

DOI: 10.13718/j.cnki.xdzk.2017.08.019

表述效应对平衡量表 内部一致性信度的影响^①

韦 嘉^{1,2}, 郭 磊², 张进辅²

1. 四川师范大学 教育科学学院, 成都 610068; 2. 西南大学 心理学部, 重庆 400715

摘要: 通过蒙特卡洛模拟研究考察平衡量表中源自条目不同计分方向的表述效应及量表内因素对测量结果内部一致性信度系数 α 系数和组合信度(CR) 系数的影响. 方差分析结果验证了研究假设: 表述效应使平衡量表的 α 系数与 CR 系数被高估, 高估程度随方法因子负荷升高 ($F_{(2, 392)} = 345.10, 344.87, p < 0.001$), 不同计分条目比例失衡 ($F_{(2, 392)} = 8.03, 8.32, p < 0.001$), 方法因子间相关 ($F_{(4, 392)} = 7.14, 7.02, p < 0.001$) 和特质因子负荷的降低 ($F_{(2, 392)} = 411.96, 410.49, p < 0.001$) 而加剧. 表明编制量表时, 须提升条目质量并尽量使不同计分条目数相等.

关键词: 表述效应; 平衡量表; α 系数; 组合信度; 蒙特卡洛模拟研究

中图分类号: B842

文献标志码: A

文章编号: 1673-9868(2017)08-0133-07

1 问题提出

1.1 内部一致性信度系数概述

信度即测量结果的可靠程度, 反映随机误差对测量结果的影响, 可从测量结果的一致性与稳定性² 方面进行评价^[1]. 测量结果的一致性反映了基于条目抽样产生的随机误差大小, 常用的评价指标为内部一致性信度系数.

内部一致性信度从操作定义看即测量同一特质条目的相关性, α 系数是其最常用的评价指标^[2-3]. α 系数可由条目间的平均相关系数求出^[4],

$$\alpha = k \cdot \bar{r}_{ij} / [k \cdot \bar{r}_{ij} + (1 - \bar{r}_{ij})] \quad (1)$$

其中, k 为条目数, \bar{r}_{ij} 为条目间的平均相关系数. 故 α 系数会随着: ① 条目数的增加而增大; ② 条目平均相关系数增大而增大. 此外, α 系数的计算还须满足条目间为正相关的理论假设^[4]. 因此, 为计算 α 系数及其他一些信效度指标, 研究者会先对量表中的反向计分条目得分进行反向编码^[5], 再计算 α 系数.

也有研究者^[6] 建议使用组合信度(composite reliability, CR) 系数来评价测量结果的内部一致性. CR 系数是在验证性因素分析框架下, 基于条目的因子负荷计算得到^[7],

$$\rho = (\sum \lambda_i)^2 / [(\sum \lambda_i)^2 + \sum \theta_{ii}] \quad (2)$$

其中: λ_i 为条目在公因子上的非标准化负荷, θ_{ii} 为非标准化的条目误差方差. 故 CR 系数会随条目因子负荷

① 收稿日期: 2016-07-16

基金项目: 国家自然科学基金项目(31400897); 教育部人文社会科学研究青年基金(15YJC190003).

作者简介: 韦 嘉(1981-), 男, 广东珠海人, 讲师, 博士, 主要从事心理与教育测量研究.

通信作者: 张进辅, 教授, 博士研究生导师.

的升高而增大。

1.2 表述效应与内部一致性信度系数概述

表述效应的操作定义为,平衡量表(同时包含直接计分与反向计分条目的量表)中采用同法测量(直接或反向计分)条目间所产生的非导因于欲测目标结构的额外共同变异^[8].其强度可通过因素分析框架下 2 个等价标准进行评价:条目的表述效应因子负荷大小和基于测量模型估计的表述效应因子方差。

先前研究者对表述效应的关注多聚焦于表述效应如何影响测量结果的结构效度,而较少关注表述效应对测量结果信度的影响.理论上,条目的测量误差间会因表述效应而产生共变,违反了计算 α 系数时条目误差彼此不相关的基本假设,因而会高估 α 系数^[6].从数理角度看,平衡量表即采用了不同方法测量相同特质,即同质异法(monotrait-monomethod, MTMM)^[9].2 个条目得分 X 和 Y 的相关系数可以由公式 3 算出:

$$\gamma_{xy} = \lambda_{Tx}\lambda_{Ty} + \lambda_{Mx}\lambda_{My}\varphi_{MxMy} \quad (3)$$

