手机与纸笔测验的心理测量学等值性：以儿童版事件冲击量表为例

黄飞[1]，祝卓宏[1]，王文忠[1]，张建新[1]，纪阳[2]，章魁[2]，刘宁*，汪浩*
(1.中国科学院心理研究所，北京 100101；2.北京邮电大学，北京 100876)

【摘要】 目的：以儿童版事件冲击量表为例考察手机测验和纸笔测验的心理测量学等值性。方法：在四川绵竹市某中学对 334 名初中学生施测手机版本和纸笔版本的儿童事件冲击量表(CRIES-13)，并考虑平衡顺序。应用经典测验理论和验证性因素分析取向对跨测验模式的等值性进行检验。结果：两种框架下的分析均得出量表具有较好的信度和结构效度，除个别条目外，在条目水平和量表水平上均存在很高的跨测验模式的等值性。结论：结果初步支持手机平台用于心理测验的测量学的可靠性以及与纸笔测验的等值性。

【关键词】 手机平台测验；纸笔测验；等值性；重复测量

中图分类号：R395.1 文献标识码：A 文章编号：1005-3611(2010)01-0031-03

Psychometric Equivalence Between Mobile Phone Based and Paper-and-pencil Tests: A Case with Children's Revised Impact of Event Scale

HUANG Fei, ZHU Zhuo-hong, WANG Wen-zhong, ZHANG Jian-xin, et al
Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

【Abstract】 **Objective:** Mobile phone can be designed as a platform to administer test, Mobile-Phone Based Test (MPBT). To examine the reliability and validity of MPBT and equivalence with paper-and-pencil test (PPT) were the purposes of this study. **Methods:** Children's Revised Impact of Event Scale (CRIES) was edited into two forms: PPT and MPBT. 334 students of 8 eight classes in grade 7 and 8 in Mianzhu city Sichuan province participated this study. Both Classics Test Theory (CTT) and Confirmatory Factor Analysis (CFA) were applied to examine the reliability, construct validity and equivalence between the two forms. **Results:** In CTT, item means, standard deviations, the Cronbach's α and inter-correlations among sub-scales according to the original three sub-scale construct were compared, and no significant difference was found. And both forms had acceptable internal consistency coefficients. In CFA, model comparison revealed strict measurement and structure equivalence between the two forms. **Conclusion:** Both CTT and CFA analysis supported the reliability and validity of MPBT and equivalence between PPT and MPBT in the case of CRIES.

【Key words】 Mobile phone based test; Paper-and-pencil test; Equivalence; Repeated measure

任何一种新生事物的出现，都可能成为一个研究对象，也可能成为研究和实践的工具，手机就像计算机一样，同时具备这两点。手机作为一种数据收集工具，具有很大的潜力。比如，手机平台与经验抽样方法(ESM)的结合使得生态即时评估设计可行且便利[1,2]。手机和网络的结合，使得手机作为测试平台与计算机具有很大的相似性。可以预见，将手机作为研究工具的尝试将会由于其强大、独特的优势而具有很大的发展空间。况且，手机已经渗透到人们社会生活诸多领域和层面，影响着人们的生活节奏[3]，而且还将继续扩大影响，包括心理研究和实践[4,5]。

所谓手机测验，是指以手机为平台实施测验的过程。相对于纸笔测验和网络测验，手机测验具有的优势在于：普及性，便捷性，个人身份可辨识性，以及实时连续性。和网络测验一样，设计良好的在线手机平台测验能提供即时的反馈，能数字化存储数据，能相对便利的修改和编辑测量内容[6]。但同样，它也面临网络测验相似的威胁：手机测验平台开发的技术性需求，应答者动机的不明确性，测验的安全，测验的伦理等问题[6]。手机的小巧也是手机平台测验的一个局限所在：由于界面和按键的有限，对于测验内容的编排的灵活性不如用计算机的网络测验和纸笔测验。

手机测验是一个新兴的事物，虽然有研究者用手机来收集数据，但是对于将手机作为一种测验平台的可行性和有效性则没有论证过。而在使用手机测验过程中，必须保证测量的内容不会受到这种测验方法的不利影响。否则，手机平台的测验可能会遭受质疑。除了操作平台的硬件和软件以及使用友好

