

# 个人拟合指标在艾森克人格测验中的应用

刘拓, 曹亦薇, 戴晓阳

(深圳大学心理学系, 广东 深圳 518060)

**【摘要】** 目的:用个人拟合指标对 EPQ 量表数据进行分析,探讨不拟合现象的可能成因和不拟合被试对参数估计的影响。方法:用  $I_z$  和 Guttman 错误检查不拟合情况,用 RMSD、MD 指标及 ICC 面积法考察参数估计结果的差异。结果:①  $I_z < -1.65$  的被试占总人数的 3.44%,  $I_z$  与 Guttman 错误的相关为  $-0.838$ ;②最不拟合的 50 人中,编号相连被试共有 32 人;③删去不拟合被试前后,特质参数差异很小,高区分度高难度和高区分度低难度项目参数差异大。结论:个人拟合指标可以找到团体测验中来自同一测试环境的异常反应被试集团;不拟合被试的存在,对被试特质参数的估计影响不大,但会造成对项目区分的低估,进而影响难度越大和越小的项目。

**【关键词】** 项目反应理论; EPQ; 个人拟合指标; ICC

中图分类号: R395.1

文献标识码: A

文章编号: 1005-3611(2011)03-0323-04

## An Application of Person-fit Indices to EPQ

LIU Tuo, CAO Yi-wei, DAI Xiao-yang

Department of Psychology, Shenzhen University, Shenzhen 518060, China

**【Abstract】 Objective:** Person-fit indices were computed on EPQ. This study was to investigate the cause of misfit and the influence of the inclusion of misfitting persons in parameter estimates. **Methods:**  $I_z$  and Guttman error were used to investigate the misfit. MD, RMSD indices and ICC method was used to evaluate the effect of deletion of misfitting persons. **Results:** ①  $I_z < -1.65$  was found in 3.44% of the examinees, and  $I_z$  was correlated  $-0.838$  with Guttman error; ② There were 32 in 50 most misfitting persons ranked in adjacent numbers; ③ It was found that person parameters changed little after deleting misfitting persons. However, the high  $\alpha$  high  $\beta$  items and the high  $\alpha$  low  $\beta$  items changed a lot. **Conclusion:** A group of aberrant response persons in the same testing environment were found out by person-fit indices in group test. The presence of misfitting person cannot influence person parameters much, but resulted in underestimating discrimination parameters, and then influenced the bias in estimation of higher or lower threshold parameters.

**【Key words】** IRT; EPQ; Person-fit Index; ICC

心理测验通过被试对量表的作答来了解被试的某种心理特质水平,在项目反应理论(IRT)中反映为特质参数<sup>[1]</sup>。但在实际测验时,由于各种原因(如猜测、粗心、无动机、社会赞许性等等因素)会造成某些被试的反应模式无法正确反映出其心理特质水平,导致心理特质水平被高估或低估的情况。例如在人格测验中就经常会出现由于被试胡乱作答、伪装等情况造成测验结果不能真实反映其人格特质的现象。这种与被试特质水平不匹配的反应模式,称为异常反应模式,在心理测量学中则表现为被试反应模式与测量模型的不拟合,称为个人不拟合(person misfit)现象。

个人拟合指标(person fit index)是探测被试反应模式与测量模型拟合程度的指标。近年来研究者开发了很多指标如  $I_z$ 、U3 等用于测量个人拟合的问题。其中较为著名的指标是  $I_z$ ,它以项目反应理论(IRT)为基础,通过计算每个被试的标准化似然函数值而得到的<sup>[2]</sup>,不少研究者认为这是一个较好的个

人拟合指标<sup>[3]</sup>。从原理上说, $I_z$  指标属于 Guttman 模型的一种<sup>[4]</sup>。所谓 Guttman 模型是指被试应该符合对高过其特质水平的项目选“否”,而对低于其特质水平的的项目选“是”的假设。若违反这个假设,就认为是错误的。在经典测量理论中常通过计算违反 Guttman 模型的错误个数来描述被试异常反应的程度,这个直观的非参数指标称为 Guttman 错误。Meijer 曾把 Guttman 错误与 U3 指标进行比较证明了它的有效性<sup>[5]</sup>,后又用它来分析被试的反应模式<sup>[6]</sup>。而  $I_z$  指标和 Guttman 错误的差异比较还未见过报告。

