

·专题：研究方法·

中国管理研究中问卷调查法的取样与测量合适性：评估与建议

□刘 洋 谢 丽

[华南理工大学 广州 510640]

[摘 要] 问卷法作为中国管理研究中最普遍的研究方法近来受到较多质疑。问卷调研法的“严谨性”本身没有问题，而是由于部分学者在使用这一方法过程中的不严谨性（特别是在取样和测量方面），导致学者对此方法产生了一定的误解。基于此，针对近十年发表在《管理世界》上的137篇采用问卷调研法的演绎性研究，对其取样和测量合适性进行评估，提出了八个常见的问题，并以一篇范文为例，提出了对应的建议，以期采用问卷调研法的管理研究提供一定的借鉴。

[关键词] 问卷调研法；取样；测量；中国管理研究

[中图分类号] F270

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2017)02-0024-08

引言

定量研究方法在中国管理学研究逐步走向正规过程中扮演了重要角色，问卷调查法由于其成本低廉、信效度较高、快速有效收集数据等优点，成为中国管理学定量研究中最为普及、最为成熟的方法之一^[1]。然而，纵观近十年采用问卷调研法的中国管理学研究可以发现，尽管其为构建和检验中国管理理论做出了重要贡献，但在采用这种方法过程中的严谨性(Rigor)仍有很多不足。事实上，不论构建何种理论，“严谨”是根基：逻辑和数据间的严格匹配做不到，信度和效度不高，那么研究的发现和构建的理论都需谨慎对待。

基于这一出发点，本文将对对中国管理研究代表性期刊《管理世界》中，近十年采用问卷调研法为主要研究方法的论文进行评估，特别是对这些论文中取样和测量的合适性这两个被忽略最多的角度进行比对，试图去发现中国管理研究中采用问卷调研法进行研究常见的问题，并提出相应的改进建议。通过对137篇文献的评估，本文发现，中国部分管理研究学者在问卷调研法的取样和测量过程中主要

存在以下八个方面的问题：普遍缺乏对抽样总体的明确界定；普遍未能对抽样过程进行详细描述；普遍缺乏对样本估计偏差评估、无回复偏差的评估；普遍缺乏对被调查对象与测量单元之间关系的评估；重点关注内部一致性，而忽略了其他提升信度的策略；对内容效度的评估缺乏；聚合效度与区分效度的做法存在的问题；普遍缺乏对共同方法偏差的关注。最后，提出了一些解决策略，以期采用问卷调研法的管理研究提供一定的借鉴。

一、取样与测量合适性的评估标准

问卷调研法是指通过抽样收集能够代表总体的数据，得出结论，在允许随机误差的界限下，推断和概括总体特征的研究方法^[2]。管理研究中，问卷调查能够有效地基于任何目地对组织利益相关者（股东、管理层、客户、雇员）的态度进行评估^[3]。问卷法的质量取决于样本设计和其调查步骤的合适性，其中抽样和测量两个部分至关重要^[4]。关于问卷调研法的具体过程和注意事项本文不在赘述，可参考相关教科书（例如，罗胜强和姜嫄^[5]）。本部分将重点阐述抽样和测量合适性的评估标准。

[收稿日期] 2017-01-05

[基金项目] 国家自然科学基金(71502065, 71502160)；广东省教育厅特色创新项目(教育科研类)(2014GXJK006)；广东省软科学项目(2016A070706002)；华南理工大学工商管理学院研究生重点课程建设项目(ykc201601)。

[作者简介] 刘洋(1987-)男，博士，华南理工大学工商管理学院副教授；谢丽(1992-)女，华南理工大学工商管理学院硕士研究生。

(一) 取样合适性的评估标准: 样本能否代表总体

抽样的目的在于对目标总体进行部分抽样得出结论反推总体特征, 抽样的过程重点关注选取样本是否能够代表研究对象总体^[6], 具体涉及目标总体及抽样总体、分析单元、样本量、抽样过程、回收及无回复情况等内容(图1)。评估问卷调研中取样合适性的六大标准是: (1) 清晰确立特定的目标总体和抽样总体; (2) 根据研究问题准确阐述分析单元; (3) 确定合理的样本量并且阐述其依据; (4) 对于取样过程进行充分描述; (5) 对于回收率和无回复率的情况描述; (6) 根据所采用的抽样策略清晰阐述合理的估计和数据分析过程, 包括无回复偏差的解决方案及对于发现的概括^[6-7]。

