

跨文化研究方法的回顾及展望

梁 觉¹ 周 帆²

(¹ 香港城市大学管理系, 香港) (² 浙江大学管理学院, 杭州 310058)

摘 要 文化现象是行为科学关注的重要课题, 跨文化研究从文化对比的角度对文化现象进行阐释, 不过这类研究在研究方法上要面对一些相对独特的问题, 本文将从研究设计、测量、数据统计等方面分析跨文化研究中值得研究者关注的方法问题。作者首先从研究设计的角度分析如何通过汇聚方法(consilience methods)加强检验跨文化研究的因果关系的能力, 然后分析跨文化研究中测量对等性(measurement invariance/equivalence)问题以及相关的统计方法, 最后简要介绍了多层线性模型(hierarchical linear modeling)在跨文化研究中的应用、并讨论跨文化研究方法的发展趋势。

关键词 跨文化研究; 汇聚方法; 概念对等性; 测量对等性

分类号 B841

1 跨文化研究中因果关系验证和汇聚方法

1.1 跨文化研究中因果关系验证问题

来自不同文化背景下的个体在认知、态度、信念和价值观等方面存在很多差异。随着国际商务和国际间交流的迅速发展, 文化现象已经成为研究者广泛关注的重要课题。跨文化研究作为研究文化现象的重要方法, 一般含有文化比较的意味, 通常在两个及以上的文化之间某个变量的水平或者是变量之间关系进行比较, 并将文化作为一个自变量来分析文化效应(梁觉, 周帆, 2008)。跨文化研究需要将文化作为自变量来解释因变量上的变异, 但由于研究者无法通过随机分组的方法来进行操纵、排除可能的混淆变量的影响, 跨文化研究只能被视为准实验研究, 其内在效度低于严格的实验研究, 难以有力地论证文化与效应之间的因果关系。

因果关系的推论通常要满足三个条件: 1) 原因发生于结果之前, 2) 原因和结果存在稳定的关系, 并且 3) 对于结果的发生, 除了这个原因以外没有其它替代性解释 (Shadish, Cook, & Campbell, 2002)。在跨文化研究中, 因果关系的推论会遇到很多困难。由于个体通常是生活于某个文化中, 所以

很难说文化作为自变量是发生在结果变量之前。但根据 Cook 和 Campbell (1979) 提出的即时效应(instantaneous effect), 原因所导致的结果可能是同时发生的, 因此对于第一个条件, 研究者只需要证明自变量的变化导致相应的因变量变化。因果关系的第二个条件比较容易检验, 最大的难点是在第三个条件, 即要排除文化以外的其它可能的替代原因对结果的解释。但由于文化一般只是导致某个现象的其中一个条件而已, 通常都不是导致结果的充分及必要条件, 因此跨文化研究的因果推论的最大难题就是如何排除混淆变量对结果的解释。

1.2 跨文化研究中常见的因果关系推论偏差

在跨文化研究中, 研究者所采取的一些研究方法可能影响研究的因果关系推断, 以下从统计推论效度(statistical conclusion validity)、内部效度(internal validity)和构念效度(construct validity)三个方面分析这些问题。

跨文化研究统计推论效度方面常犯的第一个错误是对显著性检验的忽视。一些研究仅从“观察”得出结论。例如一些研究只限于观察某个变量在两个文化平均值之间是否有“明显”差异, 但没有做进一步的统计检验; 还有一些研究观察到回归系数在一个文化样本中显著, 而在另一个文化样本中不

显著, 就认为变量间关系在两个文化间存在显著差异, 而没有检验回归系数差异的显著性。在统计推论效度方面的问题还包括对测量中的结构对等性 (structural equivalence) 和截距对等性 (scalar equivalence) 检验的忽视。例如很多跨文化研究仅在不同的样本间检查量表的一致性系数是否相当, 而没有进一步对测量对等性的问题进行检验。

