

基于多层验证性因素分析的各种信度系数方法

田雪垠¹, 郑蝉金^{2*}, 郭少阳², 贺冠瑞³

(1. 复旦大学心理学系, 上海 200433; 2. 华东师范大学教育学部, 上海 200062; 3. 江西师范大学心理学院, 南昌 330022)

摘要:近年来,心理学研究的复现性受到广泛关注,许多实证研究难以重复验证,信度较低。大量研究使用多层技术,但只报告整体信度,导致研究可重复危机。基于各种信度系数和验证性因素分析的理论,以二层模型为例,总结多层信度计算方法,通过文献综述检索应用情况,并用 MPLUS 进行实例演示,最后讨论单层信度估计存在的偏差及分层估计的好处。总之,对多层数据进行分层信度估计是很有必要的,可消除因测量工具缺乏信度而导致的研究不可重复。

关键词:可重复危机;验证性因素分析;多层信度;MPLUS

中图分类号:B841.2

文献标识码:A

文章编号:1003-5184(2019)05-0461-07

1 引言

近年来,心理学研究的复现性受到广泛关注,许多实证研究难以重复验证,信度较低(Baker, 2015; Open Science Collaboration, 2015)。信度(reliability)是测量某种心理特质所得结果的一致性程度,反映随机误差因素所带来的方差变异。同时,信度也是衡量测验质量的重要指标,被定义为观察分数与真分数间相关的平方(Lord & Novick, 1968),或表示成真分数与观察分数的方差之比(McDonald, 1999)。信度越高,测量结果越可靠。

实际应用中,由于测验真分数难以得到,测验信度无法直接计算。故而,许多学者提出了不同信度系数来估计测验信度,例如克隆巴赫 α 系数(α 系数;Cronbach, 1951),合成信度 ω 系数(ω 系数;McDonald, 1970),最大信度 H 系数(H 系数;Thomson, 1940)等。其中, α 系数使用最广,在很长时间内,几乎被误认为信度本身。Guttman(1945)在数学上证明 α 系数是测验真实信度估计的下界,只有当题目残差间不相关,且满足基本 τ 等值的假设(即潜变量对所有题目的影响相同),其信度估计值才可靠。据Cho(2016)报告,在《Journal of Applied Psychology》(JAP)与《Academy of Management Journal》(AMJ)这两个心理学和管理学领域影响力较大的期刊中,没有文献在使用信度系数时报告过基本 τ 等值,因此无法判断这些文献中信度系数的使用是否正确。与 α 系数不同,由结构方程模型(structural equation modeling, SEM)中验证性因素分析(confirmatory factor analysis, CFA)计算而来的 ω 系数和 H 系数不受该条件限制。

同时,注意到以往研究大多局限于单层模型,若

测验存在一定数据结构,则不能反映测验的真实信度(Geldhof, Preacher, & Zyphur, 2014)。目前越来越多研究使用多层数据,并运用多层线性模型(hierarchical linear modeling, HLM)进行分析,但这些研究往往只报告测验的整体信度,未对信度进行分层讨论,隐含了单层结构的假设。实际上,多层数据模型各层存在独立的信度,只报告单层信度显然是不合理的,这会给心理学变量的测量带来较大误差,导致研究的可重复危机。

已知单层结构数据在CFA下进行信度估计是较为精确的,将CFA扩展至多层模型,可对各层的协方差矩阵进行单独分析。因此,利用多层验证性因素分析(multilevel confirmatory factor analysis, MCFA)估计多层信度也是精确的。

基于MCFA,信度估计得到一系列发展。其中, Raykov 和 Du Toit(2005)提出多层信度的点估计和区间估计,考虑了组内和组间层次结构,但只提供整体信度估计,不能有效地分析多层信度。Cranford等(2006)也针对多层信度展开了讨论,提出多层信度估计理论,但不足以证明各层间有独立的真分数和误差方差。

在教育研究领域,普遍存在学生嵌套于班级,同时班级又嵌套于学校的模型。Rantanen(2013)指出在某些情况下,这些层次结构会更复杂,观测数据不一定相互嵌套,可能同时有多个层次结构。虽然该研究可解释组内和组间层次的变异程度,但仍未对信度进行分层讨论,不能反映测量工具的稳定性。