总体而言,个人拟合领域的多数研究者较为关注个人拟合指标在不同情况下对异常反应模式的检测效果或是多个指标间的效果比较<sup>[7-10]</sup>。尽管不少研究者认为测验中被试的欺骗、粗心、瞌睡、伪装、无动机等行为可能导致个人不拟合<sup>[4,11]</sup>,如 Zickar 和 Drasgow 就建议个人拟合指标可用于探查被试在人格测验中的掩饰行为<sup>[12]</sup>。但是,对个人不拟合原因的实质性探讨并不多见。国内曹亦薇曾用 MSD(多维标度法)对不拟合的被试在词汇测验中所选的干扰项特

点进行分类,进而探索这些被试在辨析词义时的思维方式<sup>[13]</sup>。但这些研究存在相当的局限性,一是多数学者使用单参数的 Rasch 模型,而很多心理数据并不适合 Rasch 模型;二是样本中不拟合被试的存在究竟会对模型的参数估计造成怎样的影响并无定论。目前可以查到的文献有 Meijer 以及 Schmitt 认为不拟合被试会影响到测验的效度<sup>[14,15]</sup>, Philips 和 Curtis 基于 Rasch 模型的数据分别探讨了不拟合被试在等值和态度测验中的影响情况<sup>[16,17]</sup>等少量资料。

本研究使用个人拟合指标对团体施测的艾森克人格问卷(EPQ)结果进行检查,主要目的在于:第一,通过  $I_z$  指标与 Guttman 错误的分析,探讨测验中个人不拟合情况出现的可能原因;第二,考察样本中的不拟合被试对被试特质参数以及项目参数估计的影响情况。

## 1 方 法

### 1.1 数据与模型

本研究使用龚耀先修订的 EPQ 问卷成人版,被试来自于深圳市 15 所中学的高中学生,共 1860 人。根据探索性因素分析结果以及前人的研究经验,本研究决定选择单因素结构的内-外向(E)分量表的数据作为分析对象,因为使用 IRT 模型必须满足单维性的前提。然后通过似然比的模型拟合检验,发现 IRT 的二、三参数模型 (2PL、3PL)<sup>[18]</sup> 拟合情况优于 Rasch 模型,但考虑到三参数模型中的猜测参数在人格测验中意义不大<sup>[19]</sup>,故在本研究中使用 2PL 模型。

### 1.2 个人拟合指标

研究选用了两个个人拟合指标  $I_z$  与 Guttman 错误。一般认为  $I_z < -1.65$  则为不拟合<sup>[20]</sup>,而 Guttman 错误指标值越大,则违反 Guttman 模型的项目对的数目越多,即越不拟合<sup>[5]</sup>。研究中采用 Guttman 错误作为  $I_z$  指标的一种更直观的佐证,并把两种指标筛选的被试进行比对。

### 1.3 使用软件

本研究中项目参数和被试能力参数的估计使用 BILOG-MG 软件(Version3.0 版),个人拟合指标  $I_z$  和 Guttman 错误的计算在 SAS 统计软件 9.1 实现,项目参数和能力参数的差异计算,以及 ICC 曲线面积的差异计算使用 RGui 软件 2.11.1 版。

## 2 结 果

### 2.1 拟合指标的检测结果

首先使用 BILOG-MG 软件估计出被试特质参数和项目参数值,然后使用自编的 SAS 程序计算被试的  $I_z$  和 Guttman 错误。根据  $I_z < -1.65$  的标准,有 64 人被检出个人不拟合,占全体被试(1860)的 3.44%。算得  $I_z$  指标与 E 量表总分的相关为 -0.080,说明基于 E 量表计算出的  $I_z$  指标与量表总分无关。而  $I_z$  指标与 L 量表总分的相关为 -0.047,这说明  $I_z$  指标与艾森克的说谎量表得分无线性关系,此结果与 Ferrando 的部分研究结果一致<sup>[9]</sup>。

