一种自我矮化与异化的行止。鉴于教育研究情境和研究对象的复杂性，我们不能迷信问卷调查和统计方法。但不可否认的是，他们确实是教育研究方法体系中非常重要的数据采集与分析工具，服务于教育研究目的，有助于提示教育活动或现象的特点和规律。所以，我们也不能因噎废食，应在深刻意识到量化研究方法局限性的同时，更慎重地使用它。

量化研究中的教育测量需要还原被调查者最真实的想法与状态，这意味着研究者要忠实于被调查者的选择和测量结果的宽松尺度，对数据的解读保持谨慎的态度。如将量表的数值作为有序变量进行操作与分析，而不是一味追求精确的数字化表达，也许这样更有利于探寻教育行为和现象背后的现实意义与内涵。此外，对量化研究浅层化的诟病一定程度上源于部分教育研究者对量化研究方法的不当使用，且量化研究的数据收集、处理与分析过程绝大部分在私下进行，往往不为人所知，这导致人们无法监督和检测其分析过程和研究结论的科学性。因此，我们应掌握扎实的问卷设计、数据处理与统计分析方法，严格遵守量化研究程序，对数据与客观事实始终保持敬畏之心，科学地操作数据，谨慎地解释结论，以此提高教育研究中量化研究方法的價值。同时加大对量化研究中学术不端行为的伦理审查与惩戒力度，遏制“为了发表而发表”的形式主义。

最后，教育研究的成功取决于研究人员在遵守基本研究原则的背景下，对其研究实践的不断反思。所以，最重要的是，在量化操作的过程中，我们不能只关注数据和计算而牺牲了思考，漠视数字背后“人”的内涵。量化研究方法的價值与局限性也提示我们，作为一名教育研究者，我们不仅要具备较完善的量化研究能力，还需重视对哲学思辨能力的培养，对于它的重要性，孔德早就做出了预判：如果缺乏某种既定的思辨观念作一贯的指引，那么人的才智就绝不可能组织、甚至不可能收集必不可缺的材料。

（蔡红红，华东师范大学高等教育研究所博士研究生，上海 200062）

（原文刊载于《中国高教研究》2020年第9期）