近几年研究中,复杂层次结构分析更易操作。Geldhof等(2014)基于MCFA对多层信度展开讨论,给出计算方法,并利用模拟数据进行验证,推动了多

* 通讯作者:郑蝉金, E-mail: noblesavage@sina.com。

层信度的发展。

综上,文章主要从以下部分展开论述:(1)探讨 CFA 下单层信度估计;(2)简介多层信度概念,在 MCFA 下估计信度;(3)检索国内外核心期刊,分析多层信度的应用情况;(4)用 MPLUS 进行实例演示,探究多层信度估计的适用性。

2 基于验证性因素分析的单层信度估计

信度被定义为真分数与观测分数方差的比值,用公式表示为:

$$\rho_x = \frac{\sigma_T^2}{\sigma_x^2} = 1 - \frac{\sigma_E^2}{\sigma_x^2} \quad (1)$$

其中 σ_x^2 表示总分方差, σ_T^2 表示真分数方差, σ_E^2 表示误差方差。分母中总分方差可直接计算,但分子中真分数或误差方差不能直接得到,要借助其他方法估计,各种信度系数的区别就在于估计方法不同。

CFA 是 SEM 的重要组成部分,主要处理观测指标与潜变量间关系,被称为测量模型(侯杰泰,温忠麟,成子娟,2004)。下面以 α 系数、 ω 系数和 H 系数这三种信度系数为例,讨论 CFA 下的信度估计。

2.1 克隆巴赫 α 系数

α 系数是信度的下限估计,只有满足特定条件, α 系数估计结果才是实际信度(温忠麟,叶宝娟,2011)。当模型是不含潜变量的完全饱和和协方差结构时,可用 CFA 进行信度估计。

α 系数是测验所有题目平均协方差($\bar{\sigma}_{ij}$)、总分方差(σ_x^2)和题量(n)的函数,公式表示如下:

$$\alpha = \frac{n^2 \bar{\sigma}_{ij}}{\sigma_x^2} \quad (2)$$

其中,平均协方差通过将矩阵 Σ (题目对称矩阵 σ_{ij}^2 和协方差矩阵 σ_{ij}) 中所有元素相加,再除以元素个数得到。总分方差 σ_x^2 是所有题目方差与两倍协方差之和(即协方差矩阵所有元素和),用公式表示为:

$$\sigma_x^2 = 1' \Sigma 1 \quad (3)$$

上式中 1 和 $1'$, 分别表示所有元素均为 1 的行向量和列向量,其行列数与矩阵 Σ 相等。

2.2 合成信度 ω 系数

ω 系数是在 SEM 基础上提出的,允许题目与其对应因子间具有异质相关性,即有不同的因子载荷,比 α 系数更精确(吴瑞林,袁克海,2012;叶宝娟,温忠麟,2011)。

当测验为标准化的单维结构时, ω 系数可用以下公式估计:

$$\omega = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k \theta_{ii}} \quad (4)$$

其中, λ_i 指题目 i 的因子载荷, θ_{ii} 指题目 i 的方差;分子部分等价于 $1' \Lambda \Lambda' 1$, 即题目真分数所有潜在模型协方差矩阵之和,分母部分表示真分数方差与所有残差方差之和。

计算合成信度 ω 系数时,无法得知因素载荷、因子间协方差及题目残差方差的真实值。实际代入式中的参数均为 SEM 下的估计值 $\hat{\Lambda}$ 、 $\hat{\Phi}$ 、 $\hat{\Psi}$, 其与具体模型及样本有关。

当测验满足基本 τ 等价时, ω 系数和 α 系数在数值上相等,即 α 系数是 ω 系数在一定条件下的特例。

2.3 最大信度 H 系数

与 ω 系数类似, H 系数是基于 SEM 计算测验总分的估计方法,但题目加权方式不同。

H 系数法具有最佳加权合成分数(Bentler, 2007; Raykov, 2004), 公式可表示为:

$$H = \frac{\sum_{i=1}^k \frac{l_i^2}{1 - l_i^2}}{1 + \sum_{i=1}^k \frac{l_i^2}{1 - l_i^2}} \quad (5)$$

