

【统计调查与分析】

# 网络测验中可靠性控制变量的完全因素不变性检验

王道阳<sup>1,2</sup>, 刘争光<sup>2</sup>

(1. 安徽师范大学 教育科学学院, 安徽 芜湖 241000; 2. 北京师范大学 中国基础教育质量监测协同创新中心, 北京 100875)

**摘要:**网络测验相对于传统测验最重要的问题是对被试答题时的情境与条件无法控制。通过对网络测验中被试是否有急事处理(无 vs. 有)、测试时的情绪体验(积极情绪 vs. 消极情绪)、噪音影响(无影响 vs. 有影响)以及测试场所(学校 vs. 其他地方)等情境变量进行完全因素不变性检验, 实现网络测验可靠性控制的目的。以主观健康量表(SHC)为例, 针对 3 667 名青少年(年龄 14~26 岁,  $M \pm SD = 19.31 \pm 1.62$ ), 进行上述情境变量不同组别的 SHC 完全因素不变性检验。结果表明: 是否有急事需要处理、答题时情绪体验、噪音影响、答题场所等完全因素不变性都在一定程度上不能成立, 说明上述指标可以作为网络测验中可靠性控制变量。

**关键词:**完全因素不变性; 测量不变性; 结构不变性; 主观健康量表

中图分类号: F845.11 文献标志码: A 文章编号: 1007-3116(2017)10-0110-06

## 一、问题的提出

网络测验相对于传统的纸笔测验有着成本小

(如不需要纸张印刷、数据录入等)、取样大、互动性强等优势, 但网络测验相对于传统的纸笔测验对答题时情境、条件控制要更加困难, 如果不加以控制可

收稿日期: 2017-03-31; 修复日期: 2017-06-23

基金项目: 国家社会科学基金教育学青年课题《不同类型流动青少年心理健康教育的衔接与贯通研究》(CBA120108)

作者简介: 王道阳, 男, 安徽宣城人, 博士生, 副教授, 硕士生导师, 研究方向: 心理发展与测评;

刘争光(通讯作者), 男, 河南郑州人, 博士生, 研究方向: 心理发展与测评。

analysis is further adopted to analyze the spatial cross-correlation between the " 5 Modernizations" coordination degree and the efficiency of the " 5 Modernizations" coordinated development. Finally, the dynamic evolution of the efficiency of the " 5 Modernizations" coordinated development is discussed based on the model of non-parameter kernel density estimate. The finding indicates: the coordination degree of the " 5 Modernizations" in Anhui province has been improving year by year in the sample period. The efficiency of the " 5 Modernizations" coordinated development is on the decline and the spatial difference is shrinking, but they fail to present a good interaction; the efficiency of the " 5 Modernizations" coordinated development is on a comparatively low level; coordination degree of the " 5 Modernizations" has an obvious spatial correlation and has a tendency of aggregation, with the Wanjiang Urban Belt as its aggregation area; the double variants of coordination degree of the " 5 Modernizations" and the efficiency of the " 5 Modernizations" coordinated development don't have a conspicuous aggregation effect. Based on this, put forward the corresponding recommendations.

**Key words:** the coordinative development of " 5 Modernizations"; efficiency; spatial correlation; dynamic evolution

(责任编辑: 郭诗梦)

能会带来测量不变性问题。已有研究报告网络测验中被试测验场地噪音影响、测试时的情绪体验、答题时长、答题时是否有急事处理以及答题场所等都是影响测验特性的重要影响因素<sup>[1]</sup>。基于此, 可以对于上述网络测验中的情境变量进行不变性检验。如果不变性不成立, 则说明某种条件下问卷或量表的测量特性发生变化, 这种条件下的数据需要进行处理或清理。