$I_z$  指标与 Guttman 错误的相关为 -0.838,意味着  $I_z$  指标越小,则 Guttman 错误的个数就越大,就相关系数的绝对值来说属于高相关。进一步比对 Guttman 错误的最大值和  $I_z$  最小值的前 20、30 和 50 人,其中编号相同的被试人数分别为 12、21 和 34 人,各占 60%、70%、68%。抽取两个指标检测出的最不拟合的 50 名被试进行分析,发现其中有两批被试的序号相连或邻近,他们是 A 学校的 22 人,序号在 131~158 号之中;B 学校的 10 人,序号在 127~142 号之中。根据原始数据编码规则认定,相邻编号的被试不仅属于同一个班,更是同教室某个区域座位紧邻的一批学生。结合当时团体施测的实际情况,故考虑这种异常反应情况可能与当时作答环境有关,如喧闹或相互交谈,班级气氛或某种情绪使这些学生缺乏答题兴趣、胡乱作答。

### 2.2 不拟合被试对参数估计的影响

2.2.1 不拟合被试对特质参数估计的影响 尝试删除由  $I_z$  指标检出的 64 名不拟合被试,观察删除前后特质参数估计的变化。见表 1。表 1 中的 MD(平均差异)与 RMSD(差异均方根)指标都是用于评价两个变量之差的指标。前者通过计算两变量差异绝对值的平均数,后者是计算两变量差异平方和的平均方根,它们的值越接近于 0 则表示差异越小。表 1 结果显示:特质参数的全距略微变小,MD 为 0.043、RMSD 为 0.055。这一结果反映了数据中异常反应被试的存在对正常作答被试的特质估计的影响可能不会太大。

表 1 删除异常反应被试前、后被试特质参数的比较

	平均值	标准差	全距	MD (RMSD)
删除前	0.005	0.972	4.646	0.043
删除后	0.000	0.944	4.625	(0.055)

2.2.2 不拟合被试对项目参数估计的影响 考察不拟合被试对项目参数估计的影响,首先分别估计删除不拟合被试前后的区分度与难度参数( $\alpha_1$  与  $\alpha_2$ 、 $\beta_1$

与  $\beta_2$  分别为删除前、后的区分度和难度的估计值)。比较方法有两种:一是直接观察,通过区分度倍率( $\alpha_2/\alpha_1$ )以及难度之差( $\beta_2-\beta_1$ )以及 MD、RMSD 值来观察删除不拟合被试的影响(见表 2);二是面积法,主要是模仿项目功能差异的做法,来调查区分度与难度变化的综合效果。直接比较区分度的结果:①区分度倍率变化最大的是 17 题与 53 题,分别为 1.527 和 1.774 倍,然后是 49 题与 33 题分别为 1.340 和 1.209 倍,最小的是 65 题,显示出区分度越大倍率变化越大的倾向。②除 65 题外,删除后的项目区分度全部有提高,反映了数据中不拟合被试的存在容易出现区分度低估的现象。难度变化方面,难度参数差( $\beta_2-\beta_1$ )最大的是 13 题为 0.309,其次是 41 题与 49 题分别为 0.123 和 0.138,难度差总平均为 0.034,

MD、RMSD 分别为 0.077 与 0.103。对照表 1,不拟合被试的存在对难度参数总体也有低估的影响,并且较之特质参数,其变化要明显得多。我们再借用 DIF(项目功能差异)研究中经常使用的面积法进一步探讨这两个指标的变化规律。面积法指的是比较 ICC(项目特征曲线)下的面积差,项目特征曲线是项目区分度和难度的共同体现,若项目参数相等,则 ICC 曲线重合,ICC 下的面积差为 0;反之项目参数差异越大,则 ICC 下的面积差越大。见表 2。表 2 中题号是按  $\alpha_1 \times \beta_1$  的值由大到小排列的。结果显示,题目的  $\alpha_1 \times \beta_1$  值越大或越小,其相应的面积差异也越大,即受到不拟合被试的影响越严重。结合直接比较的结果,发现若两项目区分度接近,则难度越大或越小的项目受到不拟合被试的影响越严重。