跨文化研究在内部效度方面常出现的问题包括样本的选择性、以及被试对研究所使用的测量或操纵的反应方式上面的文化差异。样本的选择性指的是研究者选取的样本之间除了文化差异以外, 还存在其它方面的差异, 而正是这个差异而不是文化差异本身是导致结果的原因。例如 Leung, Lau, 和 Lam (1998) 的一个研究发现香港家长的威权式教养方式与孩子的学业成绩呈正相关, 而在澳洲和美国则发现负相关的结果。但进一步分析发现香港家长的教育程度要低于澳洲和美国, 在澳洲和美国教育程度较低的家长群体中, 也发现威权教养方式与孩子的学业成绩呈现正相关, 因此教养方式与学业成绩之间的关系是教育程度而不是文化差异的某个方面所导致的。除样本的选择性以外, 被试反应方式的文化差异也可能是降低研究内部效度的原因。例如 Hui 和 Triandis (1989) 发现相对于其它族群, 拉丁裔美国人在回答问卷的时候倾向于一种两极化的反应方式, 因此一些研究结果可能是由于反应方式而不是研究者所设想的某个文化维度上的差异所导致的。

跨文化研究在构念效度上常见的问题主要包括构念概念定义的不对等、概念操作定义的不对等、跨层次问题等。首先概念在文化之间的定义可能存在很大的不同, 例如 Sternberg (2004) 发现不同文化对于智力的理解存在很大的差异。在非洲一些文化中服从被看作是智力较高的表现, 跟西方文化截然不同 (Serpell, 1993)。即使概念的定义在不同文化是非常接近的, 概念的操作定义也可能出现很大的偏差。例如在中国文化中, 将“粗鲁”的操作定义定为打断别人谈话的频率是合理的, 但是巴西文化中显然不能这样定义, 因为插话是普遍的可接受的会谈方式。概念的跨层次问题也是经常遇到的构念效度问题, 很多跨文化研究涉及个体和文化两个层面, 如果简单地将一个层面的概念或结论类推到另一个层面可能是错误的。最后, 测量题项的不对等、对研究材料和情境的熟悉度、对研究情境的反应性以及实验者的期望等因素也可能削弱跨文

化研究的构念效度。

1.3 跨文化研究中汇聚方法及其类型

在其它一些学科, 研究者同样也面临不能进行严格实验研究的难题。例如在传染病学中, 伦理的约束使得研究者不可能通过随机分组这样的严格实验操作来检验因果关系。在生物进化论的研究也不能以实验方法进行。在进化心理学中, Caporael (2001) 提出研究者可以借用进化生物学家提出的汇聚归纳方法。没有单一证据可以证明自然选择理论, 但是不同来源的众多证据的汇集排除了许多替代解释, 从而使这一理论难以被质疑。通过“汇聚”的多方面的证据将大大加强研究结果对所论证关系的说服力, 弥补跨文化研究本身存在的一些固有的不足。下列对汇聚方法的分析源自 Leung 和 van de Vijver (2008)。

跨文化研究中的汇聚方法可以分为 4 种类型:

1) 多重情景的汇聚 (contextual consilience) 收集多个不同文化情景或文化群体的数据, 如果这些结果都与理论预期吻合, 则为所论证的关系提供了有力的支持;

2) 多重方法的汇聚 (methodological consilience) 是使用多种方法, 例如调查法、实验法、纵向研究等方法, 来验证某个研究假设, 考察这些不同方法是否得到相同的结论;

3) 多种预测的汇聚 (predictive consilience) 指从该因果关系理论中推导出一些预期假设, 如果这些假设都得到验证, 那就大大增强了对这个因果关系的确信程度;

4) 排他的汇聚 (exclusive consilience) 指除研究者所论证的因果假设以外, 研究者提供证据排除其它多种替代性假设可以解释观测到的结果的可能性。

1.4 跨文化研究中汇聚方法的应用策略

在跨文化研究设计中, 研究者可以采取一些具体的研究策略来强化研究结果对所论证研究假设的说服力, 这些策略可以被归为 3 类。

第一类策略是系统比较策略 (systematic contrast strategy), 包括 1) 多重比较策略 (multiple contrast strategy), 研究者根据研究假设的某个文化维度依照大小强弱顺序抽取 3 个以上的文化样本进行研究, 例如研究者考察“集体主义-个人主义”这个维度来解释可能的研究结果, 则可以抽取在这个维度水平上的 3 个文化组, 比如日本、法国、美国, 如果研究结果也呈现这样的次序, 则该研究结论比