【基金项目】 中国科学院抗震救灾应急研究项目 (O8CX112011)；国家高技术研究发展计划重点项目(2008AA021204)；国家自然科学基金(30840001)

* 诺基亚研究中心

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

性之外，心理测量学上关注的很重要的一点就是测验的信效度问题，其中包括与传统的纸笔施测的测量等值性问题[7,8]。本研究将就此问题在已有的平台操作软件的基础上，集中探讨跨测验模式的心理测量学等值性。

1　对象与方法

1.1　研究设计与被试

采用被试内设计，并结合 ABBA 策略来平衡顺序效应。从四川省绵竹市汉旺学校的初一初二年级各抽取 4 个班级，其中每个年级中两个班级先进行纸笔条件下的测验，再进行手机平台的测试（A 顺序），另外两个班级则以相反的顺序进行（B 顺序）。手机测验平台线上版是北京邮电大学移动生活与新媒体实验室制作的，在 Nokia6120 手机界面上呈现。操作基本流程：在手机端安装一个程序，在线注册，然后登陆之后选择测验内容进入测验，完成后提交，最后退出。手机界面每个版面容纳一个条目，条目选项是以汉字表示不同程度，通过上下键选择相应的选项。测试程序为：对先参与手机平台测验的被试在班主任老师的指导下熟悉手机平台测试的基本操作，然后进入正式测验，整个过程 5-10 分钟完成；完成之后回收手机，进行正常教学活动，在课程结尾发放纸笔形式的测验，指导学生完成并当场回收，不到 5 分钟。而对于先参与纸笔测验的被试则以上述相反的顺序进行。

最后纸笔形式的测验和手机平台测验能匹配上的被试数为 334 名。其中男 178 人，女 156 人；年龄 12-15 岁，初一 161 人，初二 173 人。A 顺序组 169 人，B 顺序组 165 人。

1.2　测量工具

事件冲击量表（儿童版）中文修订版，该问卷的条目内容修改为与地震相关。该问卷共有 13 个条目，原始结构是包含 3 个维度：侵入性（Intrusion），回避（Avoidance），唤醒（Arousal）。采用四点量表，“0=完全没有”，“1=很少”，“3=有时”，“5=经常”。

2　结　　果

2.1　条目水平上的检验

从逐个条目水平上的均值比较结果来看，手机测验和纸笔测验在不同的条目上表现出不同的大小趋势，但是差异的效应量均较小，若将显著性水平采用 Bonferroni 法加以校正，所有条目得分的差异（包括 I1）均不具有统计显著性。见表 1。

2.2　量表水平上的检验

从量表水平上的均值比较结果来看，两种测验模式在三个分量表和总量表得分差异很小，不具有统计学显著性。从相关来看，两种测验模式下得分具有高度相关。见表 2。

从各个分量表之间相关模式和内部一致性的跨模式比较来看，从数值趋势上来看，手机和纸笔测验在不同量表上的各有高低。采用 Feldt 检验法并因相关样本校正自由度，对内部一致性的差异检验[9]表明手机测验侵入的内部一致性低于纸笔测验，而唤醒量表的内部一致性则高于纸笔。整个量表的内部一致性分别为 0.86，0.86，差异不显著。见表 3。

对两种测验下 3 个分量表形成的 6 个变量的相关模式进行检验[10]，逐一检验表明侵入与回避之间的相关在纸笔条件下更高，唤醒量表的跨测验方式的相关低于侵入和回避。

表 1　条目水平上比较

项目	手机测验	纸笔测验	t 值
I1	1.64 ± 1.37	1.77 ± 1.37	-2.00*
I2	1.96 ± 1.61	2.02 ± 1.73	-0.68
I3	1.34 ± 1.40	1.36 ± 1.41	-0.29
I4	1.49 ± 1.46	1.52 ± 1.44	-0.41
I5	1.55 ± 1.47	1.69 ± 1.58	-1.75
I6	1.78 ± 1.60	1.75 ± 1.59	0.34
I7	1.96 ± 1.65	1.85 ± 1.56	1.37
I8	1.92 ± 1.53	1.86 ± 1.47	0.81
I9	1.55 ± 1.37	1.54 ± 1.41	0.08
I10	2.01 ± 1.70	1.87 ± 1.64	1.70
I11	1.69 ± 1.60	1.70 ± 1.55	-0.11
I12	1.93 ± 1.59	1.89 ± 1.58	0.42
I13	0.68 ± 1.13	0.69 ± 1.14	-0.14

