

艾森克个性问卷(成人版)计算机自适应测验的信效度研究

雷辉^{1,2}, 戴晓阳², 朱熊兆¹

(1.中南大学湘雅二医院医学心理研究所,湖南 长沙 410011;

2.深圳大学师范学院心理学系,广东 深圳 518060)

【摘要】 目的:检验艾森克个性问卷(成人版)计算机自适应测验(EPQ-A-CAT)在实际应用中的测量精准度、效度和测试效率。方法:选取 204 名大学生被试完成了 EPQ-A-CAT 的测试,其中 120 名被试接受了 CAT 和纸笔两种版式的施测,并对 31 名被试进行了 CAT 版式的重测。结果:CAT 版式的估计标准误均值在四个分量表中都小于终止规则预设的最小误差值;CAT 重测的相关系数分别为 0.79(E)、0.85(N)、0.67(P)、0.71(L);CAT 版式总体能够节省一半以上的项目,各分量表节省了 70%~30% 的项目;两个版本测试结果的积差相关系数在 E、N、P、L 四个分量表中分别为 0.85、0.81、0.75 和 0.74。结论:研究结果证明了 EPQ-A-CAT 版式是高效、可靠和有效的测量工具。

【关键词】 艾森克个性问卷;计算机自适应测验;实际应用

中图分类号: R395.1

文献标识码: A

文章编号: 1005-3611(2012)03-0325-03

Reliability and Validity of the Computerized Adaptive Test of Eysenck Personality Questionnaire (Adult) (EPQ-A-CAT)

LEI Hui, DAI Xiao-yang, ZHU Xiong-zhao

Department of Psychology, Shenzhen University, Shenzhen 518060, China

【Abstract】 Objective: A Computerized Adaptive Test of Eysenck Personality Questionnaire(Adult) (EPQ-A-CAT) was administrated under real testing condition to evaluate and validate measurement accuracy, validity, and efficiency. **Methods:** The EPQ-A-CAT was administered in a sample of 204 college students, of which 120 subjects received the two versions CAT and paper-and-pencil test and 30 subjects received the CAT test-retest. **Results:** The mean of estimated standard error of CAT under four subscales were all less than the prespecified minimum error values; and the correlations of test-retest were 0.79 (E), 0.85 (N), 0.67 (P), 0.71 (L) respectively; item savings were above a half of the total scale, and ranged from 70% to 30% for the four subscales; the Pearson correlation coefficients of the two versions scores were 0.85, 0.81, 0.75, and 0.74 for the E, N, P, and L subscales. **Conclusion:** The EPQ-A-CAT proved to be an efficient, reliable and valid measurement tool.

【Key words】 Computerized adaptive testing; Eysenck Personality Questionnaire; Practical application

基于项目反应理论的计算机自适应测验(computerized adaptive test, CAT)是一种比传统测验方式更加快速有效的测验方式^[1]。自从上世纪 80 年代, CAT 开始应用于大型能力和成就测验,并且也有研究者尝试将项目反应理论和技术应用于人格测验^[2,3]。但直到本世纪初,随着相关技术的进一步发展,才有研究者开发出基于项目反应理论的计算机自适应人格测验,结果证明将 CAT 应用到人格测验可以有效地提高测试效率^[4-8]。艾森克个性问卷(Eysenck Personality Questionnaire, EPQ)是国内外常用的人格量表之一,20 世纪 80 年代引入我国,多位学者对其进行了修订^[9-12],并得到了广泛的使用。2010 年,我们曾经通过模拟研究探讨了将 CAT 方式用于 EPQ 的可行性和有效性。结果显示采用 2PLM 模型来拟合数据,在最大信息量选题策略和最小误差法终止条件下, CAT 方式能够节省一半以上的项目,同时

还能达到与纸笔测验相当的测量精准度, CAT 分数和纸笔分数在四个分量上的相关系数也达到或接近高相关^[13]。我们在此基础上发展出艾森克个性问卷(成人版)计算机自适应测验(A Computerized Adaptive Test of Eysenck Personality Questionnaire[Adult], EPQ-A-CAT), 本研究即对这个新型测验的实际应用进行测量学检验。

1 对象与方法

1.1 对象

本研究样本来自深圳大学学生,共 204 名被试,年龄在 18~29 岁之间,其中男生 101 名,平均年龄 22.04±2.74 岁,女生 103 名,平均年龄 21.93±2.58 岁。其中 120 名被试还同时进行了纸笔版的测试,作为两种测验方式的效度分析。另选取 31 名被试,其中男生 14 名,女生 17 名,进行了间隔 3~4 周的 CAT 重测,对其结果进行重测信度的分析。

1.2 工具

1.2.1 艾森克个性问卷(成人版) 包括内—外向(Extraversion,E)、情绪稳定性(Neuroticism,N)、精神质(Psychoticism,P)和掩饰性(Lie,L)四个分量表,共88个项目。采用纸—笔方式测试,计分方法为0、1记分。1984年由龚耀先等修订并建立全国常模,2008年再次采样更新常模。