仅抽取两个文化样本的研究结论的确信程度要高。

2) 纵向比较策略(temporal contrast strategy), 即考察某个文化维度上的变化是否与因变量的变化有关。通常文化的变化是缓慢而难察觉的, 因此很多研究者通过旅居者或移民这个视角来考察文化变化的效应, 例如可以比较移民时间不同的群体、新移民和移民后裔等在所关注的因变量上的差异。例如 Heine 等 (1999) 在自尊的文化差异的研究里对比了交换学生等群体, 探索文化适应导致自尊的变化, 从而验证西方文化下存在较高自尊水平的假设。有 3 种取样方式可以用来强化系统比较策略。第一种取样方式是随机重复(random replication), 即通过一个足够大的随机的多文化样本来排除混淆变量对结果的解释, 但通常这样的取样较难做到。第二种取样方式是多样化取样(diverse sampling), 通过多对文化样本来重复检验观察到的文化差异。假如在一对文化之间出现预期的文化差异, 但在另一对文化之间可能在自变量以外的其它的文化特征方面存在差异, 如果在这对文化之间可以验证先前观察到的结果, 则自变量差异更可能是导致结果差异的唯一原因。第三种取样方式是系统取样(systematic sampling), 即在取样的时候考虑混淆变量, 选取在混淆变量上匹配的两个文化进行比较, 从而排除混淆变量对可能得到的因果关系的解释。

第二类策略是协变量策略(covariate strategy), 包括: 1) 单一协变量策略(simple covariate approach), 即仅引入所关注的文化维度变量作为协变量, 如果控制这个变量以后, 文化效应不再显著, 则表明这个变量可能是文化效应的原因; 2) 多项协变量策略(complex covariate approaches), 在这种方法中, 研究者纳入了多个可能的文化维度变量, 探索文化效应的原因、或是排除其它替代性假设而强化假设的解释力, 例如可以将个人主义—集体主义维度的多个层面作为协变量, 探索哪些层面与组内偏向(in-group favoritism)之间存在关系 (Chen, Brockner, & Katz, 1998); 3) 单文化拓展(monocultural extension), 在考察文化效应可能的原因时, 研究者可以在一个文化样本内探索文化维度与因变量之间的关系, 例如 Brockner 等 (2000) 在比较程序公平效应在中美文化之间的差异后, 发现在美国被试样本中独立我—互依我维度上不同强度之间的公平效应呈现出类似不同文化样本之间的差异, 因此支持了自我构念作为解释程序公平文化效应原因的说服力。

第三类策略是多重方法策略(multimethod strategy), 包括 1) 使用多个因变量指标, 不同的因变量指标可能与不同的混淆变量有关, 如果使用不同的因变量指标都可以得到类似的结果, 那么这个文化效应被无关变量所导致的可能性会大大降低。2) 多种实验策略, 尽管在跨文化研究中不能进行完全意义的实验, 但研究者可以采用一些准实验方法, 近年来提出“实验人类学(experimental anthropology)”方法 (Cohen, Nisbett, Bowdle, & Schwarz, 1996), 即操纵一些变量来加强或减弱文化效应, 从而推证因果关系, 首先假设某个文化效应是由于某个变量所导致的, 然后操纵这个变量的水平, 观察因变量是否会出现预期的变异。实验操纵方法可以区分为两类: 在操控外部情境变量的实验(ecological experiments), 研究者可以操纵外部情境变量来探讨预期的效应, 例如改变环境中的某些因素来观察被试的反应与操纵因素之间的关系; 在操控个体的文化维度的实验(idiographic experiments), 研究者可以在个体水平上“操纵”文化, 例如使用“文化启动”(culture priming)方法, 向被试呈现代表不同文化的刺激物, 比如国旗、文化名人等等 (Hong, Morris, Chiu, & Benet-Martinez, 2000); 当然也可以在具体的文化维度变量上进行启动, 例如可以通过让被试回答自己与家人相同或不同之处, 来操纵独立我—互依我这个维度 (Trafimow, Triandis, & Goto, 1991)。