近年来, 西方的量表、问卷大量引入国内使用, 以及跨文化比较研究、追踪研究等, 都需解决测量等值性或测量不变性问题。Byrne 等将测量不变性与结构不变性 (Structure Invariance) 做了区分, 认为测量不变性包括因子负荷、截距以及误差方差不变性, 而结构不变性包括因子方差与协方差、潜均值等不变性<sup>[2]</sup>。测量不变性与结构不变性统称为完全因素不变性 (Factorial Invariance)。完全因素不变性验证主要基于多组验证性因素分析 (Multi-group Confirmatory Factor Analysis, MCFA) 理论<sup>[3]</sup>。MCFA 检验的参数模型主要有 (图 1): 第一, 无任何约束的分组验证性因素分析 ( $M_{\text{group}}$ )。即, 分别对  $M_{\text{group}1}$  和  $M_{\text{group}2}$  (或更多组) 进行没有任何约束的验证性因素分析, 考察各组验证性因素分析的拟合指数。第二, 协方差不变性 ( $M0$ , Invariant Covariance)。该模型是在零假设前提下, 协方差矩阵之间的不变性检验, 即验证  $\Sigma^g = \Sigma^{g'}$  ( $g$  与  $g'$  表示不同组, 下同)。第三, 形态不变性 ( $M1$ , Configural Invariance), 又称为基线模型 (Baseline Model)。该模型只要求各组潜变量、显变量基本结构对等, 不设定任何参数的限制<sup>[4]</sup>。第四, 负荷不变性 ( $M2$ , Metric Invariance), 又称为弱不变性 (Weak Invariance)。该模型主要是验证因子的负荷在两组之间是否对等, 即验证  $\Lambda_i^g = \Lambda_i^{g'}$ 。第五, 截距不变性 ( $M3$ , Scalar Invariance), 又称为强不变性 (Strong Invariance)。该模型是验证观测变量的截距是否具有跨组的不变性, 即验证  $\tau_i^g = \tau_i^{g'}$ 。第六, 严格不变性 ( $M4$ , Strict Invariance), 又称为误差方差不变性 (Error Variance Invariance) 或误差协方差不变性 (Error Covariance Invariance)。该模型是验证误差方差是否具有跨组的不变性, 即验证  $\Theta_i^g = \Theta_i^{g'}$ 。第七, 因子方差不变性 ( $M5$ , Factor Variance Invariance)。该模型是验证潜变量 (因子,  $\xi$ ) 的方差是否跨组的不变性, 即验证  $\Phi_i^g = \Phi_i^{g'}$ 。如果因子方差的不变性成立, 则说明各组潜变量上的离散程度是对等的<sup>[5]</sup>。第八, 因子协方差不变性

( $M5'$ , Factor Covariance Invariance)。该模型是验证潜变量的协方差是否跨组的不变性, 即验证  $\Phi_{ii'}^g = \Phi_{ii'}^{g'}$ 。在实际应用中, 一般把因子方差的不变性与因子协方差的不变性合并验证, 统称为因子方差-协方差不变性 (Factor Variance/Covariance Invariance)<sup>[6]</sup>。第九, 潜均值不变性 ( $M6$ , Latent Mean Invariance)。该模型是验证潜变量的均值是否具有跨组的不变性, 即验证  $\kappa_i^g = \kappa_i^{g'}$ 。