其中, l_i^2 表示单维下题目 i 的完全标准化因子载荷。Hancock 和 Mueller(2001) 将上式简化为:

$$H = \left(1 + \frac{1}{\sum_{i=1}^k \frac{l_i^2}{1 - l_i^2}} \right)^{-1} \quad (6)$$

H 系数提供题目最佳加权,且是平方项,具有一定优点。首先, ω 系数受到负因子载荷影响,但 H 系数可使其有意义。其次, H 系数在计算总分时对题目进行最优加权,计算出的信度值最大。最后, ω 系数对所有题目进行等加权,受到质量差的题目影响,而 H 系数对大信息量题赋予高权重,反之亦然。因此, H 系数可在一定程度上减少误差的干扰(叶宝娟,杨强,2011; Penev & Raykov, 2006)。

3 多层信度

3.1 多层验证性因素分析

以二层模型为例,对多层信度估计展开讨论。把观察分数(y_{ik} , 下标 i 和 k 分别表示组内和组间层次)划分为组内真分数、组间真分数、组内误差和组间误差。各层具有不同真分数方差,故应对各层进行独立信度估计。

其中,组内信度是组内真分数方差与总方差的比值,计算公式为:

$$\rho_{wi} = \frac{\sigma_{T_{wi}}^2}{\sigma_{T_{wi}}^2 + \sigma_{E_{wi}}^2} \quad (7)$$

同理,组间信度用组间真分数方差在组间总方差中所占比例表示,公式如下:

$$\rho_{bk} = \frac{\sigma_{T_{bk}}^2}{\sigma_{T_{bk}}^2 + \sigma_{E_{bk}}^2} \quad (8)$$

此外,组内与组间方差的解释程度可用组内相关系数(intraclass correlation coefficient, ICC)描述,该系数表示组间方差占总变异的比值,公式如下:

$$\rho_{wi} = \frac{\sigma_{T_{bk}}^2 + \sigma_{E_{bk}}^2}{\sigma_{T_{wi}}^2 + \sigma_{E_{wi}}^2 + \sigma_{T_{bk}}^2 + \sigma_{E_{bk}}^2} \quad (9)$$

ICC通常作为HLM的评判指标,应用于多层结构的检验(Heck & Thomas, 2015)。当题目ICC较小,如小于0.05时,估计的残差方差几乎为0,可忽略组间差异对整体的影响,只计算整体信度(段锦云,王娟娟,朱月龙,2014)。当相应的组间真分数方差也接近0时,接近0的残差方差在组间变异中占很大比例,不可忽略组间层次的影响(Hox, van de Schoot, & Matthijsse, 2012)。需要强调的是,组间信度与ICC是不同概念,组间信度反映了个体观察分数在组间层次的差异,ICC则反映了分组所带来的变异,通过公式也能看出两者的差异。

3.2 多层信度的计算

随着对多层结构模型的探索,多层信度的理论越来越完善,其中Geldhof等(2014)对多层信度系数的估计进行了详细探讨。下面对不同信度系数在MCFA下的估计进行总结。

3.2.1 多层 α 系数

根据 α 系数估计法及MCFA的要求,多层数据必须满足完全饱和的协方差模型。

将等式(2)分别应用于组内和组间层次,得到独立的组内及组间 α 系数。各层 α 系数的分子是该层题数平方乘以该层平均协方差;分母是该层完全协方差矩阵中所有元素之和,可通过该层所有题目方差加上两倍的独立协方差得到。

在二层结构数据的信度估计中,各层计算公式分别表示为:

Level-1:

$$\alpha_{wi} = \frac{m^2 \overline{\sigma_i}}{\sigma_i^2} \quad (10)$$

Level-2:

$$\alpha_{bk} = \frac{n^2 \overline{\sigma_k}}{\sigma_k^2} \quad (11)$$

其中, σ_i 、 σ_k 分别表示组内和组间层次所有题目的协方差矩阵, m 和 n 表示组内和组间层次的题数。

3.2.2 多层 ω 系数

当各层均满足单维结构时,可用MCFA估计多层 ω 系数。将等式(4)应用于各层,则各层的 ω 系数分别是各层题目因子载荷及其残差方差的函数。该方法要求分别估计各层的残差方差,除非有充分理由说明没必要进行组间信度估计,否则不能将组间残差方差固定为0(Gottfredson, Panter, Daye, Allen, & Wightman, 2009)。