2 跨文化研究中概念和测量对等性问题

对等性是跨文化研究中的一个非常重要的问题。跨文化研究通常需要在不同的文化之间比较因变量的差异, 基于完全不对等概念进行跨文化比较, 就如同将西瓜和芝麻进行比较一样荒谬。同样, 同一个测量条目可能在不同的文化中具有不同的意义指向, 例如, 骑自行车在美国被试中可能被当作一种娱乐运动, 而在中国被试中可能被当作一种交通方式, 因此如果将这个条目用来作为休闲活动测量的跨文化比较可能会带来偏差。因此, 在跨文化研究中, 需要采取一些分析方法和统计手段, 检验所研究的概念在概念涵义上及其测量(或操纵方式)上的对等性。

2.1 概念对等性

概念对等性(construct equivalence)指概念是否在对比的文化之间具有相同的涵义。我们知道管理

科学以及其它社会科学的很多概念都来自于美国文化下的研究,其中一些概念可能在中国文化中具有不同的内涵。Farh, Zhong 和 Organ (2004)探讨了中国文化下组织公民行为的结构(Organizational Citizenship Behavior, OCB),研究者通过对中国文化下与 OCB 相关的行为的收集和分类,并对照西方文化下 OCB 的各个维度,发现在中国文化下一些 OCB 维度的内涵与西方文化下 OCB 的一些对应维度接近,但一些维度只在中国文化中出现。这个研究表明,如果在中国文化下的研究直接套用西方 OCB 的测量工具,那么就忽视了 OCB 中的那些在中国文化下重要的内容。同样的,在中国文化下建立的一些概念,可能在西方文化中也具有不同的内涵。家长式领导风格(paternalistic leadership, Cheng et al., 2004; 郑伯坝,周丽芳,樊景立,2000)研究提出“立威”、“树德”、“施恩”3种领导行为。尽管没有直接的研究证据,但在美国文化下,这3种行为不一定会构成家长式领导风格。

对概念对等性的分析是一个比较复杂的问题,研究者可以从以下几个方面着手:首先,研究者需要探讨概念的结构,如果是多维度结构,可以比较维度结构在不同文化之间是否一致,例如,对于人格这个概念,可以探讨在西方文化下发现的大五因素人格模型是否在中国文化下也成立(e.g., Cheung et al., 2001; Wang, Cui, & Zhou, 2005);其次,研究者可以考察这个概念与其它相关概念之间的关系(例如相关系数)在不同文化是否存在显著的差异,如果是的话,则可能表明这个概念的涵义存在文化间差异;最后,可以考察与这个概念存在因果关系的变量之间的关系,如果这些关系存在文化间差异,并且研究者可以排除其它可能的情景变量的解释,则也可能表明概念在不同文化之间并不对等。

此外,为全面考虑概念可能存在的文化差异,研究者在研究设计的时候,宜采取一种文化平衡的方式(decentered approach),即同时从所关注的不同文化的角度来进行概念分析和研究设计,而不是以某个文化为主导。早期的一些研究并没有特别注意这个问题(Hofstede, 1980),但近年来这种做法较为普遍。例如,在一个全球进行的领导行为研究中(GLOBE, House & Javidan, 2004),研究者从多个文化中分析概念的涵义和抽取测量条目,而不是只在某个文化中进行分析 and 确定概念内涵和测量方式,然后将其用于其它文化内;Leung 和 Bond