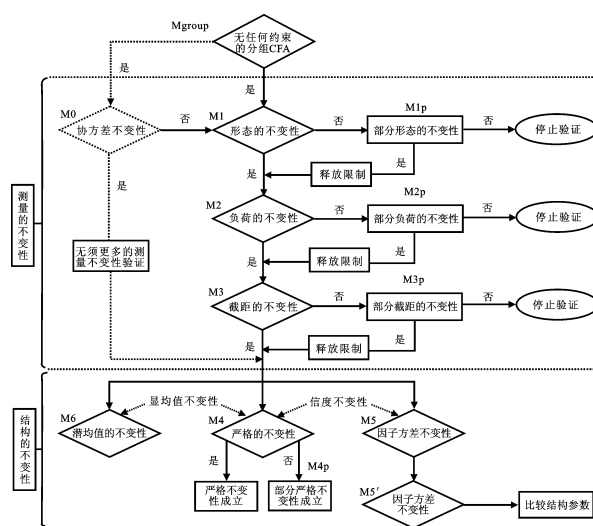


图 1 完全因素不变性检验流程图

主观健康 (Subjective health) 反映的是个体对于自身健康状况的理解和评估<sup>[7]</sup>。主观健康量表常常作为效标变量, 在人格、心理健康研究中广泛使用。但随着网络测验的普及, 也有研究者通过网络答题方式开展主观健康量表调查<sup>[8]</sup>。然而, 在网络测试中面临着测试环境、条件的无法控制, 如测试场所、答题时噪音影响、答题时被试的情绪状态、答题时被试是否着急 (有其他要紧的事等着要做) 等, 这些将直接影响主观健康测验的可靠性。因此, 通过对收集被试答题时情境、条件的状况, 并对这些变量进行不变性检验, 验证这些变量是否可以作为主观健康量表可靠性指标, 并以此作为清理数据的依据。

## 二、研究方法

### (一) 研究对象

运用计算机和手机网络平台, 征集答题调查了年龄为 14~26 岁 ( $M \pm SD = 19.31 \pm 1.62$ ) 的青少年 3 667 名。其中, 男 1 627 人 (44.4%), 女 2 040 人 (55.6%); 户口为非农业 1 035 人 (28.2%), 农业 2 632 人 (71.8%)。

### (二) 研究工具

研究工具主要包括背景问卷与主观健康量表。

背景问卷包括年龄、性别、户口等人口学变量,也包括被试对答题时周围环境与心理感受的评估。主要有“你在有事急需处理吗?”“你当前情绪体验是什么?”“你的周围是否有噪音?噪音是否影响答题?”“你在哪里答题?”等。

主观健康量表(Subjective health complaints, SHC)适用于个体对自身健康状况的主观评估,包括头痛、胃痛、腰痛、郁闷、易怒或脾气暴躁、感到紧张、感觉头晕。按过去半年发生上述症状频率分为5个等级(每天、每周超过1次、每周1次、每月1次、很少或从不)进行评估,总分在7~35之间,分数越高说明主观自评健康越满意(Haugland & Wold, 2001)。本次研究的 $\alpha$ 系数为0.84。

### (三)数据分析

采用 Harman 单因素检验对共同方法偏差进行检验。结果表明,特征值大于1的因子共3个,第一因子变异解释率为34.00%,小于40%的临界标准,共同方法偏差不显著。

针对主观健康量表,进行情境变量不同组别的

全因素不变性检验。验证模型包括  $M_{\text{group}}$ 、M1、M2、M3、M4、M5(因子方差-协方差的不变性)以及 M6。拟合指数使用  $\chi^2/df$ 、CFI、90% CI、RMSEA、SRMR;模型之间比较使用  $\Delta\chi^2$ ,即  $\chi^2$  变化量。也有研究者建议使用  $\Delta\text{CFI}$  指标<sup>[9]</sup>。较为宽松的观点认为  $|\Delta\text{CFI}| > 0.01$  时,两模型不变性检验假设不成立<sup>[10]</sup>,而严格的观点认为  $|\Delta\text{CFI}| > 0.002$  或者  $\Delta\chi^2 (p < 0.05)$  时,两模型不变性检验假设就不能成立<sup>[11]</sup>。本文同时使用  $\Delta\chi^2$  和  $\Delta\text{CFI}$  进行综合判断。

## 三、结果与分析

### (一)不同组别 SHC 的描述性统计

不同组别 SHC 得分、SHC 总分及各个题目总的均分和标准差,以及峰度和偏度值见表1。结果显示,偏度在0.47~1.90之间,峰度0.00~3.88之间,数据完全符合正态分布的标准(峰度绝对值在0~2、偏度绝对值在0~7)<sup>[12]</sup>。总样本的内部一致性 $\alpha$ 系数为0.84,不同组别的内部一致性 $\alpha$ 系数也在0.83~0.87之间。

