

个人拟合指标在Likert型人格测验中的应用

王昭¹, 郭庆科²

(1.大连医科大学人文与社会科学学院,大连 116044;2.山东师范大学心理学院,济南 250014)

【摘要】 目的:以 Guttman 理想模型和等级计分项目反应理论作为基础,探讨似然估计个人拟合指数 L 在 Likert 人格量表中的有效性。**方法:**模拟研究采用 Monte Carlo 方法,项目数为 40,0,1,2 三等级计分。项目参数包括难度和区分度,样本容量为 1000,一次实验中重复模拟 100 次;实际数据以大五人格简化版(FFI)为研究工具,对 1108 名大学生和高中生进行施测,考察 L 对 FFI 信效度的影响。**结果:**模拟研究表明 L 在 Likert 量表中接近正态分布,L 对各类偏差个体均具有较好的检测效果,对各类偏差反应正确拒绝率均在 95%左右,对于随机反应和折中反应报准率为 100%;实际研究结果表明删除不同比例拟合较差个体后量表 α 信度和结构效度均明显增加,删除 50% 不拟合个体后 A、C、E、N、O 量表 α 信度分别由原来的 0.635、0.789、0.725、0.830、0.714 提升至 0.727、0.871、0.823、0.913、0.841, χ^2/df 分别由原来的 7.594、6.999、10.872、16.445、14.047 降至 3.113、2.432、3.635、3.784、7.093。**结论:**个人拟合指数 L 在 Likert 人格量表中能够有效地检测反应偏差个体。

【关键词】 个人拟合指标; 等级项目反应理论; Likert 量表; Monte Carlo 模拟

中图分类号: R395.1

DOI: 10.16128/j.cnki.1005-3611.2016.03.019

The Application of Person-fit Statistics in Likert Personality Scales

WANG Zhao¹, GUO Qing-ke²

¹Humanities and Social Sciences Department of Dalian Medical University, Dalian 116044;

²Department of Psychology, Shandong Normal University, Jinan, 250014

【Abstract】 Objective: To explore the fit of L in Likert-type scales. This article proposes L for assessing individual fit based on two theory models: the Guttman's psychometric model and Graded response mode. **Methods:** Carlo Monte data simulation method was used, the number of items was 40, the score was 0, 1 and 2. Project parameters included difficulty and discrimination. The sample size was 1000 and 100 was repeated under an experimental condition. In the practical research, using the Big Five Personality Questionnaire(FFI), 1108 college and high school students were tested to investigate the effect of L on the reliability and validity of FFI. **Results:** Simulation studies showed that L was close to normal distribution in the Likert scale. L could detect all kinds of aberrant response. Correct rejection rate for all kinds of aberrant response was about 95% and the accuracy rate of random response and compromise response were 100%. Practical research showed that the reliability and construct validity were promoted when different proportions of aberrant individuals were removed. When 50% aberrant individuals were deleted, the reliability of A, C, E, N and O increased from 0.635, 0.789, 0.725, 0.830, 0.714 to 0.727, 0.871, 0.823, 0.913, 0.841; χ^2/df decreased from 7.594, 6.999, 10.872, 16.445, 14.047 to 3.113, 2.432, 3.635, 3.784, 7.093. **Conclusion:** L can effectively detect the aberrant responses in Likert scales, and improve the reliability and validity of the test.

【Key words】 Person-fit statistics; GRM; Likert-type scale; Monte Carlo simulation

在心理测量领域,人格测验在心理统计学上的适应性要低于能力测验,这是由于人格测验中通常包含多个维度,各维度之间很难相互独立,除了项目的内容,人格还会影响被试的反应模式。研究表明反应模式会在很大程度上影响人格测验的结果^[1]。因此,越来越多的研究者关注人格测验的得分是否能反应所要测量的人格特征,或其他人格特征受非量表能力的影响^[2]。个人拟合指标(Person-fit statistics, PFS)即是对测验中的偏差反应个体进行筛查的一系列指标。因此,人格测验中的个人拟合研究逐渐受到了研究者们的重视。Zickar 和 Drasgow 检验