(2004) 进行的社会通则研究也是从不同文化中抽取条目。

2.2 测量的对等性

在水平指向的研究中,不仅要检查概念对等性,还要探究测量对等性(measurement equivalence)。即使确认了目标变量在不同文化间的概念对等性,也不能将测得的分数直接进行比较,因为我们不能确定在不同文化下的测量在单位上是一致的,就如我们不能认为在长度上100肯定大于90,因为前者的单位可能是毫米而后者的单位可能是千米。在对目标变量进行跨文化间水平比较之前,我们需要考察这个变量的“测量对等性”。一些物理的测量,例如长度和重量,可以比较容易建立测量对等性。但行为/心理的测量对等性的建立则较为复杂。以下我们将讨论在跨文化研究中建立测量对等性的方法和一些值得注意的问题。

测量的不对等性可能是由于条目偏差引起的,条目偏差是在跨文化研究中经常遇到的问题。即使研究的概念在不同文化间具有概念对等性,测量这些概念的问卷中的一些条目可能在不同文化之间存在偏差:即来自不同文化的、在这个概念上具有同等水平的受试者,在这些条目上的分值并不等同。如果对研究概念进行测量的工具存在具有文化偏差的条目,则从这个测量工具得到的跨文化研究结论可能不可靠。所以,在跨文化研究中需要对所采取的测量工具进行分析,探讨条目偏差是否存在及其原因。

用来分析跨文化研究中测量的对等性问题大概有三种方法,包括方差分析(ANOVA, van de Vijer & Leung, 2000)、项目反应理论(Item response theory, IRT, e.g., Maurer, Raju, & Collins, 1998)、验证性因素分析(confirmatory factor analysis, CFA)等。方差分析和项目反应理论从探查测量条目是否存在文化偏差的角度来探讨测量对等性问题;验证性因素分析方法从总体上探查测量在不同文化间对等性及对等性的水平。在以下我们对每种方法逐一简单介绍。

方差分析的做法是将条目上的分值作为因变量,文化和量表总分(分为若干个组)作为自变量,进行方差分析,如果文化和总分的交互作用显著,即表示在不同文化之间,条目分值与总分的关系模式不一致,代表这个条目存在偏差。

另一种考察测量对等性的方法建基于项目反应理论(Item Response Theory, IRT)。项目反应理论

对被试在条目上的反应概率的模型进行分析(形成项目特征曲线, item characteristic curve, ICC), 由该条目的区分度(discrimination)、难度(difficulty level, 或称为普遍度, popularity)和猜测系数决定, 它们之间的关系为:

$$P_i(\theta) = c_i + (1 - c_i) \frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]} \quad (1)$$

由(1)式表达的项目特征曲线为一个 S 型的、上下两端分别趋近于 1 和 0(概率)的曲线。其中, θ 表示在该条目所测量的概念(如某种能力、态度等)上的高低, $P(\theta)$ 表示该条目在所测量的概念为 θ 时的概率(注意上述方程表达的是概率而不是概率密度)。D 是一个常数, 通常取 1 或 1.7; a 表示该条目的区分度, b 表示项目的难度(或普遍度); c 是猜测系数, 表示被试猜测作答的程度。

对于态度、信念等常见的心理学变量, 可以采用简化的 Rasch 模型 (Fischer & Molenaar, 1995), 即假设不存在猜测, 并且固定条目的区分度为 1:

$$P_i(\theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)} \quad (2)$$

使用项目反应理论分析条目跨文化偏差的思路为: 如果条目的区分度或难度在不同文化之间存在显著差异, 则表示这个条目存在文化偏差。在运用项目反应理论检验文化偏差的时候, 可以采用两种方式, 参数比较方式和模型比较方式。参数比较的方式是在各个文化下分别估计参数(区分度和难度)、然后比较它们在不同文化下的差异; 模型比较的方法是检验设定参数为跨文化相等的模型、与设定参数存在跨文化差异的模型之间, 在数据拟合上有无差异。