1.2.2 艾森克个性问卷(成人版)计算机自适应测验(EPQ-A-CAT) 其题库保留了全部量表的88个项目。具体算法是:四个分量表的初始能力估计值设为0;选题策略是最大测验信息量法;采用期望后验估计方法估计潜在特质水平;终止规则为最小误差法。EPQ-A-CAT采用人机对话方式施测,程序依照E→N→P→L的顺序呈现项目,直到四个分量表的终止规则都满足时方可停止测验。

2 结 果

2.1 测量精度

在CAT中不同的特质水平其估计标准误也不同,表1中列出了四个分量表的估计标准误。

CAT特质水平估计标准误均值在四个分量表中都小于CAT终止规则预设的最小误差值。

31个被试在间隔3~4周后进行重测,两次测试结果的相关系数分别为0.79(E)、0.85(N)、0.67(P)、0.71(L)。

2.2 测试效率

表2表明,CAT总体上能够节省一半以上的项目,从各分量表来看,E和N分量表节省了60%以上的项目,L分量表能节省70%以上的项目,而P分量表节省了30%以上的项目。

表1 CAT特质水平估计标准误的描述性统计

分量表	最小误差值	均值	标准差	最大值	最小值
E	0.45	0.42	0.02	0.50	0.37
N	0.41	0.38	0.03	0.47	0.33
P	0.58	0.51	0.05	0.69	0.42
L	0.43	0.40	0.02	0.43	0.36

表2 施测的项目数量

	总体	E分量表	N分量表	P分量表	L分量表
纸—笔版的项目数	88	21	24	23	20
CAT使用项目数均值	37.7	7.7	9.3	15.3	5.3
CAT使用项目数标准差	13.5	5.7	7.2	6.5	2.9
CAT使用项目数所占百分比	42.8	36.8	38.9	66.6	26.4
CAT使用项目数的最小值	15	3	4	4	3
CAT使用项目数的最大值	82	21	24	23	20

2.3 效度分析

120名被试同时进行EPQ纸—笔版和EPQ-A-CAT测试,为了平衡的顺序效应,60名被试先做

纸—笔版后做CAT版,另60名顺序相反。首先对顺序效应进行检验,结果见表3。

从表3来看,四个分量表中组1和组2无论纸笔分数还是CAT分数,均值差异都没有达到显著性,说明顺序效应不明显。

两个版本测试结果的积差相关系数在E、N、P、L四个分量表中分别为0.85、0.81、0.75和0.74。

表3 两种不同的测试顺序的结果比较

	纸笔分数				CAT分数			
	组1	组2	P值	效果量	组1	组2	P值	效果量
E分量表	48.89	50.41	>0.05	0.086	0.076	0.010	>0.05	0.116
N分量表	47.46	49.17	>0.05	0.065	-0.209	0.076	>0.05	0.013
P分量表	41.17	44.78	>0.05	0.003	0.258	0.377	>0.05	0.048
L分量表	43.45	40.85	>0.05	0.025	-0.971	-1.112	>0.05	0.038

注:组1为先纸笔后CAT,组2为先CAT后纸笔。效果量指标为Cohen's d。

3 讨 论

从特质水平估计标准误的均值来看,每个分量表的均值都小于最小误差值,而最小误差值是从纸—笔测验的信度计算得来,这说明与EPQ纸—笔测验相比,EPQ-A-CAT对于被试者特质水平估计的准确性要更高。这可能因为在CAT测试中,选取的项目是跟特质水平相匹配的,匹配得越好,测量的误差就会越小,因而在这样的自适应性测验中,大多数被试特质水平的估计标准误都低于最小误差值。从特质水平估计标准误的标准差来看,四个分量表的值都很小,说明该测验对每名被试特质水平估计的准确性比较接近,这也反映了EPQ-A-CAT有较好的公平性。

本研究结果显示EPQ-A-CAT的重测信度为E=0.79,N=0.85,P=0.67,L=0.71。表明EPQ-A-CAT版本测试结果的稳定性较好。但如果以戴晓阳、曹亦薇推荐的0.75作为评价分量表重测信度的标准^[14],EPQ-A-CAT的E和N量表都能达到,而P量表则稍差一些。可能原因一是与重测样本量过少有关,在样本量不大的情况下,相关系数的稳定性不好,而且相关系数易受极端值的影响。但是更有可能与P分量表的测量准确性不高有关,当P分量表不能很准确地估计出潜在特质水平的情况下,其估计值并不能很好的代表真值,因此两次测试结果的相关也不会很高。

从EPQ-A-CAT施测的项目数量来看,CAT版式明显提高了测试效率。E、N和L三个分量表能够节省一半以上的项目,这可能由于这三个分量表的项目难度分布比较广,对于潜在特质水平比较宽范围的被试都有比较好的测量效果。而P分量表只节