了 PFS 对人格测验中的不诚实行为的鉴别能力,发现 PFS 可以鉴别出不诚实行为^[3]。Ferrando 和 Eliseo 也进行了相似的研究,但发现 PFS 的检测效果并没有社会期望量表和测谎量表好,他们认为这是由于存在不诚实行为的被试占总人数的比率过高^[4]。Ponocny 和 Klauer 考察了 Rasch 模型对人格测验偏差行为的检测效果,但须对测验的结构进行认真的分析,同时还必须掌握可靠的关于诚实与不诚实反应的信息后 Rasch 模型才可应用于人格测验^[5]。国内学者将个人拟合指标用于艾森克人格测验,证实了 PFS 在人格测验中的有效性^[6,7]。研究表明在众多 PFS 中,似然估计指数 L 和 Lz 对偏差反应的检验效果最好^[7,8]。

似然估计指标 L 由 Levine 和 Rubin 于 1979 年提出, 是 PFS 中研究最为广泛的指数^[9], Drasgow 等于 1985 年提出了 L 的标准化形式 L_z ^[10], 当前更多的研究者将研究的重点针对于 L 和 L_z 的特征和检测效果。Raiche 等探讨了 L_z 的分布特征及影响因素^[11], Armstrong 也进一步探讨了 L_z 的特征及适用性^[12]; Torre 等采用不同分布来提升 L_z 的准确性^[13], Dodeen 将 L_z 应用于数学测验的研究, 发现学习态度、考试焦虑与 L_z 存在显著相关^[14]; Magis 等将 L_z 和 L_z^* 进行详细比较, 并将 L_z^* 用于言语技能测验^[15]; 刘拓等探讨了 L_z 对项目参数估计的净化效果^[16]。虽然大量研究表明 L 和 L_z 对偏差反应具有较好的检测效果, 并广泛应用于各种实际测验中。但目前为止, 多数研究局限于二级评分。当前所使用的很多人格测验都采用 Likert 量表形式, 而 Likert 量表中的个人拟合研究很少, 这也成为 L_z 应用于人格测验的重大局限。近年来, 部分研究者将研究重点关注于多级评分个人拟合指标的可行性。Emons 采用多级项目反应理论对不拟合个体进行检测, 研究发现 L_z 能够筛查出较明显的局部偏差反应^[17]; Ferrando 曾在线性因素分析的基础上对个人拟合指标在连续反应模型中的分布特征进行探讨^[18,19]; Conijn 等采用多层逻辑回归法探讨 PFS 在等级项目的适用性^[20]; Tendeiro 采用 CUSUM 方法探讨 PFS 在 Likert 量表的适用性^[21]。虽然这些研究致力于 Likert 量表偏差反应的鉴别, 但较少有研究针对具体偏差反应类型和实际数据进行探讨。本研究的主要目的就是考察 L 应用于 Likert 量表的有效性。

本研究以 Guttman 理想模型^[22]和等级计分项目反应模型 (Graded response mode, GRM) 作为基础。Guttman 理想模型在 Likert 量表中应转化为: 心理测验可以被看作是一个线性标尺, 每个个体都可以在该标尺上找到各自的能力梯度点, 理想状态下, 个体在该能力梯度点以下所有相对简单项目上的得分应高于能力梯度点以上所有相对较难项目上的得分。GRM 是 1969 年 Samejima 提出的^[23], 把二分反应资料分析推广为对多级及连续反应资料的分析, 从那时起项目反应理论得以迅猛发展。二级计分模型可以被看作是 GRM 的特例^[24]。近几年来, 研究者发现, GRM 应用于 Likert 形式的人格测验中同样有效^[25]。

GRM 参数估计方法与 2 级评分项目反应理论相同, 其基本原理建立在 2 级计分的基础上。以 3 级分量表为例 (得分数分别为 0, 1, 2), $P_1(\theta)$, $P_2(\theta)$ 分别表示能力为 θ 的被试在该项目各计分等级的正确反

应概率, 即 $P_1(\theta)$ 表示得 1 分或 2 分的总概率, $P_2(\theta)$ 则表示得 2 分的概率, 那么各等级得分的概率应为^[24]:

得 0 分: $P(X=0|\theta)=1-P_1(\theta)$

得 1 分: $P(X=1|\theta)=P_1(\theta)-P_2(\theta)$

得 2 分: $P(X=2|\theta)=P_2(\theta)$

L 在二级计分测验中计算公式表示为所有项目得分概率的自然对数值和, 即当项目得分为 0 时, 取该能力水平个体得 0 分概率的自然对数 $\ln[1-P_g(\theta)]$, 当项目得分为 1 时, 取该能力水平个体得 1 分概率的自然对数 $\ln P_g(\theta)$ 。Likert 量表中 L 的含义同样为所有项目得分概率的自然对数和, 其计算公式表示为 $L = \sum_{g=1}^G \ln P_g(X_g|\theta)$