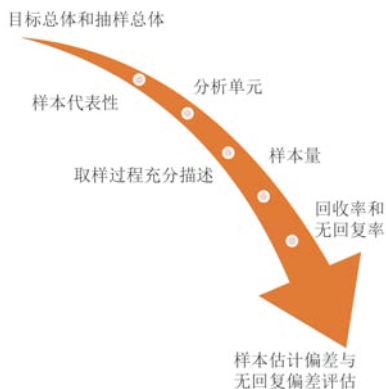


图1 取样合适性评估标准

第一, 对目标总体和抽样总体的明确界定。在问卷调查研究中, 目标总体指的是研究者感兴趣并且尝试通过抽样获得信息进行研究的部分。在问卷调研中, 第一项标准要求总体通常是限定的总体, 这个总体是明确表示的, 数量有限的, 它可以是个人或者集合, 如家庭、组织、企业等^[7]。而样本总体指从总体中提取的一个子集并且参与实际研究的群体, 因为收集所有总体的数据通常是不可能或不实际的。然而, 在某些情况下, 由于资源限制和可行性约束, 某些目标总体不得不予以排除, 这意味着需要总体效度来证实抽样总体对目标总体的合理概括, 即任何两者间的差异都需要被讨论和考虑, 需要妥协或者修正^[7]。另外, 无法明确描述总体的文献大量存在^[7], 这将影响研究各个部分的有效性, 所以清晰界定特定的目标总体及抽样总体, 提高抽样总体对目标总体的解释合适性。而在创业研究领域, 总体难以识别导致抽样成为难题^[8]。

第二, 根据研究问题准确阐述分析单元。在管理研究中分析单元可能是个人、群体、部门或者组

织, 又或者是产品、应用、系统、项目, 分析单元可以是任何研究者在提出的问题和假设中的任何主体, 可能会出现多于一个分析单元的情况^[6]。

第三, 确定合理的样本量并且阐述其依据。在管理学的问卷调研法中一贯的错误是确定不合理、不足够的样本量^[9], 此外就是现实中样本的基数相对偏小, 导致抽样的样本量相对较小, 这是在刚起步的创业研究学者普遍反映的问题^[8]。然而样本量是问卷调查研究设计中影响显著差异、关系和相互作用识别的重要因素, 所以能否确定且获得合理的样本量成为衡量取样合适性的重要标准一致。通常认为测量的准确度在100~200之间随着样本量的增加而提高, 达到200以后随着样本量增加而降低^[10]。

第四, 对于取样过程进行充分描述。即对整个样本选择的过程及步骤进行详细描述, 帮助其他研究者获得评价研究和发现的信息, 也让其他研究者可以在对比的条件下重复此研究, 这就是当今管理学界越来越强调的研究方法的可复现性^[11]。由于篇幅限制, 可能无法对每一个技术细节进行批露, 但是也存在某些方法足够可以达到此目的^[7]。

第五, 对于回收率的情况描述。由于问卷调查非常依赖受访者的参与程度, 低回复率是学者在分析调查结果主要关注的问题。低回复率会引起样本量不足, 降低统计功效, 限制可应用的统计方法类型, 同时降低读者对文章主观认知的可信度评价, 还会引起无反应偏差, 导致产生误导性的结论^[3]。学者发现在1975年顶级组织研究期刊(例如《Journal of Applied Psychology》《Academy of Management Journal》)发表的研究中, 回复率为64.4%, 而1995年则下降到50%^[12]。

第六, 根据所采用的抽样策略清晰阐述合理的估计和数据分析过程, 主要包括样本估计偏差评估、无回复率偏差评估等。比如在采取群体样本时, 如果采用简单随机抽样的方差估计方法是不合适的, 这就意味着要根据抽样的设计来采取合适的估计和分析方法^[4]。

(二) 测量合适性的评估标准: 信度、效度与同源偏差

测量代表科学的概念、发展测量工具、对相关和不相关差异来源进行评估的过程。本文评估测量合适性的标准包括效度、信度、共同方法偏差三大方面。信度反映测量结果免受误差影响的程度, 表征测量结果的一致性、稳定性和可靠性, 主要可以从复本信度、重测信度、折半信度等角度进行评估^[5]; 效度用于检验测量工具能有效测量到所要测

量的构念的程度即测量的准确性^[1],主要可以从内容效度、结构效度(聚合效度与辨别效度)等角度进行评估;共同方法偏差指的是因为同样的数据来源或评分者、同样的测量环境、项目语境及项目本身特征所造成的预测变量与效标变量之间人为的共变。这种人为的共变对研究结果产生严重的混淆并对结论有潜在的误导,是一种系统误差。