在二层结构数据中,各层计算公式分别表示为:

Level-1:

$$\omega_{wi} = \frac{(\sum_{i=1}^m \lambda_i)^2}{(\sum_{i=1}^m \lambda_i)^2 + \sum_{i=1}^m \theta_i} \quad (12)$$

Level-2:

$$\omega_{bk} = \frac{(\sum_{k=1}^n \lambda_k)^2}{(\sum_{k=1}^n \lambda_k)^2 + \sum_{k=1}^n \theta_k} \quad (13)$$

其中, λ_i 、 λ_k 分别表示组内和组间层次题目 i 和 k 的因子载荷, θ_i 和 θ_k 表示题目 i 和 k 所对应的方差, m 和 n 是组内和组间层次的题数。

3.2.3 多层 H 系数

在多层数据结构下,估计 ω 系数的方法同样适用于 H 系数。

H 系数根据标准化因子载荷的平方(用 l_i^2 表示)来估计信度,排除负因子载荷的干扰,并提供各题的最佳加权。但MCFA一般不提供标准化的因子载荷,要先用原始矩阵进行参数估计,公式如下:

$$\frac{l_i^2}{1 - l_i^2} = \frac{\lambda_i^2}{\sigma_i^2 - \lambda_i^2} \approx \frac{\lambda_i^2}{\theta_i} \quad (14)$$

在二层结构数据中,各层计算公式分别表示为:

Level-1:

$$H_{wi} = \left(1 + \frac{1}{\sum_{i=1}^m \frac{\lambda_i^2}{\theta_i}}\right)^{-1} \quad (15)$$

Level-2:

$$H_{bk} = \left(1 + \frac{1}{\sum_{k=1}^n \frac{\lambda_k^2}{\theta_k}}\right)^{-1} \quad (16)$$

其中, λ_i 、 λ_k 分别表示组内和组间层次题目 i 和 k 的因子载荷, θ_i 和 θ_k 表示题目 i 和 k 所对应的方差, m 和 n 是组内和组间层次的题数。

4 多层信度的应用情况

为评估多层信度在心理学研究中的应用情况,借助“中国知网”、“Web of Science”和“Google Scholar”等平台对五本国内外核心期刊近五年

(2014 年 1 月至 2018 年 4 月)的文章进行检索。

4.1 检索期刊

选取期刊分别是《Journal of Applied Psychology》(JAP)、《Journal of Organizational Behavior》(JOB)、《心理学报》、《心理科学》和《心理发展与教育》。

其中,JAP 和 JOB 是应用心理学核心外刊,具有较高影响力。“Web of Science”期刊引用报告显示,2017 年 JAP 的影响因子为 4.643,在应用心理学类中排第 6,管理学类中排第 23;JOB 的影响因子为 4.229,在应用心理学类位列第 7,管理学类位列第 30。《心理学报》、《心理科学》及《心理发展与教育》是国内心理学核心期刊,在中国知网“2017 中国最具国际影响力学术期刊(人文社会科学)”排名前三。

4.2 检索过程

首先,用“Web of Science”和“中国知网”检索近五年的文章。根据标题、关键词,通过“Google Scholar”对文献综述和概念论文进行排查,核对期刊官网,统计各期刊应用文章总数。

接着,利用关键词检索,统计使用多层技术的文章数量。外刊主要借助“Google Scholar”高级搜索功能。考虑到 ICC 是 HLM 的评判标准,使用“ICC”、“HLM”作为关键词,浏览标题、摘要,必要时查看原文,确定文章数量。中文期刊借助“Google Scholar”和“中国知网”检索。在“Google Scholar”高级搜索中用“多层”关键词查找,并用“ICC”关键词再次验证。同时,在“中国知网”中结合“跨层”、“多层”、“ICC”等关键词检索,综合多次结果进行统计。