X_g 为被试在项目 g 上的得分, $P_g(X_g|\theta)$ 则表示能力水平为 θ 的被试得分为 X_g 的概率。

本研究目的是考察似然估计指数 L 应用于 Likert 人格量表的有效性。模拟数据考察 L 在 Likert 量表中的分布特征及其对人格测验中常见四种行为的检测效果; 实际数据使用大五人格问卷, 通过考察使用 L 作为筛查指标后, 测验信效度的变化, 来探测 L 对偏差反应的检测效果。

1 方 法

1.1 模拟数据

采用 Monte Carlo 数据模拟方法, 使用 SAS9.1 模拟数据, 项目数为 40^[26], 每个项目为 0, 1, 2 三级计分。项目参数包括难度和区分度。样本容量为 1000, 一次实验条件下重复模拟 100 次。

1.1.1 拟合样本 第一步: 获得一组样本容量为 1000 的标准正态分布 θ 值; 第二步: 定义难度参数 b_1 和 b_2 , 区分度参数 a 。根据等级评分项目反应理论, 为了保证 3 级评分的 $b_1 < b_2$, b_1 的取值范围为 $[-2, 1.9]$, b_2 的取值范围为 $[-1.9, 2]$, 相邻项目的难度间隔为 0.1; 区分度参数 a 为 $(0, 1)$ 的均匀分布。第三步: 根据 2 参逻辑斯蒂模型得到 p 矩阵, 若样本量用 n 表示, 项目数用 k 表示, 得到 $n \times k$ 的 p_1 矩阵和 p_2 矩阵。第四步: 根据 p 矩阵得到 $n \times k$ 的项目反应矩阵。第五步: 根据 p 和项目矩阵计算 L 值。

1.1.2 偏差样本 掩饰行为: 指被试为了掩饰症状, 倾向于选择较低的反应水平。即使是难度水平很低的项目, 其反应水平也很低。模拟数据中, 使难度水平最低的 8 个 (20%) 项目的得分为 0。

期望反应: 指为了符合社会期望, 被试在一些项目上表现出很高的反应水平, 模拟数据中, 难度水平最高的 8 个项目上的得分为 2。

随机反应:指被试不加区分的对项目进行随机反应,模拟数据中使偏差个体在各项目上的反应为联合分布,与 a, b, θ 值无关。

折中反应:指被试倾向于选择中间等级的选项,模拟数据中使所有奇数项目上的0分和2分改为1。

1.2 实际数据

1.2.1 被试 既参加初测又参加重测的总有效样本为1108人,男生328人,女生780人。大学生640人,来自两所普通高校,高中生468人,来自两所普通高中。

1.2.2 工具与施测 本研究采用大五人格测验简化版NEO-FFI(NEO Five Factor Inventory,以下简称FFI)。FFI包括神经质N,外倾性E,经验开放性O,宜人性A,认真性C共5个人格维度,包括60个项目,每个维度12题,采用5级评分。FFI已在中国得以应用。为了使项目长度适宜做个人拟合分析,研究中将FFI中的一半题目进行语义反转(即原来肯定陈述句改为否定陈述句,原来否定陈述句改为肯定陈述句)。而奇数题不变,形成量表用于初测,再把FFI中的另一半题目做语义反转形成重测量表。首先用初测量表对被试施测,施测时告诉被试要进行两次性格测试,以帮助他们了解自己的性格。初测后半个月再次进行重测。统计结果时将初测量表和重测量表合并,对合并量表进行探索性因素分析,将载荷小于0.3的项目成对删除,形成N、E、O、A、C五个分量表,保留项目数分别为20、20、18、20、20,总项目数为98。

2 结果

2.1 L的分布特征

根据模拟数据,拟合数据样本容量为1000,重复模拟次数100,总样本容量为100000。偏差样本中拟合数据样本容量950,偏差样本容量50,重复模拟次数100,总样本容量为100000,偏差样本比例为5%。拟合样本和各类偏差样本分布特征见表1。

结果表明,拟合样本L得分接近于正态分布,略呈负偏态,这与二级计分研究结果相接近。偏差样本L的分布形态与拟合样本有着显著不同。虽然样本中含偏差个体比例仅为5%,但对L得分的峰度和偏差产生了明显影响。