第一,对信度的评估。早期研究强调测量工具和信度检验^[13]。信度强调测量是稳定性、准确性,即多大程度可以相信此测量结果。信度越高的测量,随机误差越小。关于问卷调查的信度,一般通过平行测试获得样本,随后进行信度测量。最常见的三种信度检验包括复本信度(Parallel forms reliability)、重测信度、折半信度。复本信度是指同时发展两份等效但不完全相同的测验(相同的平均数方差以及和其他概念相同的相关系数),每个被试需要同时回答这两份问卷。测试结果之间的相关系数越高,信度就越高;这种测试最接近平行测试,总体的代表性强,但是设计可相互替代问卷的难度较大,而且一旦被调查者掌握题目的互替规律,可能失去复本意义。而重测信度是在不同的时间点,使用同一个测验,对同一组参与者进行测量,两次测量的相关系数,也被称为稳定性系数;重测信度因操作明确简单、省时省力而受到追捧,但也因前后两次作答相互不独立而使得根据题目调整间隔时间非常重要,比较少适用于异质性调查。折半信度是同质信度的一种,用来测量的问题或测验分为数目相等的两半,独立几份,然后计算这两组题目得分的相关性。它们之间的内部一致性越强,代表所测量的内容集中考察相同特质,说明问卷质量好。折半信度的缺点是分半的方法不同,导致计算出来的信度估计值是不唯一的。另外,考虑内部一致性时,要同时考虑量表条目的数量和条目的内容范围。Cronbach's α 系数是最为广泛应用的同质信度指标,同时接受的批评也是最多的^[14]。Lance等提出追求Cronbach's α 太高会削弱内容效度^[15]。由于不同的假定、前提条件及信度的相对性,建立统一的问卷可靠性的信度系数标准是非常困难的。

第二,对效度的评估。效度检验起源于20世纪50年代,逐渐发展到今天,关注的是被测量的主体特征,测量数值的含义,数据的有效性,回答了我們是否正在测量我们想要测量的对象这一问题,被认为是“检验的检验”。效度检验重要的原因是它对于研究实现了框架的统一,保证推理过程是研究

者基于他们的测量、设计、分析的选择和集合得出的,以得出科学有效的结论。早有组织研究学者开发了检查清单对效度相关研究进行评估,提出四个关键的评估标准,分别为内部效度、构念效度、统计效度、外部效度^[16]。在管理学中,最常见的效度测度为构念效度(Construct validity),包括聚合效度及区分效度。构念效度反映测量内容多大程度上反映研究中所要测量的构念,包括:(1)遗漏了一些反映构念内容的测量指标;(2)增加了一些与构念无关的指标;(3)构念的不同成分的比例结构是否偏颇^[17]。Campbell和Fiske^[18]提出采用多特质——多方法矩阵的方法,来检验聚合效度和区分效度。聚合效度是指不同的观察变量是否可以用来测量同一潜变量,旨在通过不同方式测量同一构念时,所观察到的数值之间应该高度相关,可通过潜变量CFA模型从相关测量指标路径系数显著性进行判断;而区分效度则是指不同潜变量之间是否存在显著差异,如果两个潜变量的相关系数95%的置信区间涵盖了1,另外如果效度模型没有显著优于完全模型,则表示两个构念间缺乏辨别力。此外管理学中及内容效度(Content validity),效标关联效度(Criterion validity)的应用较少。

第三,对共同方法偏差的评估。共同方法偏差是组织管理领域引起误差的常规问题,在行为科学(尤其是问卷研究)中广泛存在,学者提出采用多质多法的设计来预测共同方法变异的重要性^[15]。Podsakoff等通过多质多法对70篇覆盖心理学、市场学、教育学的文献进行共同方法偏差的研究,得出在研究测量的误差里面,大概有四分之一的偏差来自共同方法偏差^[19]。以下一个或者多个要素影响都有可能引起的系统性误差,包括共同评定者效应、一致性倾向、内隐观、社会称许性、宽待偏向、默认偏向、心境、短暂的心情等^[19]。正如Siemens等所指,用同种方法测量变量得出变量之间的关系会因为共同方法变异而被夸大,导致由于观察测定数受影响而引起的构念效度妥协,同时引起协方差变异,及如果构念A和构念B采用了相同的方法进行测量,那么方法会对观测变量之间的关系产生系统性的影响^[20]。Podsakoff等详细提出了系列程序和统计的方法进行共同方法偏差的控制^[19]。

二、取样和测量合适性的现状评估

本文以《管理世界》近十年发表的使用问卷调查法的论文为评估的样本。选择这个期刊的原因是其作为国内管理学界引用率最高、影响力最广的期

刊,具有代表性。具体抽样标准为:第一,研究领域为组织与管理研究;第二,研究目的为检验或构建理论(描述性的调查分析排除在外);第三,问卷调研法为主要的研究方法(案例研究法等其他方法中用到问卷调研进行补充数据收集的排除在外)。根据这三个标准,本文第二作者逐篇阅读了从2007~2016年6月份的所有论文摘要,选取了共143篇,而后就有争议的部分文章两位作者进行讨论,最终确定评估的样本为137篇。图2展现了每年的问卷调研法的发表趋势,近几年基本保持在10篇左右每年,是一个大概合适的比例;图3展现了不同研究领域的发文数量,其中人力资源与组织行为领域最多,服务与营销、创新与创业、战略与治理相对比较持平。