表1 不同组别 SHC 的描述性统计表

组别		人数	%	M	SD	偏度	峰度	$\alpha$
有事急需处理	无(group1)	2 771	75.60	28.98	5.23	-1.04	0.97	0.84
	有(group2)	896	24.40	27.94	5.81	-1.12	1.35	0.85
答题时情绪体验	积极情绪(group1)	2 775	75.70	29.51	4.92	-1.20	1.59	0.83
	消极情绪(group2)	892	24.30	26.29	6.04	-0.69	0.37	0.84
噪音影响	无影响(group1)	3 398	92.70	28.90	5.28	-1.12	1.34	0.84
	有影响(group2)	269	7.30	26.52	6.25	-0.61	0.02	0.86
答题场所	学校(group1)	3 251	88.70	28.91	5.24	-1.08	1.16	0.84
	其他地方(group2)	416	11.30	27.33	6.30	-0.92	0.71	0.87
全样本	总分	3 667	100.00	28.73	5.39	-1.08	1.18	0.84
	题目1	—	—	4.45	0.97	-1.83	2.53	—
	题目2	—	—	4.54	0.87	-1.90	3.88	—
	题目3	—	—	4.39	0.96	-1.70	2.32	—
	题目4	—	—	3.63	1.21	-0.47	-0.88	—
	题目5	—	—	3.89	1.15	-0.78	-0.45	—
	题目6	—	—	3.73	1.18	-0.54	-0.82	—
	题目7	—	—	4.09	1.14	-1.05	0.00	—

### (二)不同组别 SHC 完全因素不变性验证结果分析

由表2可知,从不同组别的无任何约束验证性因素分析的结果来看,均表现出  $\chi^2 (p < 0.05)$ 、 $\text{CFI} > 0.85$ (仅噪音影响中有影响组、答题场所其他地方组  $\text{CFI} < 0.85$ ),虽然 RMSEA 的值大于0.10,但大多数 SRMR  $< 0.10$ ,而 RMSEA 更易于受到样本数量的影响。因此,总体上看,分组模型拟合较好。

具体来看 SHC 可靠性控制的几个变量不变性检验结果:第一,是否有急事需要处理(无 vs. 有),M1 与 M2、M3~M6 模型差异的  $\Delta\chi^2$  均表现为  $p < 0.05$ ,完全因素不变性不成立。第二,答题时情绪体验(积极情绪 vs. 消极情绪),M1~M6 模型差异的  $\Delta\chi^2$  均表现为  $p < 0.05$ ,且 M3 与 M4、M5 与 M6 模型的  $\Delta\text{CFI} > 0.01$ ,完全因素不变性不成立。第三,噪音影响(无影响 vs. 有影响),M1、M2 模型差异的  $\Delta\chi^2$

均表现为  $p > 0.05$ , M1、M2 不变性成立; 而 M3 ~ M6 不变性检验不成立 ( $\Delta\chi^2_{(7)} = 63.25, \Delta\chi^2_{(1)} = 13.26, 51.08$ , 均  $p < 0.05$ )。第四, 答题场所(学校 vs. 其他地方), M1 ~ M6 模型差异的  $\Delta\chi^2$  均表现为  $p < 0.05$ , 且 M2 ~ M6 的  $\Delta CFI > 0.002$ , 完全因素不变性不能成立。