表 2 删除不拟合被试前、后项目参数比较表

题号	直接比较						面积法	
	$\alpha_1$	$\alpha_2$	$\alpha_2/\alpha_1$	$\beta_1$	$\beta_2$	$\beta_2-\beta_1$	$\alpha_1 \times \beta_1$	面积差
17	0.715	1.092	(+)1.527	1.536	1.637	0.101	1.098	0.495
49	1.396	1.871	(+)1.340	0.613	0.751	0.138	0.856	0.272
13	0.397	0.440	(+)1.108	0.997	1.306	0.309	0.396	0.256
21	0.838	0.873	(+)1.042	0.076	0.190	0.114	0.064	0.110
41	1.151	1.265	(+)1.099	0.026	0.149	0.123	0.030	0.145
25	1.100	1.192	(+)1.084	-0.059	0.024	0.083	-0.065	0.112
65	0.497	0.487	(-)0.980	-0.132	-0.061	0.071	-0.066	0.055
55	0.942	0.992	(+)1.053	-0.235	-0.193	0.042	-0.221	0.071
61	0.389	0.421	(+)1.082	-0.732	-0.662	0.072	-0.285	0.096
37	1.387	1.500	(+)1.081	-0.241	-0.240	0.001	-0.334	0.073
84	1.475	1.595	(+)1.081	-0.316	-0.291	0.025	-0.466	0.072
80	0.837	0.913	(+)1.091	-0.693	-0.768	-0.075	-0.580	0.125
14	1.486	1.627	(+)1.095	-0.435	-0.428	0.007	-0.646	0.079
29	1.133	1.230	(+)1.086	-0.640	-0.655	-0.015	-0.725	0.089
45	1.190	1.265	(+)1.061	-0.617	-0.675	-0.058	-0.734	0.079
05	1.526	1.651	(+)1.082	-0.531	-0.578	-0.047	-0.810	0.078
71	1.022	1.106	(+)1.082	-0.915	-0.978	-0.063	-0.925	0.104
01	0.814	0.947	(+)1.163	-1.357	-1.466	-0.109	-1.105	0.203
10	2.462	2.568	(+)1.043	-0.507	-0.503	0.004	-1.248	0.023
33	1.471	1.779	(+)1.209	-0.919	-1.000	-0.081	-1.352	0.172
53	1.119	1.985	(+)1.774	-1.902	-1.822	0.080	-2.128	0.468

注:表中“(+)”、“(-)”表示变化方向,“(+)”说明  $\alpha_2 > \alpha_1$ ,反之  $\alpha_2 < \alpha_1$

### 3 小 结

本研究应用个人拟合指标  $I_x$  对艾森克人格测验的内-外向分量表的数据进行检测,发现 3.44% 的被试存出异常反应模式。本研究发现  $I_x$  与 Guttman 错误具有相当高的一致性(-0.838)、但是与 L 量表(说谎量表)得分几乎没有相关(0.047)。调查不拟合最严重的 50 名被试,发现有 60% 的被试编号是连续的,这为今后考察个人不拟合的原因可能由测验环境、被试无动机等提供了有力的线索。

从本研究的分析结果来看,不拟合被试的存在

虽然对正常作答被试的特质估计影响不大,但对项目参数估计(无论是区分度与难度)都会出现低估倾向,尤其是对区分度的影响更为明显。这些结果是否具有普遍性,还须进一步确认。

#### 参 考 文 献

- 1 刘仁刚. 条目反应理论简述. 中国临床心理学杂志, 2009, 17(1): 37-41
- 2 Drasgow F, Levine M, Williams EA. Appropriateness measurement with polychotomous item response models and standardized indices. British Journal of Mathematical and