图2 2007~2016年6月《管理世界》采用问卷法的论文发文趋势

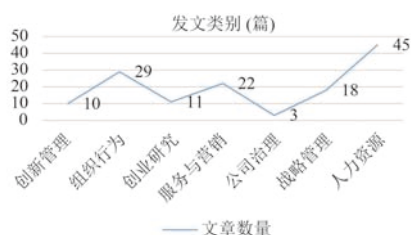


图3 2007~2016年《管理世界》采用问卷法的论文类别

(一) 取样合适性的评估

第一,目标总体和抽样总体的清晰界定。被评估样本中早年的研究鲜有文献对目标总体和抽样总体进行清晰界定,这就导致读者根本无法判断抽样总体是否能够代表目标总体,样本能否代表抽样总体。例如,被评估样本中常见做法有两类:第一类是通过便利抽样的原则,给社会网络中的企业家(例如认识的人,MBA/EMBA学员等)发放问卷。由于没有清晰界定目标总体和抽样总体,致使无法判断最终的样本是否可以代表总体;第二类是选择一家或者几家企业的员工收取了一定数量的问卷,这种做法的大部分研究都未能清晰界定符合条件的抽样总体,进而也无法判断样本是否能够代表总体。

近两年发表的论文中在这一方面有了极大改

善,例如姚晶晶等^[21]暗含了其目标总体为中国的民营企业,经过“两步骤”分层随机抽样——选择了3个沿海省份(广东、福建、浙江)和3个内地省份(湖北、四川和山西),而后挑选了每个省份的一个地级市里的两个区/县,随机从当地注册的企业中选择一定数量的企业——最终确定了抽样总体为400家。赵瑜等^[22]界定了其抽样总体为浙江省某卫视集团中自愿参与研究的40个节目团队中154名员工。

第二,根据研究问题准确阐述分析单元。组织管理领域分析单元主要包括个人层面的企业家、领导、高管、经理人、员工,团队层面的创业团队、高管团队、研发团队等,组织或企业,跨组织或多个分析单元并存的情况。就个人层面的大部分文献,如早期研究对企业家胜任力^[23]及近年对领导与员工关系,上下级关系的研究^[24],都对分析单元给出了相对准确的阐述。就团队、组织层面的文献,部分文献对分析单元的描述有所欠缺,但是近年来情况有所改善,如赵瑜等^[22]明确指出研究中的伦理型领导是一个团队层面变量,并对个体层面的员工评价的伦理型领导数据进行聚合,实现个体层面与团队层面变量的跨层嵌套。另外,朱苏丽等^[25]研究员工与组织时,摒弃了将“员工”“组织”看作相互独立的利益主体,而是从社会交换理论对中国企业员工与组织关系的亲情交换关系进行研究,对分析单元给出了相对明确的界定。

第三,确定合理的样本量并且阐述其依据。大部分文献都直接指明了样本数量,但在阐述样本量的合理性及依据有失偏颇,说服力不足,特别是缺乏样本量相对总体的量的充分阐述及在特殊情况下样本量的合适性,以及样本量与回复率之间的关系。如姜翰和金占明^[26]主要以运动用品(鞋服)制造业为样本研究企业间关系,经过两个步骤,前期确立业内97家重点企业为对象获取相关数据,后期对142家其他企业进行问卷调查,文章只指出最终得到151个企业关系的有效数据,但没有清晰指明样本量设置的合理性。又如陈卫旗^[27]在正式抽样时在全国18家企业的人力资源部各自抽取20名员工作为组织样本,同样没有准确阐明其依据,这就暗含着样本量是随机的。

第四,对于取样过程进行充分描述。早期的文献普遍存在缺乏取样过程的详细描述的情况,例如部分文献未清晰呈现调查对象与测量单位之间关系,其中一个例子是卫武等^[28]在研究企业对利益相关者压力反应情况时,选取500家企业的主管进行

问卷填写,但是未对调研对象及测量单位的关系进一步准确描述。当然也有部分学者在这方面做的比较妥当,如庄贵军等^[29]问卷发放的对象是熟悉关系营销的销售人员,从而保证了调查对象与测量单位(关系型渠道治理、关系营销导向、跨组织人际关系)之间的关系;另外,部分学者未对地理位置、所属行业、预测试的技术细节进行准确披露,阐述取样过程都存在一定的缺失和不足,导致难以对样本代表总体特征的有效性进行评估,无法确定取样偏差可能对获得样本的本质造成的影响,同时也很难让其他研究者在可比较的条件下复制该研究。当然,也有一些学者在这方面做得相对较好,如姜翰等^[30]研究创业企业机会主义行为时,样本空间选定为运动用品(鞋服)制造企业集群,明确界定抽样时间及创业企业范围,分两个阶段进行抽样,阶段一收集402家集群企业,阶段二排除了不符合既定创业企业标准及36个信度明显不高或数据缺失严重的样本后得到366家企业作为样本。又如姚晶晶等^[21]对数据搜集分层抽样的技术细节、仔细披露在调查前中后采取的多种手段的过程和细节来保证调查可靠性。