表 2 不同组别 SHC 完全因素不变性模型的拟合指数表

模型	$\chi^2$	df	$\Delta\chi^2$	$\Delta df$	CFI	$\Delta CFI$	90%CI	RMSEA	SRMR
<b>有事急需处理</b>									
M <sub>group1</sub>	915.02	14	—	—	0.868	—	0.144~0.161	0.152	0.061
M <sub>group2</sub>	346.91	14	—	—	0.864	—	0.148~0.178	0.163	0.066
M1	1262.52	30	—	—	0.867	—	0.143~0.157	0.150	0.062
M2	1272.59	34	10.07*	4	0.867	0.000	0.134~0.148	0.141	0.065
M3	1280.50	40	8.09	6	0.867	0.000	0.124~0.136	0.130	0.066
M4	1337.04	47	56.54***	7	0.861	-0.005	0.117~0.128	0.122	0.084
M5	1349.48	48	12.44***	1	0.860	-0.001	0.116~0.127	0.122	0.105
M6	1375.14	49	25.66***	1	0.857	-0.003	0.116~0.127	0.121	0.123
<b>答题时情绪体验</b>									
M <sub>group1</sub>	886.15	14	—	—	0.865	—	0.142~0.158	0.150	0.063
M <sub>group2</sub>	326.06	14	—	—	0.860	—	0.143~0.173	0.158	0.065
M1	1225.36	30	—	—	0.862	—	0.140~0.155	0.147	0.064
M2	1249.64	34	24.28***	4	0.860	-0.002	0.133~0.146	0.140	0.069
M3	1262.45	40	12.81*	6	0.859	-0.001	0.123~0.135	0.129	0.069
M4	1630.37	47	367.92***	7	0.818	-0.041	0.130~0.141	0.136	0.146
M5	1680.94	48	50.57***	1	0.812	-0.006	0.131~0.142	0.136	0.197
M6	1938.45	49	257.51***	1	0.783	-0.029	0.140~0.151	0.145	0.270
<b>噪音影响</b>									
M <sub>group1</sub>	1083.29	14	—	—	0.873	—	0.142~0.158	0.150	0.061
M <sub>group2</sub>	159.02	14	—	—	0.821	—	0.169~0.224	0.196	0.081
M1	1242.44	30	—	—	0.868	—	0.141~0.156	0.148	0.062
M2	1246.51	34	4.07	4	0.868	0.000	0.133~0.146	0.139	0.063
M3	1251.38	40	4.87	6	0.869	0.001	0.122~0.135	0.129	0.064
M4	1314.63	47	63.25***	7	0.862	-0.007	0.116~0.127	0.121	0.076
M5	1327.89	48	13.26***	1	0.861	-0.001	0.115~0.126	0.121	0.096
M6	1378.97	49	51.08***	1	0.856	-0.005	0.116~0.127	0.122	0.122
<b>答题场所</b>									
M <sub>group1</sub>	1004.71	14	—	—	0.875	—	0.140~0.155	0.148	0.060
M <sub>group2</sub>	237.48	14	—	—	0.832	—	0.174~0.218	0.196	0.076
M1	1242.69	30	—	—	0.869	—	0.141~0.156	0.148	0.062
M2	1259.02	34	16.33**	4	0.868	-0.001	0.134~0.147	0.140	0.065
M3	1281.34	40	22.32**	6	0.866	-0.002	0.124~0.136	0.130	0.067
M4	1352.61	47	71.27***	7	0.859	-0.007	0.117~0.129	0.123	0.094
M5	1374.42	48	21.81***	1	0.857	-0.002	0.117~0.128	0.123	0.122
M6	1402.13	49	27.71***	1	0.854	-0.003	0.117~0.128	0.123	0.141

注: M5 是因子方差-协方差不变性; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ 。

## 四、讨论

### (一) SHC 的完全因素不变性检验结果

主观健康是一个简单而重要的评价指标, 主观健康较差可能是负向心理社会状况(比如离群、负性

生活事件、沮丧、工作压力)的一个共同特征。反之, 当被试处于不同的测试情境时可能会使得不同组被试之间对自身健康理解产生差异, 这需要通过完全因素不变性检验才能确定。完全因素不变检验结果表明: 第一, 是否有急事需要处理(无 vs. 有)的完全