2.2 L对各类型偏差反应的检测效果

用信号检测理论来检验L的检测效果。根据L的计算公式,L得分越小表明偏差程度越高。根据小概率原理和拟合数据结果,将单侧检验第I类错

误率为0.05的个体,即 $L < -34.20$ 的个体定义为鉴别偏差个体用 y 来表示,则 $L \geq -34.20$ 的个体定义为鉴别拟合个体,用 n 表示;将模拟的偏差个体定义为实际偏差个体用SN来表示,模拟拟合个体被定义为实际拟合个体,用N表示,则信号检测率的计算方法见表2。L对各类偏差反应的检测率及判别指数 d' 见表3。研究发现L对各类偏差反应均有较高的报准率和正确拒绝率,特别是对随机反应和折中反应的报准率和正确拒绝率几乎达到理想状态。由此可见,L对于各类偏差行为均有较好的检测效果。

表1 L描述统计指标与分布特征

	平均数(M)	标准差(SD)	偏度(skew)	峰度(kurt)
拟合样本	-22.57	7.05	-0.15	0.27
掩饰反应	-23.95	10.87	-2.54	10.90
期望反应	-24.15	10.95	-2.50	10.64
随机反应	-26.98	21.60	-4.03	19.06
折中反应	-26.84	20.08	-3.58	13.87

表2 L检测率的计算方法

	Y	n
SN	f1 报准率 $P(y/SN)=f1/(f1+f2)$	f2 漏报率 $P(y/SN)=f2/(f1+f2)$
N	f3 虚报率 $P(y/N)=f3/(f3+f4)$	f4 正确拒绝率 $P(n/N)=f4/(f3+f4)$

表3 L对各偏差反应的检测率

偏差类型	f1 (%)	f2 (%)	f3 (%)	f4 (%)	d'
期望反应	81.6	18.4	5.3	94.7	2.520
掩饰反应	79.1	20.9	5.0	95.0	2.450
随机反应	100	0	5.2	94.8	5.615
折中反应	100	0	5.5	94.5	5.59

表4 删除不同比例不拟合个体后各量表 α 信度系数变化

删除比例	A量表	C量表	E量表	N量表	O量表
原始量表	0.635	0.789	0.725	0.830	0.714
删除5%	0.649	0.810	0.744	0.850	0.741
删除10%	0.668	0.821	0.758	0.861	0.757
删除20%	0.692	0.834	0.783	0.875	0.783
50%拟合	0.727	0.871	0.823	0.913	0.841
50%不拟合	0.573	0.700	0.616	0.703	0.560

2.3 L对FFI信度的影响

分别删除拟合较差的5%,10%,20%和50%的个体。考察量表信度的变化,结果见表4。结果表明删除不同比例拟合较差个体后量表 α 信度呈递增趋势。

2.4 L对FFI结果效度的影响

使用验证性因素分析考察PFS对结构效度的影响,选择 χ^2/df ,SRMR,RMSEA三个绝对拟合指数和两个相对拟合指数,结果见表5。随着非拟合个体删除比例的增加,模型拟合度逐渐升高,50%拟合个