注：*P<0.05

表 2　量表水平上的检验

量表	手机测验	纸笔测验	t 值	相关
侵入	6.60 ± 4.48	6.69 ± 4.69	-0.57	0.79***
回避	7.71 ± 5.14	7.49 ± 5.20	1.17	0.79***
唤醒	7.19 ± 5.11	7.33 ± 4.63	-0.73	0.72***
总分	21.49 ± 12.01	21.52 ± 12.07	-0.07	0.85***

注：***P<0.001

表 3　亚量表间的相关及内部一致性的比较

量表	手机测验			纸笔测验			手机-纸笔比较		
	侵入	回避	唤醒	侵入	回避	唤醒	侵入	回避	唤醒
侵入	0.79			0.84			-0.05**		
回避	0.43	0.79		0.51	0.81		-0.08*	-0.02	
唤醒	0.69	0.39	0.75	0.68	0.43	0.63	0.01	-0.04	0.12***

注：每种测验模式下的量表的内部一致性呈现在主对角线上；*P<0.05，**P<0.001。所有相关 ***P<0.001。

2.3　验证性因素分析对等值性的检验

采用 LISREL8.75 对重复测量的量表进行等值性检验[11]，检验的步骤如下。

基线模型 M_0 的设定：相同测验模式下的各个结构相关自由估计，不同模式相同结构相关自由估计，所有指标误差自由估计。数据重新度量的方法为：将

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

两种模式下的各个维度的第一个指标的参数设定为1,其它指标的负载自由估计。基线模型 M_{01} 的设定:在上述模型的基础上,相同指标在两种模式下的对应指标的误差协方差自由估计模型,即相同指标的重复测量之间除了共同的潜在特质的影响之外,还存在重复测量效应。负载等同模型 M_1 的设定:两种模式下的相同指标的负载设定为等同,若发现不能接受等同设定,则分量表单独检验是否等值,这过程中鉴别出差异不大的条目,然后这些条目设定为等值,其它条目的负载自由估计。截距等同模型 M_2,将不同模式下的相同指标的截距设置为等同。参考CTT 指标均值比较的结果,依次检查各个指标的截距是否等同,将发现有显著差异的指标自由估计。指标的误差方差等同模型 M_3。因子方差等同模型 M_4,因子协方差等同模型 M_5。因子均值模型 M_6:纸笔模式下的均值设定为 0,网络条件下的均值自由估计。

各模型的拟合情况见表 4。

再参考 Ferrando 和 Lorenzo-Seva 所推荐的程序[12],在模型 5 的基础上将三个分量表的斜方差同时限定为 1,这个模型的拟合情况和模型 5 相当。

CFA 下等值性检验小结:13 个条目除去最后一个条目,其他条目的测量与结构在两种测验模式下具有完全的(强)等值性:条目与所欲测量的潜在变量之间的关系、测量的误差、潜变量的结构、指标的截距、潜在因子得分均满足等值性。I13 与唤醒的关系在两种测验模式下不均等,相对来说在手机测验条件下的关系更密切,另外 I4 的误差方差不均等,在手机测验下的误差方差更大,这可以解释上面发现的内部一致性的差异。

表 4 测量参数和结构参数等值过程各模型的拟合情况

模 型	df	χ^2	RMSEA	NNFI	CFI	ΔCFI	$\Delta\chi^2(\Delta df)$	ECVI
M_0	290	1123.77	0.093	0.931	0.939	—	—	3.90
M_{01}	277	753.65	0.072	0.952	0.959	0.02	370.12(13)***a	2.86
M_1	287	827.81	0.075	0.952	0.958	-0.001	74.16(10)***b	3.03
M_{11}	286	769.23	0.071	0.953	0.959	0.000	15.58(9)c	2.86
M_2	299	781.13	0.070	0.955	0.959	0.000	11.90(13)d	2.81
M_3	312	806.44	0.069	0.955	0.958	-0.001	25.31(13)*e	2.81
M_4	315	808.11	0.069	0.956	0.958	0.000	1.67(3)f	2.80
M_5	318	812.17	0.068	0.957	0.958	0.000	4.06(3)g	2.79
M_6	315	810.34	0.069	0.956	0.958	0.000	1.83(3)h	2.81