因素不变性检验,除了 M2 成立,其余不变性均不成立。但是, M1~M3 模型差异的  $\Delta CFI$  值过小(均 $< 0.002$ )。所以初步来看,有急事与无急事两组主观健康的潜变量具有等同的因子负荷,观测变量的截距也基本是对等的;但是 M3 与 M4、M5 与 M6 差异显著,不变性不成立,即有急事与无急事两组主观健康的误差方差以及潜均值是不对等的。综合来看,这两组的主观健康是不对等,是不能进行比较的。第二,答题时情绪体验(积极情绪 vs. 消极情绪)的完全因素不变性不成立。这说明,答题时积极情绪与消极情绪两组主观健康在因子载荷、截距、误差方差等观测变量不具有对等性,当然在潜变量方差协方差、潜均值更不能对等。因此,答题时被试情绪状态使被试对主观健康理解有着显著影响,应该把该变量作为主观健康量表可靠性控制的重要变量。第三,噪音影响(无影响 vs. 有影响)的完全因素不变性只有部分成立。其中,答题时情绪体验的测量不变性基本成立,说明噪音无影响与有影响两组主观健康的因子载荷、截距具有对等性。但结构不变性不能成立,说明两组主观健康潜变量上的离散程度、协方差不是对等的。第四,答题场所(学校 vs. 其他地方)完全因素不变性不成立。这说明,学校答题与其他地方答题两组主观健康在观测变量与潜变量的因子载荷、截距、误差方差、均值等都是不对等的。也就是说,答题场所使得被试对主观健康的理解有着显著影响,也应该把该变量作为主观健康量表可靠性控制的重要变量。

## (二)SHC 的完全因素不变性检验流程

在问题提出中,已经列出完全因素不变性检验的各种模型,但在实际应用中并不是所有的模型都需要检验,而是根据研究需要选择合适的部分模型进行检验。实际上,关于选择哪些模型来检验,研究者存在着一些争议,但也存在基本共识。大致有几种思路进行完全因素不变性检验(见图 1):第一,  $M_{\text{group}} \rightarrow M0 \rightarrow M4 \rightarrow \dots$ 。如果协方差矩阵(M0)不变成立,则不需要再进行接下来的形态不变性、负荷

不变性以及截距的不变性验证,可以直接进行误差方差不变性验证。如果总体协方差不变是不成立,则需要对形态不变性、负荷不变性以及截距的不变性逐步验证。从理论上,应该从协方差矩阵不变性(M0)检验作为多组验证性因素分析的逻辑起点,但在实际使用中 M0 与 M1、M2、M3 都不存在嵌套关系,相互比较变得困难。因而,从 M1 为起点开展的完全因素不变性检验更常使用。第二,  $M_{\text{group}} \rightarrow M1 \rightarrow M2 \rightarrow M3 \rightarrow M4 \rightarrow \dots$ ,就是常用的以形态不变性为逻辑起点的完全因素不变性检验。严格意义上来说,形态不变性不能算是测量不变性检验的一部分<sup>[13]179-181</sup>。但是,该模型作为后续不变性检验的基线模型,为判断不变性提供了参照标准模型,所以在实际应用中非常重要。这种思路的好处是模型有着嵌套关系,易于进行模型比较,也就易于验证相关假设;弊端是步骤较繁琐,条件较为严格。第三,仅仅进行测量不变性检验,不进行结构不变性检验,即  $M_{\text{group}} \rightarrow M1 \rightarrow M2 \rightarrow M3 \rightarrow M4$ 。综上,为了便于模型之间的比较,同时对 SHC 有一个较为系统和完整的完全因素不变性验证,故本文采用  $M_{\text{group}} \rightarrow M1 \rightarrow M2 \rightarrow M3 \rightarrow M4 \rightarrow \dots$  思路,进行完全因素不变性检验。

## (三)研究的意义和局限性

本文的意义在于探索网络测验的情境下如何控制测试可靠性,结果表明,测试时被试是否匆忙、测试时噪音影响、测试时被试情绪、测试场地等是测试可靠性的关键变量。实际上,完全因素不变性应该是多组比较、量表或问卷翻译中必须进行的验证过程,希望以此推动更多研究者重视该问题。本文采用完全因素不变性检验的方法与步骤较为完备,但也显得繁琐,后续研究可以探索更加简洁的方法与步骤。另外,对于如何实现网络测验的可靠性控制,还有一些如被试互联网计算机操作能力、测试完成时间、测试网络途径(计算机网络、移动网络)等因素也需考虑。