第五,对于回收率和无回复率的情况描述。大部分文献都清晰阐明了问卷的回收率,这是值得肯定的部分。如胡望斌等^[31]明确指出在对590家企业进行问卷调研时,回收问卷150份,有效问卷回收率为25.42%。无回复问卷会导致外部效度下降,那么这就对基于样本得出的结论的解释性产生质疑,然而鲜有文献对无回复情况做出的补救措施。

第六,根据所采用的抽样策略清晰阐述合理的估计和数据分析过程。早期有社会调查性研究,主要用于政策制定,这类文章普遍缺乏进行样本估计偏差评估。当然,不乏一些学者开始采取了手段对样本估计偏差进行评估,比如张杰等^[32]考虑到样本可能存在非随机性或异常值对回归产生影响,去掉5%比例企业规模最高和最低的样本进行稳健性检验,并且对苏南、苏中、苏北三个地区样本各自进行回归,得出与总体样本变量稳健一致的结果,这样可以有效避免了样本估计偏差。钱丽萍等^[33]进行多次无差异性检验,确定寄回问卷和无回复问卷的企业在规模、处理与零售商关系中使用的营销战略等问题没有显著差异,保证了样本的代表性。又如赵可汗等^[34],由于样本企业来自不同地区(4个),不同行业(6个),建立多层次模型以考察团队信息深度加工在团队、企业、产业和地区各层次的方差,得出在产业、地区层次方差解释比例不显著的

结果,由此选择仅在企业层次进行聚类回归分析。但是,忽视样本估计、无回复偏差评估的学者仍占很大比例,不利于帮助判断样本的总体代表性,值得学者提高警惕。

(二) 测量合适性的评估

第一,对信度的评估。早期研究特别是07、08年间忽视信度评估,极少文献采取信度检验,后来信度检验引起了越来越多学者的重视,但是也有部分学者忽视这一问题。即便部分学者采取了一定的信度检验,但是存在普遍仅仅关注内部一致性(折半信度为主),而忽视其他信度的评估的情况,例如复本信度、重测信度的评估。如杨玉龙等^[35]则对问卷涉及的9个变量进行了内部一致性测试,为获得最高的Cronbach's α 值,从整套量表删除了一些项目以保证较高信度,但也有可能威胁量表的效度。

第二,对效度的评估。效度评估在近年来越来越受到重视,学者们普遍关注结构效度,缺乏内容效度的评估,做的比较好的有贾建锋^[36]通过借鉴成熟量表设计、反复推敲每个问题、采用严格的探索性因子和验证性因子这3个环节保证了内容效度,此外通过对3个变量(战略导向、企业绩效、高管胜任特征)的验证性因子分析的各项拟合指标,通过NFI值基本接近经验值得出聚合效度良好的结论,此外作者采用AVE值的平方根与维度之间的相关系数的大小关系判别区分效度,保证了效度评估的全面性。

第三,对共同方法偏差的评估。纵观《管理世界》137篇采取问卷法的文献中,进行共同方法偏差评估的文章只有较少一部分,需要更多的学者对其重要性予以关注。当然,陆续有学者开始考虑到问卷都由同一个被试者填写可能会引起共同方法变异的问题,而采用Harman单因子检验法对回收数据进行检验是常用手段。例如焦豪等^[37]在研究创业导向与企业动态能力关系时,贾建锋等^[36]在分析高管胜任特征与战略导向的匹配对企业绩效的影响时,均采用该方法排除了共同方法变异的问题。但是,仍有大量学者忽视对问卷数据进行同源偏差检验。

三、取样与测量合适性常见问题总结与建议

通过上文的简要评估,本文接下来将简要总结中国管理研究中问卷调查法研究在取样与测量合适性方面存在的主要问题,并以发表在顶级期刊上的

最佳实践来提出解决这些问题对应的建议。

问题一:普遍缺乏对抽样总体的明确界定。

缺乏对抽样总体的明确界定导致读者无法清楚识别样本的对象、范围及特征,同样无法合理评估样本是否可以代表目标总体,无论抽样的其他环节表现得再好,最终也会影响研究的效度^[4]。在创业研究领域,总体难以识别导致抽样成为难题^[8]。例如,Zhang和Li^[38]在研究新创企业与服务中介(如会计和金融服务企业、技术服务企业、法律服务和人才搜索企业)之间的连带与该企业产品创新之间的关系时,明确界定以广东省一个高新技术产业集群中随机抽取的500个新创企业为抽样总体,并且清晰阐述所抽取样本必须满足的三个标准,提高抽样总体对目标总体的解释合理性。