其中: $\lambda_{Tx}\lambda_{Ty}$ 是 2 个条目 X, Y 在共同特质因子上负荷的乘积, $\lambda_{Mx}\lambda_{My}$ 是 2 个条目在各自方法因子上负荷的乘积, φ_{MxMy} 是 2 个条目方法因子间的相关.在平衡量表中,若条目计分方向相同,则 φ_{MxMy} 为 1,条目的相关系数会在其真实相关(此处及后文所指的真实相关是指,在其他测量学条件不变的情况下的基础上增大 $\lambda_{Mx}\lambda_{My}$ (如条目的特质因子负荷),不存在表述效应时条目间的相关).若条目计分方向不同,反向计分后 φ_{MxMy} 变成正值(但通常小于 1,除非不同的表述效应间呈完全相关),此时条目的相关系数还是会膨胀 $\lambda_{Mx}\lambda_{My}\varphi_{MxMy}$,但增幅较同法条目间低.因此理论上异法条目间的相关低于同法条目间的相关,但二者均高估了条目间的真实相关^[1, 9-10].可见,平衡量表中,条目的平均相关系数会被高估,进而高估 α 系数。

Reise 等人^[11]的研究表明,用单因子模型拟合多因子数据时,条目的因子负荷会出现偏差.而出现偏差的原因是由于额外因子的存在使条目间出现了额外的相依性,可能高估条目的负荷,即在单因子模型下的条目因子负荷会更高^[12].

虽然 Reise 等人的研究选用的工具并非平衡量表(CAHPS2.0,均为直接计分条目),条目间额外的共变是由于相似的内容而非相同计分方向产生.但 CAHPS2.0 的不同因子间同为正相关关系,与反向计分后平衡量表不同表述效应因子的相关方向相同,故该结论可以谨慎地推论到平衡量表中.因此可以合理假设,当研究者基于单因子模型得到的因子负荷计算 CR 系数时,因子负荷同样可能被高估,导致 CR 系数的高估。

不仅在经典测量理论框架下,Wang 及其同事^[13]利用双因子项目反应模型在项目反应理论框架下同样证明表述效应会高估平衡量表的信度系数.此外,计分方向相同的条目也可以视为具有潜在的题组效应,忽略这些效应同样会高估信度^[14].

综合来看,平衡量表的内部一致性信度指标(α 系数与 CR 系数)会受表述效应的影响.单维平衡量表中,条目分数的平均相关系数与单因子模型条目的因子负荷均可能被高估,导致 α 系数与 CR 系数的高估.研究假设 H_1 :条目的方法因子负荷越高, α 系数与 CR 系数被高估的程度越大。

1.3 与表述效应相关的影响平衡量表内部一致性信度的量表内因素

除条目的方法因子负荷外,由公式 3 可知,还有其他与表述效应直接相关的量表内因素可能影响内部一致性系数。

首先,条目的特质因子负荷.无论表述效应存在与否,大小几何,条目的特质因子负荷理论上仍是信度系数的决定性因素^[9].如果条目的特质因子负荷足够高,表明真分数变异已占了总变异的很大比例,此时非真分数变异所占比例自然就相对减小,对信度系数的高估程度也就相应降低.研究假设 H_2 :条目的特质因子负荷越高, α 系数与 CR 系数被高估的程度越小。

其次,不同计分方向条目的比例与不同方法因子间的相关强度.在假设直接表述效应与反向表述效应强度相当的前提下,即公式 3 中的 $\lambda_{Mx} = \lambda_{My}$,所有的计分方向相同条目的相关系数会增大 λ_{Mx}^2 (λ_{My}^2),计分方向不同条目相关系数则增大 $\lambda_{Mx}^2\varphi_{MxMy}$ ($\lambda_{My}^2\varphi_{MxMy}$).以 10 条目的单维平衡量表为例,其观测分数一共会产

生 90 个相关系数. 反向计分后, 同法与异法的条目比例、同法与异法相关系数个数比例以及在方法因子间不同相关的情形下, 平均相关系数与不存在表述效应时的高估量见表 1 所示.