注:对用于单位化的指标的等值性进行估计:将其它的指标设定,检验之前设定四个条目的各个参数的等值性,结果发现所有的条目的各个参数都具有强的等值性。ΔCFI, $\Delta\chi^2(\Delta df)$的来源:a 是 M_{01} 相对于 M_0 的变化,b 是 M_1 与 M_{01} 的差异,c 是 M_{11} 与 M_{01} 的差异,d 是 M_2 与 M_{11} 的差异,e 是 M_3 与 M_2 的差异,f 是 M_4 与 M_3 的差异,g 是 M_5 与 M_4 的差异,h 是 M_5 与 M_6 的差异。

3 结 论

测验模式的革新,会带来一些扰动,手机测验和计算机化辅助测验、网络测验一样,也面临同样的情境。在这过程中,硬件技术、软件技术、各类测验的发展经验,检验的技术也得到发展和积累[6]。随着手机更深入全面的融入人们的日常生活,以及技术的革新,手机的功能和影响日益凸显。手机与心理服务的结合将会是一个发展趋势,手机测评将会成为其中一员,但需要合理的评估与监督才能发挥最佳的效用。

本文以儿童版事件冲击量表为例,从心理测量学角度对手机测验和纸笔测验的等值性进行探讨。本研究使用的是被试内设计,及相应的重复测量的等值性分析方法。设计和分析的基本原理是:较短时间间隔下,个体属性的变化不显著,这样检验出来的结果就更能归为测验模式的效应。本文从不同的分析框架出发对等值性进行检验,从而可以提供多方位的信息,为工具的开放和发展提供重要的信息。本研究表明手机测验方式对于事件冲击程度的自我评估基本上没有影响,但其中最后一个条目和第 4 个条目,在两种条件下似乎功能有所差异,这可能是手机相对于纸笔更“唐突”,对于初次使用者来说可能引发兴趣,提高兴奋程度从而影响到作答。

本研究只是建立在一个样本上使用一个量表进行的检验,进一步的研究还有必要检验测验模式和其他可能影响等值性因素的交互作用,比如性别,年龄、测量工具的类型等。心理测量学的等值性只是手机平台测验和纸笔测验关系中的一个方面。关于手机测验的技术问题、成本问题等都会影响手机测验的推广。这些因素将影响手机测试的可靠性和有效性。手机测验只有很好的满足要求和克服障碍,才能为实践应用提供一个可靠的工具。

参 考 文 献

1 Song ZL, Foo M, Uy MA. Mood spillover and crossover among dual-earner couples: A cell phone event sampling study. Journal of Applied Psychology, 2008, 93: 443-452

(下转第 72 页)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

即拥有 D 型人格的幸存者更有可能遭受创伤性应激障碍(PTSD)[18]。

参 考 文 献

1 温盛霖,王相兰,陶炯,等. 四川江油太平镇安置点北川、平武地震灾民 1 周后心理症状分析. 中国神经精神疾病杂志,2008,34(9):525-527

2 李洁,郭建雄,徐文军,等. 四川汶川地震伤员的心理卫生状况调查. 中国神经精神疾病杂志,2008,34(9):523-524

3 罗兴伟,高雪屏,蔡太生,等. 汶川地震亲历者的心理健康状况调查. 中国临床心理学杂志,2008,16(6):571-573

4 Bhushan B, Kumar JS. Emotional distress and posttraumatic stress in children surviving the 2004 tsunami. Journal of Loss and Trauma, 2007, 12(3): 245-257

5 Giannopoulou I, Smith P, Ecker C, Strouthos M, et al. Factor structure of the children's revised impact of event scale (CRIES) with children exposed to earthquake. Personality and Individual Differences, 2006, 40(5): 1027-1037