## 参考文献:

- [1] Hardre P L, Crowson H M, Xie K. Examining Contexts-of-Use for Web-Based and Paper-Based Questionnaires[J]. Educational & Psychological Measurement, 2012, 72(6).
- [2] Byrne B M, Shavelson R J, Muthén B. Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance[J]. Psychological Bulletin, 1989, 105(3).
- [3] Adam W M, Gary J L. A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance[J]. Structural Equation Modeling: A Multidisciplinary Journal, 2004, 11(1).

- [4] Horn J L, Mcardle J J. A Practical and Theoretical Guide to Measurement Invariance in Aging Research[J]. *Experimental Aging Research*, 1992, 18(3-4).
- [5] Schaubroeck J, Green S G. Confirmatory Factor Analytic Procedures for Assessing Change during Organizational Entry [J]. *Journal of Applied Psychology*, 1989, 74(6).
- [6] Vandenberg R J, Lance C E. A Review and Synthesis of the Management Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research[J]. *Organizational Research Methods*, 2000, 3(1).
- [7] Currie C, Alemándíaz A Y. Building Knowledge on Adolescent Health: Reflections on the Contribution of the Health Behaviour in School-aged Children (HBSC) Study[J]. *European Journal of Public Health*, 2015, 25(2).
- [8] 孙其昂, 李向健. 中国城乡居民自感健康与社会分层——基于(CGSS)2008年的一项实证研究[J]. *统计与信息论坛*, 2013(12).
- [9] Meade A W, Johnson E C, Braddy P W. Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance[J]. *Journal of Applied Psychology*, 2008, 93(3).
- [10] Cheung G W, Rensvold R B. Evaluating Goodness-of-fit Indexes for Testing Measurement Invariance[J]. *Structural Equation Modeling*, 2002, 9(2).
- [11] Fournier J, Gaudreau P, Demontrond-Behr P, et al. French Translation of the Flow State Scale-2: Factor Structure, Cross-cultural Invariance, and Associations with Goal Attainment[J]. *Psychology of Sport and Exercise*, 2007, 8(6).
- [12] Curran P J, West S G, Finch J F. The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis[J]. *Psychological Methods*, 1996, 1(1).
- [13] 王孟成. 潜变量建模与 Mplus 应用[M]. 重庆: 重庆大学出版社, 2014.

### Testing Factorial Invariance in Online Test Based on Reliability Controlling Variables

WANG Dao-yang<sup>1,2</sup>, LIU Zheng-guang<sup>2</sup>

(1. College of Educational Science, Anhui Normal University, Wuhu 241000, China; 2. Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing 100875, China)

**Abstract:** With the prevalence of internet, the online test shows great advantage for psychological assessment, and would be likely to replace the paper and pencil test. During the process, the issue whether the online test has invariance with the paper and pencil test should be considered and solved. Compared with the traditional test, the online test is facing the challenge from the uncontrollable environment and condition in test. In order to control the reliability of the online test, the researchers need to test Factorial invariance as to the variables, such as, hurry or not, noise or not, mood in the test, testing place and so on. The study took the subjective health complaints (SHC) as an example, with the subjects of 3 667 adolescents and youths (age: 14-26,  $M \pm SD = 19.31 \pm 1.62$ ), to test SHC factorial invariance according to different groups as below: hurry (hurry vs. no-hurry), noise (noise vs. non-noise), mood in the test (positive mood vs. negative mood), testing place (school vs. other place). The results show that factorial invariance is untenable across the hurry group and no-hurry group or positive mood group and negative mood group or noise group and non-noise group or School group and other-place group. Therefore, different groups divided by variables above have inequivalent responses for subjective health, and these reliability controlling variables should be controlled as key interfering variables to keep a high reliability for online test.

**Key words:** factorial invariance; measurement equivalence; structural invariance; subjective health complaints

(责任编辑:于 茜)