问题二:普遍未能对抽样过程进行详细描述。

如何提供获取样本的充足信息导致在评估总体特征时难以判断样本的有效性,无法评估抽样偏差可能对获得样本的性质产生的影响。另外好的研究能够被精准设计并开展,缺乏抽样流程的详细阐述不利于推广该研究^[4]。例如Li等详细描述了反向翻译、前测、1000个企业的抽样总体、抽样过程、采访过程等^[39];Zhang和Li也详细描述了从开始对抽样总体、标准的界定,现场调查各企业邀请两位响应者的安排,确保受访者的提供准确数据等技术细节及流程;详细的描述可以确保抽样过程可复现性,帮助读者掌握抽样的细节,对其结论作出评判^[38]。

问题三:普遍缺乏对样本估计偏差评估、无回复偏差的评估。

样本估计偏差及无回复偏差是帮助读者判断样本抽样信度的重要手段,大部分学者都忽视了这一点。Zhang和Li通过严格筛选、比对受访者在决策制定的参与程度、工作年限、知识背景、职位等来确保参与水平无显著差异,通过每家企业的两位受访者同时回答两个的问题来对响应相关性进行检测,通过对比响应及无响应新创企业的规模和年龄方面的特征,均发现并无偏差,由此多层次地保证了抽样的有效性^[38]。Li等则通过相应企业和无响应企业在企业特征(企业所有权、企业年龄、产业类型和利润)进行MANOVA分析发现无显著偏差^[39]。Flynn等则通过对最早收回的最晚收回的两批问卷的企业固定资产、销售额和员工数等基本信息进行T检验,发现并无显著差异来检验无响应偏差^[40]。

问题四:普遍缺乏对被调查对象与测量单元之间关系的评估。

此不足导致抽样数据的可信度下降,而Zhang和Li对于因变量(产品创新)、自变量(初创创业与中介机构的关系、感知产业增长等)单位与调查对象资质和匹配度有严格把控^[38]。比如被调查对象需要是新创企业的管理主管、CEO、业务或产品经理,并且在战略决策中参与程度是4(1=完全不参与,5=非常高参与度),这些方法都确保了调查对象能够有效提供准确数据。Flynn等汇报了访谈对象的职位、工作年限^[40],而Li等则通过现场访谈来确定应答者是否为高管且对访问内容的熟悉程度^[39]。

问题五:重点关注内部一致性,而忽略了其他提升信度的策略。

除了内部一致性,重测信度、复本信度等都可以作为提升信度的一些策略。例如Zhang和Li不仅仅关注内部一致性,用Cronbach's alpha系数评估构念的可靠性,还把问卷分为两部分,让两位高管分别填写一部分,并核实了两位响应者之间在同时回答的两个问题上的评判一致性^[38]。相类似的,除了Cronbach's alpha, Li等在问卷收回过程中收集了访谈对象的名片,并事后随机电话回访了60位对象,每人重新回答了10个问题,并评判重测信度^[39]。

问题六:对内容效度的评估缺乏。

在问卷层面,Zhang和Li^[38], Li等^[39]和Flynn等^[40]都选择已经使用过的有效问卷,并将问卷进行反向翻译避免文化偏差,通过8~30左右高管面对面访谈进行预测试,识别出有效的测量题项,对有问题的问题进行修改或者删除,这些技术细节确保测量的内容有效性。当然,还可以通过专家就某个构念的测量是否符合她们对此构念的认识进行主观判断,看看构念内容的测量指标是否遗漏、是否无关或者比例结构是否偏颇来判断内容效度^[17]。

问题七:聚合效度与区分效度的做法存在的问题。

在采用验证性因子分析进行聚合效度评价是,需要把涉及到的所有潜变量均放在一个测量模型中进行分析,区分效度则主要借鉴多特质多方法模式进行分析。例如,Zhang和Li^[38]采用验证性因子分析评估聚合效度,测量模型的拟合值优秀,且所有路径系数均显著,说明聚合效度较好;区分效度的识别则采用 ϕ_s 在95%的置信度上都不包含值1和卡方差异性检验来判定。Li等^[39]和Flynn等^[40]也采用了相类似的做法。