研究者还可以使用验证性因素分析方法探讨跨文化研究中的测量对等性问题。从验证性因素分析的角度, 跨文化研究的测量对等性是一个多样本比较的问题: 将不同文化作为不同的样本, 比较样本之间有关参数是否相同。验证性因素分析中假定存在潜变量结构、并估计潜变量与观测变量以及残差之间的各种关系。在多样本验证性因素分析中, 观测变量与潜变量、截距和残差之间存在以下关系:

$$X_k^g = \tau_k^g + \Lambda_k^g \xi^g + \delta_k^g \quad (3)$$

在等式(3)中, X_k^g 表示样本 g 中第 k 个观测变量, ξ 是其对应的潜变量, τ 表示截距, Λ 表示潜变量 ξ 对观测变量 X 的回归斜率, δ 表示残差。这个等式表示在组 g 中潜变量对观测变量的回归方程。

我们假定潜变量和残差是相互独立的, 上述等

式(1)可以转换为:

$$\Sigma^g = \Lambda_X^g \Phi^g \Lambda_X^{g'} + \Theta_\delta^g \quad (4)$$

在等式(4)中, Σ 表示观测变量的方差协方差矩阵, Φ 表示潜变量的方差协方差矩阵, Λ 表示潜变量到观测变量的系数矩阵, Θ 表示残差方差矩阵。讨论跨文化研究中测量对等性的问题, 从多样本验证性因素分析来看, 即探讨不同文化样本之间相关参数矩阵是否可以认为相等。如不能拒绝虚无假设, 就可以认为样本之间在某个层次上具有跨文化的测量对等性。

Vandenberg 和 Lance (2000) 总结了大量使用 CFA 进行测量对等性分析的研究, 并归纳了一系列的建议。我们将这些步骤列于下:

1) 分析不同文化样本之间观测变量的方差协方差矩阵是否全等? 如果这两者全等, 表明这个测量具有完全的跨文化一致性, 不需要进行以后的步骤。由于几乎没有可能在观测变量上找到具有跨文化全等性的例子, 很多研究并没有进行第一步检验。但由于确认整体全等等同于确认了测量对等性, 因此先进行此步检验不失为一个节省的方法;

2) 探讨不同文化样本之间是否具有“结构对等性(configural equivalence)”, 结构对等性即指不同文化样本之间是否可以认为有相似的因素结构, 如果不能确认结构对等性, 则不能进行进一步的测量对等性探讨;

3) 检验不同文化样本之间是否具有“量度对等性(metric equivalence)”, 即是否不同文化样本之间 Λ_X 可以认为是相等的;

4) 检验“截距对等性(scalar equivalence)”, 即不同文化样本之间 τ 是否相等;

5) 检验不同文化样本之间残差 Θ 是否相等;

6) 检验不同文化样本之间潜变量(因子)的方差、以及协方差是否相等(Φ), 后者探讨潜变量之间的相关是否具有对等性;

7) 检验不同文化样本之间潜变量均值是否相等。

3 跨文化研究中多层线性模型的应用

在跨文化研究当中, 研究者可能面对数据的层级结构问题。跨文化研究通常包括两个数据层级, 例如由被试的态度、观念等个体差异变量所构成个体水平上的数据层次, 同时, 不同文化之间在平均水平上可能存在差异, 构成了文化水平上的数据层次。在一个层级上得到的结论并不一定可以类推到

另一个层级中 (Klein & Kozlowski, 2000)。将在宏观层次上得到的结果直接类推到微观层次, 可能发生“区群谬误(ecological fallacy)”。同样的, 将微观层次上得到的结果直接类推到宏观层次, 可能出现所谓“微体谬误(atomistic fallacy)”。在不同数据层级上, 概念可能具有不同的内涵。例如, Triandis(1996)提出个人主义—集体主义维度在个人和国家层次上可能会具有不同的涵义; 在国家层次上, 他提出“个人主义—集体主义(individualism - collectivism)”的概念, 而在个人层次上, 他提出“个人中心—集体中心(idiocentrism - allocentrism)”的概念。