最后,统计分层计算信度的文章数量。应用研究必定会引用相应的理论文章,故以多层信度的核心理论文章为中心,对其参考文献、引证文献和相关文献进行扩展搜索。在此选取 Raykov 和 Marcoulides(2006)及 Geldhof 等(2014)的研究,这两篇文章是多层研究的重要理论来源,介绍了 SEM 下的多层信度概念及如何计算多层信度。通过多个平台交叉检索,确定数量。

4.3 检索结果

对各期刊进行检索后,整理各部分结果如下表 1。

表 1 五本期刊多层信度的应用情况

	JAP n = 397		JOB n = 251		《心理学报》 n = 552		《心理科学》 n = 728		《心理发展与教育》 n = 345	
	n	%	n	%	n	%	n	%	n	%
运用多层技术文章	142	35.77	95	37.85	66	11.96	30	4.12	24	6.96
计算多层信度文章	1	0.25	3	1.20	0	0	0	0	0	0

由数据可见,各期刊应用研究文章总数分别为:JAP397 篇,JOB251 篇,《心理学报》552 篇,《心理科学》728 篇,《心理发展与教育》345 篇。

其中,使用多层技术的情况如下:JAP142 篇,占期刊总量的 35.77%;JOB95 篇,占期刊总数的 37.85%;《心理学报》66 篇,在期刊中占比 11.96%;《心理科学》30 篇,占期刊的 4.12%;《心理发展与教育》24 篇,在期刊中有 6.96% 的比例。中文期刊含较多实验研究,故文章比例较少。由此可见,多层技术的文章数量相对可观,说明其具有一定普遍性。

然而,在这五本期刊中,仅有 4 篇报告了多层信度:JAP 中 1 篇研究团队绩效文章(Koopmann, Lanaj, Wang, Zhou, & Shi, 2016),以及 JOB 中 3 篇应用文章——研究心理契约模式对组织承诺恢复影响(Solinger, Hofmans, Bal, & Jansen, 2016);情绪智力和情感过程中的个体差异(Minbashian, Beckmann, & Wood, 2017);感知掌握气氛、感受信任和知识共享的关系(Nerstad et al., 2018)。其他文章均未分层报告信度。

因此,即使多层结构在技术层面得到推进,但很

少有学者根据多层理念,对信度进行分层讨论。

5 实例

利用《An Introduction to Multilevel Modeling Techniques》(Heck & Thomas, 2015)第八章数据在 MPLUS 中进行演示。该例是关于员工组织领导力的研究,被试覆盖 36 个区,是来自 105 个组织的 650 名员工。通过六个测量指标,两个潜在因子,构建二层双因子模型。

Heck 和 Thomas(2015)给出具体模型如图 1 所示,有两个潜在因子(决策风格 DEC,评估程序 EVAL),每个因子指向三个观测指标,且因子间相关。决策风格包括管理者参与度(Shdec),客户参与度(Invcli),团队领导力(Team)。评估程序包括管理者评估(Evprog),使用和系统评估(Sysas),员工绩效标准(Evstan)。

不论哪种信度系数估计法,在多层测量模型下各层均能解释足够的变异(即 ICC > 0.05)。因此,可在 MPLUS 中用 MCFA 对数据进行分析。下面针对不同系数估计法,进行整体信度估计和分层信度估计,比较异同。