由表 1 可知, 表述效应强度恒定的前提下: ① 条目平均相关系数随直接计分条目与反向计分条目数量比例的失衡而增大; ② 2 种表述效应间的相关程度越高, 条目平均相关系数被高估的程度越高. 但无论 $\lambda_{M_x} = \lambda_{M_y}$ 的假设是否成立, 计分方向不同条目相关系数的增幅需要乘以表述效应因子间的相关系数 (通常小于 1, 而同法条目间的相关系数为 1), 因此全体相关系数的平均增幅与不同计分方向条目占总条目的比例呈反比. 即如果条目的相关系数矩阵中, 异法条目的相关系数比例越高, 全体相关系数 (平均相关系数) 膨胀的程度反而越小. 由于平均相关系数的高估程度改变是由于同法条目的相关系数与异法条目的相关系数的比例改变所致, 因此就 n 题平衡量表而言, “ m 道直接计分条目 + $(n-m)$ 道反向计分条目” (m, n 均为正整数且 $m < n$, 后同) 与 “ $(n-m)$ 道直接计分条目 + m 道反向计分条目” 是等效的. 研究假设 H_3 : 不同计分方向条目比例越失衡, α 系数与 CR 系数被高估的程度越大. 研究假设 H_4 : 不同方法间的相关系数越高 α 系数与 CR 系数被高估的程度越高.

表 1 10 条目单维平衡量表不同计分条目比例表述效应关系与平均相关系数的高估量

条目比 ^a	相关系数比 ^b	平均相关系数改变/ n^c				
		$\varphi_{M_x M_y} = 0.2$	$\varphi_{M_x M_y} = 0.4$	$\varphi_{M_x M_y} = 0.6$	$\varphi_{M_x M_y} = 0.8$	$\varphi_{M_x M_y} = 1$
5 : 5	40 : 50	50/90	60/90	70/90	80/90	1
6 : 4 (4 : 6)	42 : 48	51.6/90	61.2/90	70.8/90	80.4/90	1
7 : 3 (3 : 7)	48 : 42	56.4/90	64.8/90	73.2/90	81.6/90	1
8 : 2 (2 : 8)	58 : 32	64.4/90	70.8/90	77.2/90	83.6/90	1
9 : 1 (1 : 9)	81 : 9	82.8/90	84.6/90	86.4/90	88.2/90	1

注: a: 直接计分条目数; 反向计分条目数; b: 计分方向相同的条目相关系数个数; 计分方向相反的条目相关系数个数.

c: n 为常数 = $\lambda_{M_x}^2 = \lambda_{M_y}^2 = \lambda_{M_x} \lambda_{M_y}$.

此外, 信度系数的理论取值范围是 0~1, 但无论是 Cronbach^[4] 还是 Fornell 和 Larker^[7] 都没有给出 α 系数与 CR 系数优劣的判读标准. 后续研究者^[1] 依照 CTT 倾向于认为信度系数大于 0.70 代表优秀、大于 0.60 代表良好, 而大于 0.50 表示信度尚可. 实践中, 通常研究会保留信度系数小数点后两位到三位. 那么有理由认为当信度被高估超过 0.01 (例如由 0.59 膨胀了 0.01 到 0.60) 时, 就可能对测量结果内部一致性的结论产生质的影响. 因此, 如果表述效应会从统计上影响测量结果的 α 系数与 CR 系数, 那么需要进一步从单位的角度考虑, 表述效应在何种情况下才可能从务实层面影响研究者得出有具有质性差异的结论.

先前研究者利用核心自我评价量表初探了表述效应对信度系数的影响^[11], 但如果想通过实际调查来完成上述 4 个假设的检验, 其经济性值得研究者思考. 在此情况下, 模拟研究应该作为首选方法^[15]. 据此, 本研究拟用蒙特卡洛 (Monte Carlo) 模拟研究考察表述效应及其相关量表内因素对测量结果内部一致性信度指标 α 系数与 CR 系数的影响, 希望能借此发现更一般性的规律. 在实践中, 研究者可以通过增加样本量以抵消随机误差, 提升信度系数; 但在模拟研究中, 由于总体模型的参数是确定的 (包括条目残差), 此时基于总体模型的信度系数理论上不受样本量的影响. 由于蒙特卡洛法产生的模拟数据也会存在误差 (实际产生的数据与理论模型存在偏差), 基于此, 本研究也一并考察样本量的改变是否影响误设模型 (单维模型) 的信度系数 (不做理论假设, 属于探索性研究).