对于具有层级结构的数据, 研究者可以在不同层级上分别各自进行分析。现代统计技术发展了一些新的方法来处理存在不同层级的数据结构, 其中多层线性模型(hierarchical linear modeling, HLM, Raudenbush & Bryk, 2002)是处理跨文化研究中多层次数据的一种有力的分析工具(Schaffer & Riordan, 2003)。多层线性模型首先在较低层级的数据上建立回归方程, 然后以这个方程的截距和斜率作为因变量、选择较高层次上的变量作为自变量构成第二层回归方程, 并使用广义最小二乘法估计第二层参数。在多层线性模型中, 研究者可以同时多至3个层次的数据同时进行分析, 考察文化效应在不同层级上的效应、以及跨层级交互作用。

最近的一项研究 (Van Der Vegt, Van De Vliert, & Huang, 2005)在探讨成员多样性(demographic diversity)与创新氛围(innovative climate)之间关系的跨文化差异时, 就使用多层线性模型方法, 考察权力距离这个变量对成员多样性与创新氛围之间关系的调控作用。研究者收集到一家跨国公司的24个国家的248个分支的数据, 数据结构存在两个层级水平: 当地机构水平和文化水平。在多层线性模型分析中, 研究者对分支机构数据回归方程的斜率作为因变量、在文化水平上用权力距离作为自变量对其进行回归, 结果发现这个回归结果显著, 表明权力距离这个文化维度能够调控成员多样性与创新氛围之间的关系: 在权力距离较高的文化中, 组织的成员多样性会降低创新氛围; 而在权力距离较低的社会中, 组织成员多样性会提高创新氛围。

4 跨文化研究的发展趋势展望

本文简要介绍跨文化研究中值得关注的一些方法问题。在过去的几十年中, 跨文化研究在数量

和质量上都有很大的进步, 跨文化研究的意识愈发强烈, 很多研究者在不同文化中检验和拓展他们所提出的概念和理论, 并且提出了一系列更为精密的研究方法和数据处理方法。在中国, 随着跨文化交流日益增加和研究者对文化现象的愈加关注, 跨文化研究将会成为研究主流之一。作者认为心理和行为科学的跨文化研究将出现以下趋势: 1)新的文化维度结构将不断涌现, 近年来研究者提出了多个文化维度结构, 例如社会通则 (Leung & Bond, 2004)、认知风格 (cognitive style, 见 Nisbett, 2003)等, 可以预期未来将出现更多的、新的文化理论框架; 2)越来越多的新研究方法将被引入跨文化研究当中, 一直以来, 跨文化研究主要依赖纸笔调查和行为测量, 但是近期一些新的方法, 包括反应时、感知反应等, 可能将被应用于跨文化研究领域中; 3)研究者将更多地使用多种汇聚的方法来加强研究结果的解释力, 这些方法包括使用多种因变量指标、引入多个协变量、使用多种研究方法等; 4)最后, 一些新的数据分析方法, 例如多层线性模型, 多层结构方程模型等, 由于能提供更为复杂的数据模型, 将在跨文化研究中获得越来越广泛的重视和应用。

参 考 文 献

- 郑伯坝, 周丽芳, 樊景立. (2000). 家长式领导行为. *本土心理学研究*, 14, 3-64.
- 梁觉, 周帆. (2008). 跨文化研究的方法. 见 陈晓萍, 徐淑英, 樊景立 (编). *组织与管理研究的实证方法* (pp.385-408). 北京: 北京大学出版社.
- Brockner, J., Chen, Y. R., Mannix, E. A., Leung, K., & Skarlicki, D. P. (2000). Culture and procedural fairness: When the effects of what you do depend on how you do it. *Administrative Science Quarterly*, 45, 138-159.
- Caporael, L. R. (2001). Evolutionary psychology: Toward a unifying theory and a hybrid science. *Annual Review of Psychology*, 52, 607-628.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Rand McNally: Chicago.
- Chen, Y. R., Brockner, J., & Katz, T. (1998). Toward an explanation of cultural differences in in-group favoritism: The role of individual versus collective primacy. *Journal of Personality and Social Psychology*, 75, 1490-1502.
- Cheng, B. S., Chou, L. F., Wu, T. Y., Huang, M. P., & Farh, J. L. (2004). Paternalistic leadership and subordinate responses: Establishing a leadership model in Chinese organizations. *Asian Journal of Social Psychology*, 7, 89-117.
- Cheung, F. M., Leung, K., Zhang, J. X., Sun, H. F., Gan, Y. Q., Song, W. Z., et al. (2001). Indigenous Chinese personality constructs - Is the five-factor model complete? *Journal of*