问题八:普遍缺乏对共同方法偏差的关注。

共同方法偏差的控制被大大忽略,Zhang和

Li^[38]按照Podsakoff等^[19]的建议,整合了程序方法及统计技术来评估共同方法偏差。前者主要通过每家企业设置两名受访者、访前培训等收到,后者除了采用常用的Harmen's单因素测试之外,还进行了潜变量方法对一个未测量潜变量进行控制。此外,交互效应的成立也提供了一个侧面的证据。Li等^[39]主要采用了不同数据来源来规避同源偏差:企业特征(企业所有权、绩效年龄、规模和产业类型)和产业特征(竞争强度和结构不确定性)数据来源于二手数据,而核心自变量来源于问卷;并用Harmen's单因素测试进行检验。Flynn等^[40]采用了Harmen's单因素检验外,还找到一个方法因素(Method factor),并对比了加入这一因素的模型,尽管模型拟合更优,但路径系数相似。这里需要说明的是Harman单因素检验的最大优点是简单易用,但仅仅是一种评估共同方法变异严重程度的诊断技术,而且是一种不灵敏的检验方法,并没有任何控制方法效应的作用。除了Harman单因素检验,还可以通过偏相关分析法,即将方法变异来源作为统计分析中的一个协变量;或者在结构方程模型中,将共同方法偏差作为一个潜在变量,在无可测方法情况下进行方法因素效应的控制等手段^[19]。

四、结语

问卷调研法在取样和测量方面的合适性会显著影响研究的严谨性,本文的意义在于:第一,提醒管理学者在采用问卷调研法时特别关注取样和测量的合适性,关注问卷调研过程中的严谨性;第二,为管理学者采用问卷调研法进行理论检验过程中应该规避的常见问题提供一个清单和建议。总之,期待中国的管理学者不论是构建管理的中国理论还是中国的管理理论的道路上,均能为人类管理学知识的发展做出自己的贡献。

参考文献

- [1] 陈晓萍,徐淑英,樊景立.组织与管理的实证研究方法[M].北京:北京大学出版社,2012.
- [2] BARTLETT J E, KOTRLIK J W, HIGGINS C C. Organizational research: Determining appropriate sample size in survey research [J]. Information Technology, Learning, and Performance Journal, 2001, 19(1): 43-50.
- [3] ROGELBERG S G, STANTON J M. With organizational survey nonresponse[J]. Organizational Research Methods, 2007, 10(2): 195-209.
- [4] WANG L, MCNAMARA J F. An evaluation of the

sample designs in educational survey research [R]. Chicago: ERIC, 1997.

[5] 罗胜强,姜嫄.管理学问卷调查研究方法[M].重庆:重庆大学出版社,2014.

[6] KRAEMER K L. Survey research method in management information systems: an assessment[J]. Journal of Management Information Systems. 1991, 10(2): 75-105.

[7] DATE P U B, Type P U B, Price E. Practical guidelines for evaluating sampling designs in survey studies [J]. Evaluation Methods, 1999: 25.

[8] SHORT J C, KETCHEN Jr D J, COMBS J G, IRELAND R D. Research methods in entrepreneurship[J]. Organizational Research Methods, 2010, 13: 6-15.

[9] WUNSCH D. Survey research: Determining sample size and representative response[J]. Business Education Forum, 1986, 40(5): 31-34.

[10] FOWLER F J. Survey Research Methods[M]. Beverly Hills, CA: Sage Publications, 1984.

[11] SAVALL H, ZARDET V, BONNET M, PERON M. The emergence of Implicit criteria actually used by reviewers of qualitative research articles[J]. Organizational Research Methods, 2008, 11(3): 510-540.

[12] BARUCH Y. Response rate in academic studies-A comparative analysis [J]. Human Relations, 1999, 52(4): 421-438.

[13] KUDER G F, RICHARDSON M W. The theory of the estimation of test reliability[J]. Psychometrika, 1937, 2: 151-60.

[14] CHO E, KIM S. Cronbach's coefficient alpha: Well known but poorly understood[J]. Organizational Research Methods, 2015, 18(2): 207-230.

[15] LANCE C E, DAWSON B, BIRKELBACH D, HOFFMAN B J. Method effects, measurement error, and substantive conclusions[J]. Organizational Research Methods, 2010, 13(3): 435-455.

[16] COOK T D, CAMPBELL D T. Quasi-experiments : Design and analysis issues for field settings[J]. Psychologica Belgica, 1979, 20(1): 109-112.

[17] HAYNES S N, RICHARD D, KUBANY E S. Content validity in psychological assessment: A functional approach to concepts and methods[J]. Psychological Assessment, 1995, 7(3): 238-247.

[18] CAMPBELL D T, FISKE D W. Convergent and discriminant validation by the multitrait-multimethod matrix[J]. Psychological Bulletin, 1959, 56(56): 81-105.

[19] PODSAKOFF P M, MACKENZIE S B, LEE J Y, PODSAKOFF N P. Common method biases in behavioral research: A critical review of the literature and recommended remedies[J]. Journal of Applied Psychology, 2003, 88(5): 879-903.

[20] SIEMSEN E, ROTH A, OLIVEIRA P. Common

method bias in regression models with linear, quadratic, and interaction effects[J]. *Organizational Research Methods*, 2010, 13(3): 456-476.