2 方 法

2.1 研究设计

本研究采用 $3 \times 3 \times 3 \times 5 \times 3$ 的完全随机设计. 自变量基于相关特质—相关方法模型产生, 共 10 个观测变量 (v_1-v_{10}), 代表条目得分. 每个观测变量受一个特质因子 (trait factor, TF)、一个表述效应因子 (straightforwardly wording effect, SW, 或 reversely wording effect, RW) 以及残差的影响. TF 与 SW 或 RW 间相关为 0, 但 SW 或 RW 间允许相关. 基于 Campbell 和 Fiske 提出的聚合效度与区分效度理论, 同

质异法条目的相关应大于异质同法条目的相关,违反该原则表明测量结果的结构效度较差,也无谓务实应用的价值.因此本研究在操作上,保证条目的特质因子负荷始终大于其方法因子负荷^[16].

本研究共 5 个自变量,对应研究假设 H_1-H_4 . 自变量 1 为条目的方法因子负荷,共 3 个水平: 0.10, 0.20 和 0.30. 自变量 2 为条目的特质因子负荷,共 3 个水平: 0.35, 0.55 和 0.71. 自变量 3 为不同计分方法条目数量比,共 3 个水平: 5:5, 7:3 和 9:1. 在 5:5 的水平下, v_1-v_5 受 SW 的影响, v_6-v_{10} 受 RW 的影响; 在 7:3 的水平下, v_1-v_7 受 SW 影响, v_8-v_{10} 受 RW 影响; 在 9:1 水平下, v_1-v_9 受 SW 影响, 仅 v_{10} 受 RW 影响(如前所述, SW 和 RW 影响的条目数可互换,对结果没有影响). 自变量 4 为反向计分后方法因子间的相关,共 5 个水平: 0, 0.20, 0.40, 0.60 和 0.80, 分别对应无相关、弱相关、中等相关、强相关和极强相关. 自变量 5 为样本量,共 3 个水平: 200, 500 和 1 000. 各不同的自变量组合均模拟 1 000 次,产生 1 000 个符合多元正态分布的模拟数据.

本研究的因变量基于 2 步产生. 首先,利用误设模型去拟合不同自变量水平下的 1 000 个模拟数据,并根据其结果计算出不同自变量水平下 10 00 个模拟数据的 α 系数(α 系数直接由自变量生产的模拟数据计算,等价于假设条目只测量了一个共同特质,且残差间无相关.)与 CR 系数的平均值 $mean\alpha_1$ 和 $mean\ CR_1$ (各 405 个). 误设模型为单因子模型,同样包括 10 个观测变量(v_1-v_{10}),但仅受一个特质因子(TF')与残差的影. 然后利用同样的单因子模型,基于自变量 2,5(不同样本量无表述效应存在时的真实信度系数)的不同水平各生成 1 000 个模拟样本,直接并计算这些样本的 α 系数与 CR 系数的平均值 $mean\alpha_2$ 和 $mean\ CR_2$ (各 9 个). 由于表述效应理论上会高估信度系数,因此匹配自变量 2,5 后,用 $mean\alpha_1$ 减去 $mean\alpha_2$, 用 $mean\ CR_1$ 减去 $mean\ CR_2$, 获得 $Dif-\alpha$ 与 $Dif-CR$ 作为本研究的因变量. 在自变量 2,5 相同的情况下,每种组合各有 $3\times 3\times 5$ 对 $mean\alpha_1$ 和 $mean\ CR_1$ (对应自变量 1,3,4 的不同组合),这 45 个不同的 $mean\alpha_1$ 和 $mean\ CR_1$ 会减去相同的 $mean\alpha_2$ 和 $mean\ CR_2$, 因此最终的因变量仍是 405 个 $Dif-\alpha$ 与 $Dif-CR$.

2.2 统计

本研究自变量和因变量的获得采用蒙特卡洛法,而分析则采用多因素方差分析. 由于不同的自变量水平下,因变量仅有一个 Dif 值,因此单元格离均差平方和为 0,需要采用效应可加模型进行分析,此时模型中没有交互项存在^[17].

2.3 研究工具

模拟数据利用 Mplus7.0 产生,方差分析则用 SPSS22.0 进行.