体模型拟合指数优于50%不拟合个体。

表5 删除不同比例不拟合个体后各量表结构效度变化

量表	删除比例	χ^2/df	SRMR	RMSEA	NNFI	CFI
A量表	原始量表	7.594	0.065	0.074	0.73	0.76
	删除5%	7.246	0.066	0.074	0.76	0.78
	删除10%	6.792	0.064	0.073	0.77	0.79
	删除20%	6.109	0.063	0.073	0.80	0.82
	50%拟合	3.113	0.056	0.059	0.88	0.89
	50%不拟合	4.512	0.072	0.076	0.64	0.67
C量表	原始量表	6.999	0.056	0.071	0.88	0.89
	删除5%	5.509	0.049	0.063	0.92	0.93
	删除10%	4.851	0.047	0.060	0.93	0.94
	删除20%	3.493	0.041	0.051	0.95	0.96
	50%拟合	2.432	0.042	0.049	0.97	0.97
	50%不拟合	4.935	0.072	0.081	0.73	0.76
E量表	原始量表	10.872	0.075	0.090	0.73	0.76
	删除5%	8.492	0.066	0.081	0.79	0.81
	删除10%	7.145	0.061	0.075	0.82	0.84
	删除20%	5.541	0.057	0.069	0.87	0.89
	50%拟合	3.635	0.056	0.066	0.91	0.92
	50%不拟合	8.156	0.10	0.11	0.47	0.52
N量表	原始量表	16.445	0.078	0.11	0.83	0.85
	删除5%	8.214	0.055	0.079	0.91	0.92
	删除10%	7.198	0.052	0.076	0.93	0.94
	删除20%	5.947	0.049	0.072	0.95	0.95
	50%拟合	3.784	0.044	0.068	0.97	0.97
	50%不拟合	10.085	0.11	0.12	0.59	0.63
O量表	原始量表	14.047	0.082	0.10	0.60	0.65
	删除5%	12.570	0.076	0.10	0.67	0.71
	删除10%	11.576	0.073	0.099	0.71	0.75
	删除20%	10.435	0.072	0.099	0.75	0.78
	50%拟合	7.093	0.069	0.10	0.85	0.87
	50%不拟合	8.178	0.10	0.11	0.34	0.42

表6 各分量表间拟合指标L的相关系数

	A量表	C量表	E量表	N量表	O量表
A量表	1	0.598**	0.810**	0.600**	0.586**
C量表		1	0.627**	0.585**	0.570**
E量表			1	0.631**	0.600**
N量表				1	0.549**

2.5 各人格维度间L的相关

若L能够有效鉴别偏差个体,那么L在各人格维度之间应该具有较高的一致性。结果表明各人格维度之间L的相关较高(0.556–0.812),见表6。

3 讨 论

3.1 L指数应用于Likert量表的可行性

本研究通过三个方面证明L可应用于Likert人格测验,首先,模拟数据研究表明L具有较稳定的分

布特征,当个体能力水平 θ 呈正态分布时,L在Likert量表中同样呈正态分布,这与二级计分L分布特征相一致。而个体的人格特征通常呈正态分布,因此可以利用小概率原理推测偏差个体的存在。其次,模拟数据的信号检测率表明,L对人格测验中常见的各类偏差反应报准率接近100%,即只要个体存在偏差反应,其L值必然出现异常特征。虽然正确拒绝率并非100%,但接近95%。所以在实际研究中,当剔除L异常值个体后可能会少量存在漏报状况,但可以有效剔除偏差个体。第三,实际数据中的各种偏差反应个体与拟合个体相互混淆,偏差比例并不确定,但删除不同比例偏差个体后信效度产生了显著变化,表明L能有效剔除偏差个体。但由于L并非标准化指标,其判断拟合与否的临界值尚无法确定,因此L更适合群体的比较研究,对偏差反应需要结合具体测验的项目反应指标进行确定。

3.2 L在人格测验中的应用价值

由于L具有较稳定的分布特征,对人格测验中的偏差反应敏感,且有利于测验信效度的提高。因此,L可应用于对多级计分人格测验中。具体表现在:①目前大部分人格测验多采用因素分析法编制而成^[27, 28],虽然因素分析法更直观便捷地简化观测数据,使因素高度概括,但因素分析法需要大量观测数据支撑,而这些参与测量个体多为被动参与者,更易产生大量偏差个体,导致问卷编制过程中准确度下降。因此很多问卷在不同时间和群体中使用信效度并不稳定。若能在问卷编制过程中删除适当比例的偏差个体,将有效促进人格测验质量的提升。②当人格测验用于团体比较研究时,偏差个体的存在导致很多研究结果可信度下降,使用似然拟合指数L则能对此进行有效纠正;③心理测验的研究者发现,偏差反应类似于一种定势反应,与人格特征存在关联,而L指标将为该类研究提供有效工具;④若要求L指标用于个体偏差反应的筛查,需要在测验编制时提供与L指标相关的难度、区分度等参数指标,以确定具体测验L的临界值,这为测量学研究着提出更高的要求,也为人格测验的发展提供新的研究思路。

3.3 本研究的不足

研究采用模拟数据,而模拟数据与实际数据存在很大的差别,具体表现在以下几个方面:首先,模拟数据是先有 θ 值,再由 θ 计算 $P(\theta)$,最后生成项目得分。而实际数据则是一个相反的过程, θ 值是根据被试的项目得分来估计的,因此实际偏差反应被试可