- Statistics Psychology, 1985, 38: 67-86
- 3 Nering ML, Meijer RR. A comparison of the person response function and the  $I_2$  person-fit statistic. *Applied Psychological Measurement*, 1998, 22: 53-69
  - 4 Meijer RR, Sijtsma K. Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 2001, 25(2): 107-135
  - 5 Meijer RR. The number of guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement*, 1994, 18: 311-314
  - 6 Meijer RR, Egberink IJL, Emons WHM, et al. Detection and validation of unscalable item score patterns using item response theory: An illustration with harter's self-perception profile for children. *Journal of Personality Assessment*, 2008, 90(3): 227-238
  - 7 Reise SP, Due AM. The influence of test characteristics on the detection of aberrant response patterns. *Applied Psychological Measurement*, 1991, 15(3): 217-226
  - 8 Karabatsos G. Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 2003, 16(4): 277-298
  - 9 Drasgow F, Levine MV, McLaughlin ME. Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement*, 1987, 11: 59-79
  - 10 Meijer RR. Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods*, 2003, 8(1): 72-87
  - 11 王昭,郭庆科,岳艳. 心理测验中个人拟合研究的回顾与展望. *心理科学进展*, 2007, 3: 559-566
  - 12 Zickar MJ, Drasgow F. Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 1996, 20: 71-87
  - 13 曹亦薇. 异常反应模式的识别和分类. *心理学报*, 2001, 6: 558-563
  - 14 Meijer RR, Nering ML. Trait level estimation for nonfitting response vectors. *Applied Psychological Measurement*, 1997, 21(4): 321-336
  - 15 Schmitt N, Cortina JM, Whitney DJ. Appropriateness fit and criterion-related validity. *Applied Psychological Measurement*, 1993, 17: 143-150
  - 16 Phillips SE. The effects of the deletion of misfitting persons on vertical equating via the rasch model. *Journal of Educational Measurement*, 1986, 23(2): 107-118
  - 17 Curtis DD. Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal*, 2004, 5(2): 125-143
  - 18 Embretson SE, Reise SP. *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum, 2000
  - 19 Ferrando PJ, Chico E. Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement*, 2001, 61(6): 997-1012
  - 20 Meijer RR. The influence of the presence of deviant item score patterns on the power of a person-fit statistic. *Applied Psychological Measurement*, 1996, 20: 141-154

(收稿日期:2010-12-30)

(上接第 354 页)

本研究的主要目的是探讨社会支持、自我认知与抑郁三者之间的关系。结果表明,低社会支持的个体,其自我评价也较低,抑郁发生水平较高。其中,自我认知对抑郁具有直接的影响作用,并在社会支持与抑郁之间起到了中介作用。这一结果验证了本文提出的研究假设。从预防青少年抑郁的角度来考虑,高社会支持的孩子,因为获得了足够的关心,有效的帮助,从而能够成功应对负性的生活事件,进而肯定了自己的能力,同时也因为自己被关注,被接纳而更加有自信。因此,该研究的结果对于预防青少年抑郁有一定的指导意义。在负性生活事件发生之后,父母、老师等应积极的关注青少年,让他们感受到真正的关心和帮助,成功的应对事件,提高自信心,减少抑郁的发生。

## 参 考 文 献

- 1 Liu XC, Ma DD, Kurita H, et al. Self-reported depressive symptoms among Chinese adolescents. *Social Psychiatry and Psychiatric Epidemiology*, 1999, 34(1): 44-47
- 2 冯正直,张大均. 中学生抑郁症状流行病学特征. *中国行为医学科学*, 2005, 14(2): 103-105
- 3 Marco C, Christoph B, Niels B. Gender differences in response to pictures of nudes: A magnetoencephalographic study. *Biological Psychology*, 2003, 63: 129-147
- 4 Beck AT. *Cognitive therapy and the emotional disorders*. New York: International Universities Press, 1976
- 5 Beck AT. *Cognitive models of depression*. *Journal of Cognitive Psychotherapy*, 1987, 1: 5-37
- 6 Harter S. *Manual for the self-perception profile for adolescents*. Denver: University of Denver, 1988
- 7 Harter S. *Manual for the social support scale for children and adolescents*. Denver: University of Denver, 1985. 156-159
- 8 杨荃芹. 中国湖南省城乡高中生抑郁易感因素及预测模型追踪研究. 中南大学博士论文, 2007
- 9 陈祉妍,杨小冬,李新影. 流调中心抑郁量表在我国青少年中的试用. *中国临床心理学杂志*, 2009, 17(4): 443-448

(收稿日期:2010-12-22)