3 结果

不同的自变量水平(组间效应)下 $Dif-\alpha$ 与 $Dif-CR$ 均值差异的方差分析摘要见表 2 与表 3 所示. 从表 2 可以看出,5 个自变量除样本量外,在存在表述效应时,相较于无表述效应时,均会造成显著的 α 系数膨胀 ($p<0.001$). 从效应量 $partial\ \eta^2$ 的大小来看,其中方法因子负荷与特质因子负荷的效应量均超过了 0.138 的标准,属于高效量;不同计分方法条目比例的效应量则低于 0.059,属于低效应量;而方法因子相关的效应量低于 0.138,但高于 0.059,属于中等效应量^[18]. 从表 3 可以看出,CR 系数受各自变量的影响情况与 α 系数受各自变量的影响情况大体相同(统计显著性、效应量).

采用 Scheffe 法进行事后比较发现, $Dif-\alpha$ 与 $Dif-CR$ 的均值: ① 随方法因子的升高而升高 ($p<0.001$). ② 随特质因子负荷的升高而降低 ($p<0.001$). ③ 随不同计分条目比例失衡的程度增大而升高,就 $Dif-\alpha$ 而言,条目比为 5:5 与 7:3 时差异无统计学意义 ($p=0.611$),与 9:1 的差异有统计学意义 ($p=0.01$); 就 $Dif-CR$ 而言,条目比为 5:5 与 7:3 时差异无统计学意义 ($p=0.571$),与 9:1 的差异有统计学意义 ($p<0.001$). ④ 随方法因子间相关升高而升高,就 $Dif-\alpha$ 而言,仅水平 0 与水平 0.60 和 0.80 的差异有统计学意义 ($p=0.009$ 和 $p<0.001$),而水平 0.60 和 0.80 的差异无统计学意义 ($p=0.883$); 就 $Dif-CR$ 而言,仅水平 0 与水平 0.60 和 0.80 差异有统计学意义 ($p=0.012$ 和 $p<0.001$),而水平 0.60 和 0.80 的差异无统计学意义 ($p=0.868$).

表 2 有表述效应与无表述效应 α 系数差异的组间效应方差分析表摘要表

变异来源	平方和(type III)	自由度	均方	F	partial η^2
方法因子负荷	0.185	2	0.092	345.095***	0.638
特质因子负荷	0.221	2	0.110	411.963***	0.678
不同方法条目比例	0.004	2	0.002	8.029***	0.039
方法因子相关	0.008	4	0.002	7.140***	0.068
样本量	0 a	2	0 b	0	0
误差	0.105	392			
总变异	0.522	404			

注: a: 实际数值为 5.842e-008; b: 实际数值为 2.921e-008; ***: $p < 0.001$.

表 3 有表述效应与无表述效应 CR 系数差异的组间效应方差分析摘要表

变异来源	平方和(type III)	自由度	均方	F	partial η^2
方法因子负荷	0.184	2	0.092	344.865***	0.638
特质因子负荷	0.219	2	0.109	410.493***	0.677
不同方法条目比例	0.004	2	0.002	8.320***	0.041
方法因子相关	0.007	4	0.002	7.016***	0.067
样本量	0 a	2	0 b	0.004	0
误差	0.104	392			
总变异	0.518	404			

注: a: 实际数值为 1.914e-006; b: 实际数值为 9.568e-007; ***: $p < 0.001$.

如前所述, α 系数与 CR 系数是 0~1 的小数; 从务实角度看, 并非所有的高估均会对研究结论产生实质地影响。本研究中, 当四舍五入保留小数点后两位时, 仅 60 个 $Dif-\alpha$ 小于 0.01, 对应的自变量水平中, ① 方法因子负荷均为 0.10; ② 特质因子负荷均大于 0.35。同样仅有 60 个 $Dif-CR$ 在保留小数点后两位时小于 0.01, 且自变量组合情况与 $Dif-\alpha$ 一致。

4 讨 论

表述效应作为平衡量表一种主要共同方法变异源近年来不断受到研究者关注^[19]。本研究利用蒙特卡洛模拟研究考察了表述效应及与之密切相关的量表内测量学因素对测量结果 α 系数和 CR 系数的影响。通过效应可加模型多因素方差分析, 对 4 个研究假设进行了验证。

就 H_1 而言, 测量结果的 α 系数与 CR 系数相对于无表述效应时, 在不同方法因子负荷水平下差异有统计学意义。事后比较发现高估程度随表述效应的增强而升高, 且各水平下的均值差异有统计学意义。高效应量也表明表述效应的强度对 2 种信度系数的高估程度有实质且强烈的影响。研究假设 H_1 得到了验证。由于表述效应产生于平衡量表自身结构, 对量表编制者而言, 需要寻求与表述效应相关的心理行为变量, 间接对表述效应进行控制。