省了 1/3 的项目,这可能是由于 P 分量表本身的项目难度分布不广,对于潜在特质水平比较低的被试测量的正确性不高有关。如果被试这种特质的水平较低时,CAT 从题库中挑选不到一定数量难度匹配的项目对其进行测试,因此估计标准误也就很难快速达到测验终止的最小标准误值,所以需要挑选比较多的项目来测试,有时甚至是全部的项目测试完之后还不能达到终止的误差标准。Ferrando 在研究中也发现,P 量表对于精神质水平高的被试 ($\theta=1.88$) 特质水平估计标准误最小^[15]。而本次参加 EPQ-R-CAT 的被试均属于普通人群,这些人的 P 潜在特质水平大部分都不会很高(均值为 0.336),因而需要更多的测试项目,对特质水平的估计标准误才能达到最小标准误终止规则。

参 考 文 献

- 1 Weiss DJ. Adaptive testing by computer. *Journal of Consulting and Clinical Psychology*, 1985, 53: 774-789
- 2 Waller NG, Reise SP. Computerized adaptive personality assessment: An illustration with the absorption scale. *Journal of Personality and Social Psychology*, 1989, 57: 1071-1058
- 3 Roper BL, Ben-Porath YS, Butcher JN. Comparability and validity of CAT with the MMPI-2. *Journal of Personality Assessment*, 1995, 65(2): 358-371
- 4 Forbey JD, Handel RW, Ben-Porath YS. A real data simulation of computerized adaptive administration of the MMPI

- A. *Computers in Human Behavior*, 2000, 16: 83-96
- 5 Reise SP, Henson JM. Computerization and adaptive administration of the NEO PI-R. *Assessment*, 2000, 7(4): 347-364
- 6 Fliege H, Becker J, et al. Development of a computer-adaptive test for depression(D-CAT). *Quality of Life Research*, 2005, 14: 2277-2291
- 7 Fliege H, Becker J, et al. Evaluation of a computer-adaptive test for the assessment of depression. *International Journal of Methods in Psychiatric Research*, 2009, 18(1): 23-36
- 8 Becker J, Fliege H, et al. Functioning and validity of a computerized adaptive test to measure anxiety(A-CAT). *Depression and Anxiety*, 2008, 25: E182-E194
- 9 龚耀先. 艾森克个性问卷在我国的修订. *心理科学*, 1984, 4: 11-18
- 10 陈仲庚. 艾森克人格问卷的项目分析. *心理学报*, 1983, 2: 211-217
- 11 朱昌明,刘协和,等. 艾森克个性问卷(成人)成都地区试用报告. *四川医学院学报*, 1983, 14(2): 155-159
- 12 钱铭怡,武国城,等. 艾森克人格问卷简式量表中国版(EPQ-RSC)的修订. *心理学报*, 2000, 32(3): 317-323
- 13 雷辉,戴晓阳. 计算机自适应测验方式在艾森克人格问卷中的应用. *中国临床心理学杂志*, 2011, 19(3): 306-308
- 14 戴晓阳,曹亦薇. 心理评定量表的编制和修订中存在的一些问题. *中国临床心理学杂志*, 2009, 17(5): 562-565
- 15 Ferrando PJ. The accuracy of the E, N and P trait estimates: an empirical study using the EPQ-R. *Personality and Individual Differences*, 2003, 34: 665-679

(收稿日期:2011-10-10)

(上接第 362 页)

而原本的施暴方也受到了家庭暴力。本研究对于影响因素的探究只部分解释家庭暴力的发生,对于个体的其他方面,如人格因素、社会因素等,有待后续研究者进行深入。(致谢:衷心感谢全国各地课题协作单位的负责人,胡蕾、任双成、于素兰、苏巧玲、于桂新、麻超、王桂新。)

参 考 文 献

- 1 Krug EG, et al. The world report on violence and health. *The Lancet*, 2002, 360(9339): 1083-1088
- 2 Straus MA, Gelles RJ. Violence in American families: How much is there and why does it occur? *Troubled relationships. Families in Trouble Series*, 1988, 3: 141-162
- 3 张李玺,等. 中国家庭暴力研究. 北京:中国社会科学出版社,2004

- 4 Straus MA, Douglas EM. A short form of the revised conflict tactics scales, and typologies for severity and mutuality. *Violence and Victims*, 2004, 19: 507-520
- 5 钟迪科. 家庭冷暴力和防治对策刍议. *法制与社会*, 2011, 23: 182-183
- 6 Panda P, Panda B. Agarwal, marital violence, human development and women's property status in India. *World Development*, 2005, 33(5): 823-850
- 7 曹玉萍,等. 家庭暴力的表现形式及其相关因素的比较研究. *中国临床心理学杂志*, 2008, 16(1): 28-30, 65
- 8 邹韶红,等. 夫妻暴力社会人口学及社会文化因素的研究. *中国临床心理学杂志*, 2010, 18(3): 326-328
- 9 Rodriguez E, et al. Family violence, employment status, welfare benefits, and alcohol drinking in the United States: What is the relation? *Journal of Epidemiology and Community Health*, 2001, 55(3): 172-178

(收稿日期:2011-12-14)