6 Yule W. Alleviating the effects of war and displacement on children. Traumatology, 2002, 8(3): 1-71

7 Ashraf H. Tsunami wreaks mental health havoc. Bulletin of the World Health Organization, 2005, 83(6): 405-406

8 Norris FH, Elrod CL. Psychosocial consequences of disaster: A review of past research, in methods for disaster mental health research. Edited by Norris FH, Galea S, Friedman MJ, Watson PJ. New York: Guilford, 2006. 20-42

9 Hsien WK, Wu SJ, Ma TC, Chiu MC, Chou SY. Posttraumatic symptoms were worst among quake victims with injuries following the Chi-chi quake in Taiwan. Journal of Psychosomatic Research, 2007, 62: 495- 500

10 张本,张凤阁,王丽萍,等. 30 年后唐山地震所致孤儿创伤性应激障碍现患率调查. 中国心理卫生杂志,2008,6:469-473

11 黄国平,张亚林,向慧,等. 事件影响量表-修订版(IES-R)在女性犯人中的信度、效度分析. 中国心理卫生杂志,2006,20(5):28-31

12 Ogińska-Bulik N, Juczyński Z. Distressed personality (Type D) and risk of occurrence of cardiovascular diseases. In: Kosińska-Dec K, Szewczyk L. Development. Health. Disease. Warsaw: BEL Studio, 2004. 5-17

13 白俊云,赵兴蓉,许秀峰. D 型人格量表的信效度检验. 中国心理卫生杂志,2007,21(5):329-332

14 Creamer M, Bell R, Failla S. Psychometric properties of the impact of event scale-revised. Behaviour Research and Therapy, 2003, 41: 1489-1496

15 Hsu CC, Chong MY, Yang P, Yen CF. Posttraumatic stress disorder among adolescent earthquake victims in Taiwan. Journal of the American academy of child and adolescent. Psychiatry, 2002, 41(7): 875-881

16 Green LB, Grace CM, Gleser CG. Identifying survivors at risk: Long-term impairment following the Beverly Hills Supper Club Fire. Journal of Consulting and Clinical Psychology, 1985, 53(5): 672-678

17 沃建中,向燕辉,沈莉,等. 5.12 重震区成人受灾者不良情绪反应程度与人格、应对方式等因素关系研究. 中国临床心理学杂志,2009,17(2):227-229

18 Ogińska-Bulik N. Occupational stress and its consequences in healthcare professionals: The role of Type D Personality. International Journal of Occupational Medicine and Environment Health, 2006, 19(2): 113-122

(收稿日期:2009-07-09)

(上接第 33 页)

2 Collins RL, Kashdan TB, Gollnisch G. The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption. Experimental and Clinical Psychopharmacology, 2003, 11: 73-78

3 Wajcman J. Life in the fast lane? Towards a sociology of technology and time. The British Journal of Sociology, 2008, 59: 59-77

4 Boschen MJ, Casey LM. The use of mobile telephones as adjuncts to cognitive behavioral psychotherapy. Professional Psychology: Research and Practice, 2008, 39: 546-552

5 Grassi A, Preziosa A, Villani D, et al. A relaxing journey: The use of mobile phones for well-being improvement. Annual Review of CyberTherapy and Telemedicine, 2007, 5: 123-131

6 Naglieri JA, Drasgow F, Schmit M, et al. Psychological test on the internet: New problem, old issues. American Psychologist, 2004, 59: 150-162

7 李育辉. 不同类型量表的线上调查和传统纸笔测验的比较研究. 博士学位论文. 中国科学院心理研究所,2007

8 蔡华俭,林永佳,伍秋萍,等. 网络测验和纸笔测验的测量不变性研究:以生活满意度为例. 心理学报,2008,40:228-239

9 Alsawalmeh YM, Feldt LS. A modification of Feldt's test of the equity of two dependent alpha coefficients. Psychometrika, 1994, 59: 49-57

10 Steiger JH. Tests for comparing elements of a correlation matrix. Psychological Bulletin, 1980, 87: 245-251

11 Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 2000, 3: 4-70

12 Ferrando PJ, Lorenzo-Seva U. IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and internet-administered questionnaires. Psychological Methods, 2005, 10: 193-205

(收稿日期:2009-07-09)

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net