就 H_2 而言, 测量结果的 α 系数与 CR 系数相对于没有表述效应时, 在不同特质因子负荷下差异有统计学意义。事后比较发现二者的高估程度随特质因子负荷的升高而降低, 且各水平下的均值差异有统计学意义。高效应量同样表明特质效应的大小对 2 种信度系数的高估程度有着实质且强烈地影响。研究假设 H_2 得到了验证。这一结论提示后续研究者, 尽管理论上无法根除平衡量表中的表述效应, 但为了获得尽可能准确的信度系数, 就必须筛选那些更能反映欲测特质的条目来建构量表, 这样至少也可将表述效应对信度的影响降低。

就 H_3 而言, 测量结果的 α 系数与 CR 系数相对于没有表述效应时, 在直接计分与反向计分条目不同比例下差异有统计学意义。事后比较发现, 二者的高估程度随不同计分条目比例的失衡而升高; 但在本研究的 3 个自变量水平中, 仅当比例为 9:1 时, 高估的 α 系数与 CR 系数的均值与比例为 5:5(1:1) 及 (7:3) 时差异有统计意义。中等效应量也佐证了似乎只有当直接计分条目数量与反向计分条目数量出现很大差

异时,才会对实质提升对信度系数的高估. 研究假设 H_3 得到了部分验证. 先前研究者在是否应采用完全平衡量表(直接计分与反向计分条目数量相等的量表)的问题上,多进行的是理论上的分析论证^[8, 19-20],而本研究的这一发现为采用完全平衡量表提供了直接的实证证据支持,即完全平衡量表相对不同计分条目比例不等的平衡量表而言,在其他条件不变的前提下,对信度系数的高估程度最小.

就 H_4 而言,测量结果的 α 系数与 CR 系数相对于没有表述效应时,在不同的方法因子相关下差异有统计学意义. 事后比较发现,二者的高估程度随方法因子间相关的增加而升高,但仅在方法因子为高正相关时,才较方法因子无相关或中、低正向相关时加剧对信度系数的高估. 而该自变量的效应量较低,似乎表明方法因子相关对信度系数高估的实质影响并不大.

不同样本量下,本研究测量结果的 α 系数与 CR 系数相对于没有表述效应时的高估程度未发现具有统计学意义的差异. 这可能因为研究数据基于标准正态分布产生,随机因素对条目的影响强度是一个常数. 而实践中随机因素对条目的影响是变量,随着样本量增大,不同的随机效应会相抵消,即随机误差因样本量的增加而被不断消除,因而信度系数会得到理论上的相应提升. 据此研究者为了获得更可靠的结果在实测研究中,仍然应尽量采用大样本调查.

从务实层面看,本研究中,并非所有的自变量情形都会实质影响研究者对测量结果可靠性的判断,在某些情形下,例如条目的特征因子负荷相对较高,本研究中的 0.55 与 0.71,同时方法因子负荷相对较低 0.10,即便表述存在,其对信度系数的高估从数值上看,几乎不会使研究者或量表使用者对测量结果可靠性的判断发生质的改变.

5 结 论

- 1) 表述效应会造成 α 系数与 CR 系数的高估.
- 2) 模拟条件下,二者的高估程度会随表述效应的增强,特质效应的减弱,不同表述效应间关联程度的增加以及不同计分条目比例的失衡而增加.
- 3) 样本量在模拟条件下对 α 系数与 CR 系数的高估无实质的影响.
- 4) 仅部分自变量条件下,表述效应 α 系数与 CR 系数的高估会对测量结果的可靠性判断产生务实层面的影响.