- Cross-Cultural Psychology*, 32, 407–433.
- Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An “experimental ethnography”. *Journal of Personality and Social Psychology*, 70, 945–960.
- Farh, J. L., Zhong, C. B., & Organ, D. W. (2004). Organizational citizenship behavior in the People’s Republic of China. *Organization Science*, 15, 241–253.
- Fischer, G. H., & Molenaar, I. W. (1995). *Rasch Models: Foundations, recent developments, and applications*. Springer-Verlag: Berlin.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106, 766–794.
- Hofstede, G. (1980). *Culture’s consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Hong, Y. Y., Morris, M. W., Chiu, C. Y., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, 55, 709–720.
- House, R. J., & Javidan, M. (2004). Overview of GLOBE. In R. J. House, P. J., Paul, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, Leadership and Organizations: The GLOBE Study of 62 Societies* (pp. 9–28). Thousand Oaks, CA: Sage.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 269–309.
- Klein, K. J., & Kozlowski, S. W. J. (2000). From micro to meso: Critical steps in conceptualizing and conducting multilevel research. *Organizational Research Methods*, 3, 211–236.
- Leung, K., & Bond, M. H. (2004). Social axioms: A model for social beliefs in multicultural perspective. *Advances in Experimental Social Psychology*, 36, 119–197.
- Leung, K., Lau, S., & Lam, W. L. (1998). Parenting styles and academic achievement: A cross-cultural study. *Merrill-Palmer Quarterly*, 44, 157–172.
- Leung, K., & van de Vijver, F.J.R. (2008). Strategies for strengthening causal inferences in cross-cultural research: The consilience approach. *International Journal of Cross Cultural Management*, 8, 145–169.
- Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology*, 83, 693–702.
- Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently ... and why*. New York: Free Press.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.
- Serpell, R. (1993). *The significance of schooling: Life-journeys in an African society*. Cambridge, UK: Cambridge University Press.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, 6, 169–215.
- Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, 59, 325–338.
- Trafimow, D., Triandis, H. C., & Goto, S. G. (1991). Some tests of the distinction between private self and the collective self. *Journal of Personality and Social Psychology*, 60, 649–655.
- Triandis, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, 51, 407–415.
- Van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on cultures. *Journal of Cross-Cultural Psychology*, 31, 33–51.
- Van Der Vegt, G. S., Van De Vliert, E., & Huang, X. (2005). Location-level links between diversity and innovative climate depend on national power distance. *Academy of Management Journal*, 48, 1171–1182.
- Vandenberg, R. J., & Lance, C.E. (2000). A review and synthesis of measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, 3, 4–70.
- Wang, D., Cui, H., & Zhou, F. (2005). Measuring the personality of Chinese: QZPS versus NEO PI-R. *Asian Journal of Social Psychology*, 8, 97–122.

Cross-Cultural Research Methods: Review and Prospect

Kwok LEUNG¹, Fan ZHOU²

(¹Department of Management, City University of Hong Kong, Hong Kong, China)

(²School of Management, Zhejiang University, Hangzhou 310058, China)

Abstract

The study of cultural phenomena is a major topic for behavioral sciences. Cross-cultural research adopts a comparative perspective in analyzing cultural phenomena, and is confronted with several unique problems in research methods. This paper reviews the major methodological issues in cross-cultural research, covering topics in research design, measurement, and data analysis. From a design perspective, the paper begins with the use of consilience methods to strengthen causal inferences in cross-cultural research. The paper then explores issues of measurement invariance/equivalence in cross-cultural research and the associated analytic procedures. Finally, the paper reviews the application of hierarchical linear modeling in cross-cultural research as well as trends in the development of cross-cultural research methods.

Key words cross-cultural research; consilience approach; construct equivalence; measurement equivalence