[21] 姚晶晶, 鞠冬, 张建君. 企业是否会近墨者黑: 企业规模、政府重要性与企业政治行为[J]. *管理世界*, 2015(07): 98-108.

[22] 赵瑜, 莫申江, 施俊琦. 高压工作情境下伦理型领导提升员工工作绩效和满意感的过程机制研究[J]. *管理世界*, 2015(08): 120-131.

[23] 林泽炎, 刘理晖. 转型时期中国企业家胜任特征的探索性研究[J]. *管理世界*, 2007(01): 98-104.

[24] 刘智强, 李超, 廖建桥, 龙立荣. 组织中地位、地位赋予方式与员工创造性产出——来自国有企事业单位的实证研究[J]. *管理世界*, 2015(03): 86-101.

[25] 朱苏丽, 龙立荣, 贺伟, 王忠军. 超越工具性交换: 中国企业员工——组织类亲情交换关系的理论建构与实证研究[J]. *管理世界*, 2015(11): 119-134.

[26] 姜翰, 金占明. 企业间关系强度对关系价值机制影响的实证研究——基于企业间相互依赖性视角[J]. *管理世界*, 2008(12): 114-125.

[27] 陈卫旗. 组织与个体的社会化策略对人——组织价值匹配的影响[J]. *管理世界*, 2009(03): 99-110.

[28] 卫武, 夏清华, 资海喜, 贺伟. 企业的可见性和脆弱性有助于提升对利益相关者压力的认知及其反应吗?——动态能力的调节作用[J]. *管理世界*, 2013(11): 101-117.

[29] 庄贵军, 李珂, 崔晓明. 关系营销导向与跨组织人际关系对企业关系型渠道治理的影响[J]. *管理世界*, 2008(07): 77-90.

[30] 姜翰, 金占明, 焦捷, 马力. 不稳定环境下的创业企业社会资本与企业“原罪”——基于管理者社会资本视角的创业企业机会主义行为实证分析[J]. *管理世界*, 2009(06): 102-114.

[31] 胡望斌, 张玉利, 杨俊. 同质性还是异质性: 创业导向对技术创业团队与新企业绩效关系的调节作用研究[J]. *管理世界*, 2014(06): 92-109.

[32] 张杰, 刘志彪, 郑江淮. 中国制造业企业创新活动的关键影响因素研究——基于江苏省制造业企业问卷的分析[J]. *管理世界*, 2007(06): 64-74.

[33] 钱丽萍, 刘益, 喻子达, 陶蕾. 制造商影响战略的使用与零售商的知识转移——渠道关系持续时间的调节影响[J]. *管理世界*, 2010(02): 93-105.

[34] 赵可汗, 贾良定, 蔡亚华, 王秀月, 李珏兴. 抑制团队关系冲突的负效应: 一项中国情境的研究[J]. *管理世界*, 2014(03): 119-130.

[35] 杨玉龙, 潘飞, 张川. 上下级关系、组织分权与企业业绩评价系统[J]. *管理世界*, 2014(10): 114-135.

[36] 贾建锋, 唐贵瑶, 李俊鹏, 王文娟, 单翔. 高管胜任特征与战略导向的匹配对企业绩效的影响[J]. *管理世界*, 2015(02): 120-132.

[37] 焦豪, 魏江, 崔瑜. 企业动态能力构建路径分析: 基于创业导向和组织学习的视角[J]. *管理世界*, 2008(04): 91-106.

[38] ZHANG Y, LI H. Innovation search of new ventures in a technology cluster: The role of ties with service intermediaries[J]. *Strategic Management Journal*, 2010, 31(1): 88-109.

[39] LI J, POPPO L, ZHOU K Z. Do managerial ties in China always produce value? Competition, uncertainty, and domestic vs. foreign firms[J]. *Strategic Management Journal*, 2008, 29(4): 383-400.

[40] FLYNN B B, HUO B, ZHAO X. The impact of supply chain integration on performance: A contingency and configuration approach[J]. *Journal of Operations Management*, 2010, 28(1): 58-71.

Adequacy of the Sampling Process and Measurement in Survey Studies: An Assessment and Suggestion

LIU Yang XIE Li

(South China University of Technology Guangzhou 510640 China)

Abstract Survey research becomes not only the commonly used quantitative method in psychology and sociology areas, but also the most popular data collection method in management research for its high convenience, strong operability, and low cost. However, the appropriateness of sampling and measurement are hard to guarantee. In order to improve the standardization and science of the survey research in management research, this paper evaluates the current situation of the appropriateness of sampling and measurement based on the papers published in *Management World* from 2007 to 2016, identifies research problems, and proposes corresponding suggestions.

Key words survey studies; sample; measurement; Chinese management studies

编辑 何婧