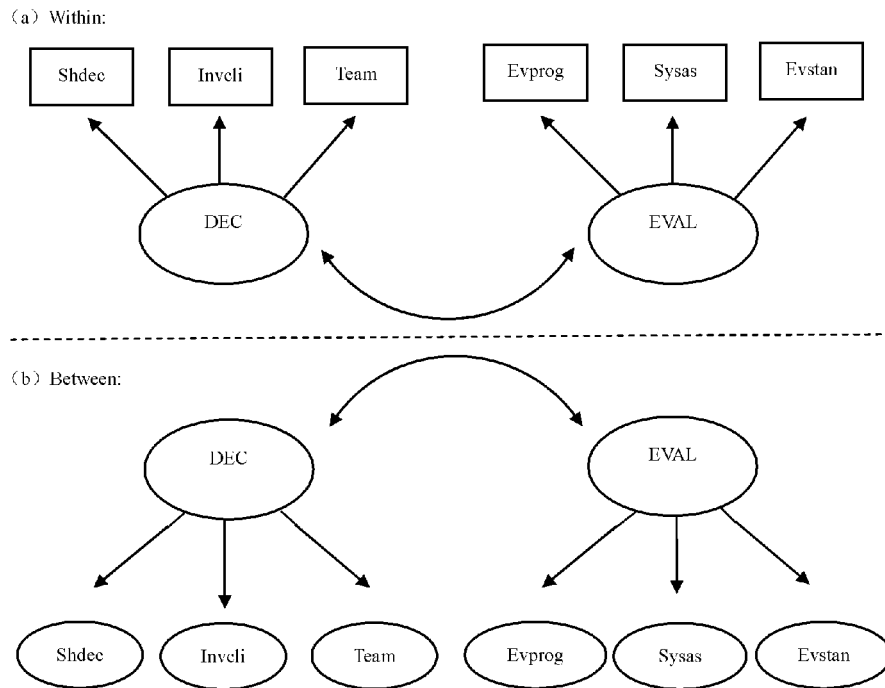


图1 二层两因子模型

单层 α 系数基于单层饱和协方差模型,将单位合成分数和 α 系数作为模型参数。在MPLUS中测验合成分数的方差被定义为:

$$\text{COMP_V} = V1 + V2 + V3 + V4 + V5 + V6 + 2 * (C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10 + C11 + C12 + C13 + C14 + C15);$$

其中, $V1 - V6$ 表示各题方差, $C1 - C15$ 表示各题间协方差,利用这些题目协方差和测验合成分数(COMP_V)可计算单层 α 系数(即整体信度),在MPLUS计算公式为:

$$\text{ALPHA} = (((C1 + C2 + C3 + C4 + C5 + C6 + C7 + C8 + C9 + C10 + C11 + C12 + C13 + C14 + C15) / 15) * 36) / \text{COMP_V};$$

根据公式和代码得到单层 α 系数为0.850,说明具有良好的整体信度。

将数据应用于二层模型,估计测验合成分数及组内、组间层次的 α 系数,结果显示各层信度均可接受,且组间信度($\alpha = 0.933$)比组内信度($\alpha = 0.825$)更可靠。

接着,在单维CFA下,用因子载荷和残差方差来估计模型参数。MPLUS中 ω 系数和 H 系数的计算公式分别为:

$$\omega = (L1 + L2 + L3 + L4 + L5 + L6) * * 2 / (((L1 + L2 + L3 + L4 + L5 + L6) * * 2) + (R1 + R2 + R3 + R4 + R5 + R6));$$

$$H = 1 / (1 + (1 / ((L1 * * 2 / R1) + (L2 * * 2 / R2) + (L3 * * 2 / R3) + (L4 * * 2 / R4) + (L5 * * 2 / R5) + (L6 * * 2 / R6))));$$

其中, $L1 - L6$ 表示各题未标准化的因子载荷, $R1 - R6$ 表示估计的残差方差。根据公式和代码计算得到单层 ω 系数为0.854,单层 H 系数为0.865,两者数值相近,均可接受。结果显示,相比单层 α 系数, ω 系数和 H 系数偏高。

将数据应用于MCFA,规定所有残差方差大于0,估计组内、组间层次的信度。结果显示, ω 系数及 H 系数的组间信度($\omega = 0.938$; $H = 0.978$)也比组内信度($\omega = 0.830$; $H = 0.838$)更可靠。

将数据应用于MCFA,规定所有残差方差大于0,估计组内、组间层次的信度。结果显示, ω 系数及 H 系数的组间信度($\omega = 0.938$; $H = 0.978$)也比组内信度($\omega = 0.830$; $H = 0.838$)更可靠。

相关代码请扫描附录二维码或发邮件获取。

6 讨论

综上所述,越来越多研究应用了多层数据,但大多关注的是跨层的理论模型。在此情况下,若只报告整体信度,未进行分层讨论,所用测量工具的可靠性就难以评估,可能会引入一些混淆变量,在一定程度上导致研究不可重复。例如,跨层的交互作用、中介调节等,通过量表收集得到的结果可能会受到质疑,因为此时测量工具不符合多层结构。因此,当数据具有多层结构时,应用HLM分析,报告分层的信度估计结果,以缓解可重复危机。