参考文献:

- [1] 邱皓政. 量化研究法(三): 测验原理与量表发展技术 [M]. 台北: 双叶书廊, 2012.
- [2] 关守义. 克龙巴赫 α 系数研究述评 [J]. 心理科学, 2009, 32(3): 685-687.
- [3] 肖坤月. 信度计算公式初探 [J]. 西南师范大学学报(自然科学版), 1989, 14(4): 161-166.
- [4] CRONBACH L J. Coefficient Alpha and the Internal Structure of Tests [J]. Psychometrika, 1951, 16(3): 297-334.
- [5] VAUTIER S, POHL S. Do Balanced Scales Assess Bipolar Constructs? The Case of the STAI Scales [J]. Psychological Assessment, 2009, 21(2): 187-193.
- [6] 温忠麟, 叶宝娟. 测验信度估计: 从 α 系数到内部一致性信度 [J]. 心理学报, 2011, 43(7): 821-829.
- [7] FORNELL C, LARCKER D F. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error [J]. Journal of Marketing Research, 1981, 18(1): 39-50.
- [8] 韦 嘉. 基于单维平衡量表的表述效应研究——以人格测验为例 [D]. 重庆: 西南大学, 2016.
- [9] LANCE C E, DAWSON B, BIRKELBACH D, et al. Method Effects, Measurement Error, and Substantive Conclusions [J]. Organizational Research Methods, 2010, 13(3): 435-455.
- [10] CAMPBELL D T, FISKE D W. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix [J]. Psychological Bulletin, 1959, 56(2): 81-105.
- [11] 顾红磊, 温忠麟. 项目表述效应对自陈量表信效度的影响——以核心自我评价量表为例 [J]. 心理科学, 2014, 37(5): 1245-1252.
- [12] REISE S P, MORIZOT J, HAYS R D. The Role of the Bifactor Model in Resolving Dimensionality Issues in Health

- Outcomes Measures [J]. *Quality of Life Research*, 2007, 16(1): 19–31.
- [13] WANG, CHEN, JIN. Item Response Theory Models for Wording Effects in Mixed-Format Scales [J]. *Educational and Psychological Measurement*, 2015, 75(1): 157–178.
- [14] 詹沛达, 王文中, 王立君, 等. 多维题组效应 Rasch 模型 [J]. *心理学报*, 2014, 46(8): 1208–1222.
- [15] 温忠麟, 刘红云, 侯杰泰. 调节效应和中介效应分析 [M]. 北京: 教育科学出版社, 2012.
- [16] 韦 嘉, 张春雨, 赵永萍, 等. 随机截距因子分析模型在控制条目表述效应中的应用 [J]. *心理科学*, 39(4): 1005–1010.
- [17] 胡竹菁. 心理统计学 [M]. 北京: 高等教育出版社, 2010.
- [18] COHEN J. *Statistical Power Analysis for the Behavioral Sciences* [M]. Mahwah, NJ: Lawrence Erlbaum Associates, 1988.
- [19] PODSAKOFF P M, MACKENZIE S B, PODSAKOFF N P. Sources of Method Bias in Social Science Research and Recommendations on How to Control It [J]. *Annual Review of Psychology*, 2012, 63(1): 539–569.
- [20] RODEBAUGH T L, HEIMBERG R G, BROWN P J, et al. More Reasons to be Straightforward: Findings and Norms for Two Scales Relevant to Social Anxiety [J]. *Journal of Anxiety Disorders*, 2011, 25(5): 623–630.

The Influence of the Method Effect Associated with Item Wording on the Internal Consistency Reliability of Balanced Scale

WEI Jia^{1,2}, GUO Lei², ZHANG Jin-fu²

1. School of Educational Science of Sichuan Normal University, Chengdu 610066, China;

2. Faculty of Psychology, Southwest University, Chongqing 400715, China

Abstract: Monte Carlo simulation was made to investigate the laws and characteristics of how wording effect (or the Method Effect Associated with Item Wording) and its relevant endogenous factors affected the internal consistency reliability coefficient, Cronbach's α and composite reliability coefficient of the outcomes of measurement. Then multifactor variance analysis was made, in which five independent variables were manipulated: method factor loading of item, trait factor loading of item, ratio of different scoring items, correlation between different method factors and sample size. The results showed that both Cronbach's α and composite reliability coefficient in the balance scale would be overestimated due to the item wording effect and that the extent of overestimation increased with the increase in the item's method factor loading ($F(2, 392)=345.10$ and 344.87 , $p<0.001$) and the decrease in imbalance of different scoring items ($F(2, 392)=8.03$ and 8.32 , $p<0.001$), correlation between different method factors ($F(4, 392)=7.14$ and 7.02 , $p<0.001$), or item's trait factor loading ($F(2, 392)=411.96$ and 410.49 , $p<0.001$). The results suggested that the completely balanced scale was better under the condition of improving the psychometric properties of items.

Key words: the method effect associated with item wording; balanced scale; Cronbach's α ; composite reliability coefficient; Monte Carlo simulation