从理论角度考虑,第一,以二层模型为例,HLM考虑了两种变异,即组内和组间变异,能解释更多的残差方差,在一定程度上可改善参数检验的结果。例如,整体信度只含一组整体参数,而多层信度中至少包含两组参数。第二,多层信度提供更多信息。当实际数据为多层结构时,整体信度不足以提供充

分信息,忽略了各层间的变异;而多层信度通过计算模型的 ICC,对数据分层讨论,报告各层信度,提供更多可靠信息。

从文献检索结果来看,多层技术得到普及,越来越多研究使用多层数据,并用 HLM 分析。然而,检索的期刊中仅有 4 篇文章报告了相应的多层信度,其他文章均未对信度进行分层讨论。由此可见,虽然多层信度的概念被接受,其理论也被相继讨论,但真正报告多层信度的研究很少,这显然是不合理的。

从应用层面出发,多层技术的理论已相对成熟,但很少有人在实际研究中对多层数据进行分层信度报告。即使使用多层数据,且 ICC 符合 HLM 的判定标准,大部分研究仍尽量回避分层信度估计,只报告整体信度。这可能是测量模型的要求较严格,检验较难通过。例如,MCFA 难以估计较小样本量,但多层回归却可以。因此,有些研究为得到预期结果,会放松对测量模型的检验,直接进行回归分析。此外,MCFA 虽然不能同时拟合所有分测验,但可对测验进行信度估计,故基于 MCFA 的多层信度估计的可靠性远优于不报告信度或单层假设下的整体信度估计。Cho(2016)也提到在对合成分数进行多层分析时,借助 MCFA 分层估计信度是很有必要的。

从实例结果分析,第一,当数据是多层结构时,仅借助单层假设估计整体信度会造成较大偏差,应考虑分层分析。第二,单层 ω 系数和 H 系数的估计结果都比 α 系数高,重复验证了 α 系数在特定条件下是信度的下限估计。第三,最大信度 H 系数与合成信度 ω 系数的方法类似,只是加权方式不同,因而两种信度系数估计法所得结果相近。第四,实例计算中,不论是哪种信度系数,其单层信度都更接近组内信度,说明当组内变异较大(ICC 较小, $ICC_{\alpha} = 0.127$, $ICC_{\omega/H} = 0.253$)时,整体信度估计更接近组内信度。

最后,文章讨论的前提是第一层满足单维且固定因子载荷的 CFA 模型,不考虑其他因子结构在不同水平上的变化。其次,研究的可重复危机包括很多方面,例如各种变量的控制和标准化等,在此探讨的是测量工具的质量保障,即保证信度的准确估计。因此,在多层数据中,对信度进行分层讨论只是在一定程度上缓解可重复危机,不能完全消除其存在的可能性。尽管有这些限制,文章总结了常见的几种信度系数方法,对多层信度的发展和应用程度进行综述,解释了单层信度估计可能存在的偏差及分层估计的好处,希望能对多层信度的推进提供一些帮助。

参考文献

- 段锦云,王娟娟,朱月龙.(2014).组织氛围研究:概念测量,理论基础及评价展望.心理科学进展,22(12),1964-1974.
- 侯杰泰,温忠麟,成子娟.(2004).结构方程模型及其应用.北京:教育科学出版社.
- 温忠麟,叶宝娟.(2011).测验信度估计:从 α 系数到内部一致性信度.心理学报,43(7),821-829.
- 吴瑞林,袁克海.(2012).基于结构方程模型的合成信度及其使用问题研究.统计与信息论坛,27(12),14-20.
- 叶宝娟,温忠麟.(2011).单维测验合成信度三种区间估计的比较.心理学报,43(4),453-461.
- 叶宝娟,杨强.(2011).用验证性因子分析估计单维测验的信度.教育测量与评价:理论版,11,8-12.
- Baker, M. (2015) Over half of psychology studies fail reproducibility test. Nature News. Retrieved July, 30, 2018, from www.nature.com/news/over-half-of-psychology-studies-fail-reproducibilitytest-1.18248.
- Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S.-Y. Lee (Ed.), *Handbook of Latent Variable and Related Models* (pp. 1-19). Netherlands: North-Holland.
- Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, 19(4), 651-682.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3), 297-334.
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, 32(7), 917-929.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91.
- Gottfredson, N. C., Panter, A. T., Daye, C. E., Allen, W. F., & Wightman, L. F. (2009). The effects of educational diversity in a national sample of law students: Fitting multilevel latent variable models in data with categorical indicators. *Multivariate Behavioral Research*, 44(3), 305-331.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, K. G. Jöreskog, & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A festschrift in honor of Karl Jöreskog* (pp. 195-216). Lincolnwood, IL: Scientific Software International.
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multi-level modeling techniques: MLM and SEM approaches using Mplus*. New York: Routledge.
- Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few

- countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*, 6(2), 87–93.
- Koopmann, J., Lanaj, K., Wang, M., Zhou, L., & Shi, J. (2016). Nonlinear effects of team tenure on team psychological safety climate and climate strength: Implications for average team member performance. *Journal of Applied Psychology*, 101(7), 940–957.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison – Wesley.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23(1), 1–21.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Minbashian, A., Beckmann, N., & Wood, R. E. (2017). Emotional intelligence and individual differences in affective processes underlying task – contingent conscientiousness. *Journal of Organizational Behavior*. Retrieved July, 30, 2018, from <https://doi.org/10.1002/job.2233>.
- Nerstad, C. G., Searle, R., Černe, M., Dysvik, A., Škerlavaj, M., & Scherer, R. (2018). Perceived mastery climate, felt trust, and knowledge sharing. *Journal of Organizational Behavior*, 39(4), 429–447.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–1 – aac4716–8.
- Penev, S., & Raykov, T. (2006). On the relationship between maximal reliability and maximal validity of linear composites. *Multivariate Behavioral Research*, 41(2), 105–126.
- Rantanen, P. (2013). The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*, 38(2), 224–239.
- Raykov, T. (2004). Estimation of maximal reliability: A note on a covariance structure modelling approach. *British Journal of Mathematical and Statistical Psychology*, 57(1), 21–27.
- Raykov, T., & Du Toit, S. H. (2005). Estimation of reliability for multiple – component measuring instruments in hierarchical designs. *Structural Equation Modeling*, 12(4), 536–550.
- Raykov, T., & Marcoulides, G. A. (2006). On multilevel model reliability estimation from the perspective of structural equation modeling. *Structural Equation Modeling*, 13(1), 130–141.
- Solinger, O. N., Hofmans, J., Bal, P. M., & Jansen, P. G. (2016). Bouncing back from psychological contract breach: How commitment recovers over time. *Journal of Organizational Behavior*, 37(4), 494–514.
- Thomson, G. H. (1940). Weighting for battery reliability and prediction. *British Journal of Psychology*, 30(4), 357–366.

Various Reliability Coefficient Methods Under Multilevel Confirmatory Factor Analysis

Tian Xueyin¹, Zheng Chanjin², Guo Shaoyang², He Guanrui³

(1. Department of Psychology, Fudan University, Shanghai 200433;

2. Faculty of Education, East China Normal University, Shanghai 200062;

3. School of Psychology, Jiangxi Normal University, Nanchang 330022)

Abstract: In recent years, the reproducibility of psychological science has received extensive attention. Many empirical studies were difficult to replicate and therefore had low reliability. It is noted that more and more researchers used a multilevel data structure and hierarchical linear modeling (HLM) for analysis. However, most of them only displayed the overall reliability, resulting in the research reproducibility crisis. Based on the confirmatory factor analysis (CFA), the article introduced estimation methods of different reliability coefficients. Taking a two – level model as an example, it also summarized the multilevel reliability estimation methods. In addition, to evaluate the application of multilevel reliability, the article investigated the researches in the past five years. Then, it illustrated how to use the software MPLUS to execute reliability estimation with an example. Finally, the article discussed the deviation of single – level reliability estimation and the benefits of hierarchical estimation. In summary, it is necessary to perform hierarchical reliability estimation in the multilevel data structure, which can eliminate the reproducibility crisis caused by unreliable measurement tools.

Key words: reproducibility crisis; confirmatory factor analysis; multilevel reliability; MPLUS