能会由于项目得分的偏差造成 θ 值估计的偏差,从而造成对L计算的不准确;第二,模拟数据中a、b为均匀分布,而实际测验中这一条件很难满足,a、b的分布形态是否对L的分布产生影响需要进一步研究。第三,模拟数据为单维量表,项目数40,而实际人格测验均为多维量表,各维度之间存在相关,且每个维度项目数通常较少。而因素相关以及项目长度对L的影响还有待进一步检验。

本研究仅考察了偏差类型、偏差个体比例对L的影响。除此之外,L还会受到很多其他因素的影响。包括样本容量、偏差项目比例、评分等级量、 θ 估计方法、测验长度、测验因素数量及因素之间的相关等,这些都有待进一步研究。

参 考 文 献

- Paulhus DL, Reid DB. Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, 1991, 60(2): 307
- 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
- Zickar MJ, Drasgow F. Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 1996, 20(1): 71-87
- 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
- Ponocny I, Klauer KC. Towards identification of unscalable personality questionnaire respondents: The use of person fit indices. *Psychological Test and Assessment Modeling*, 2002, 44(1): 94
- 刘拓,曹亦薇,戴晓阳. 个人拟合指标在艾森克人格测验中的应用. *中国临床心理学杂志*, 2011, 19(3): 323-326
- 王昭,郭庆科,韩丹. 个人拟合指数对人格测验的影响及意义. *心理科学*, 2012, 35(5): 1225-1232
- Meijer RR, Sijtsma K. Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 2001, 25(2): 107-135
- Levine MV, Rubin DB. Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics*, 1979, 4(4): 269-290
- Drasgow F, Levine MV, Williams EA. Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 1985, 38: 67-86
- Raiche G, Blais JG. Characterization of the Distribution of the Lz Index of Person Fit According to the Estimated Proficiency Level. Online Submission, 2005. 28
- Armstrong RD, Stoumbos ZG, Kung MT, et al. On the performance of the lz person-fit statistic. *Practical Assessment, Research & Evaluation*, 2007
- De La Torre J, Deng W. Improving Person-Fit Assessment by Correcting the Ability Estimate and Its Reference Distribution. *Journal of Educational Measurement*, 2008, 45(2): 159-177
- Dodeen H, Darabi M. Person-fit: relationship with four personality tests in mathematics. *Research Papers in Education*, 2009, 24(1): 115-126
- Magis D, Raiche G, Béland S. A didactic presentation of Snijders' s lz* index of person fit with emphasis on response model selection and ability estimation. *Journal of Educational and Behavioral statistics*, 2012, 37(1): 57-81
- 刘拓,曹亦薇,戴晓阳. 个人不拟合对IRT项目参数估计的影响及净化对策. *中国临床心理学杂志*, 2011, 19(5): 622-624
- Emons WHM. Detection and diagnosis of person misfit from patterns of summed polytomous item scores. *Applied Psychological Measurement*, 2009, 33(8): 599-619
- Franndio PJ. Factor-analytic procedures for assessing response pattern scalability. *Multivariate Behavioral Research*, 2007, 42: 481-508
- Franndio PJ. Some statistics for assessing person-fit based on continuous-response models. *Applied Psychological Measurement*, 2010, 34(4): 219-237
- Conijn JM, Emons WHM, van Assen MALM, et al. On the usefulness of a multilevel logistic regression approach to person-fit analysis. *Multivariate Behavioral Research*, 2011, 46(2): 365-388
- Tendeiro JN, Meijer RR. A CUSUM to Detect Person Misfit: A Discussion and Some Alternatives for Existing Procedures. *Applied Psychological Measurement*, 2012, 36(5): 420-442
- Guttman L. A basis for scaling qualitative data. *American Sociological Review*, 1944, 9: 139-150
- Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 1969
- 杜文久. IRT多级评分项目参数估计及其在测验中的应用. *应用概率统计*, 2000, 16: 350-365
- 郭庆科,周晶. Likert量表分析中不同IRT模型的有效性. *心理学探新*, 2004, 3: 67-70
- Sijtsma K, Meijer RR. The person response function as a tool in person-fit research. *Psychometrika*, 2001, 66: 191-208
- 丁强,卢家楣,陈宁. 青少年责任感问卷的编制. *中国临床心理学杂志*, 2014, 22(5): 831-834, 830
- 伍新春,刘畅,胡艳蕊,等. 父亲教投入问卷的编制及其信效度检验. *中国临床心理学杂志*, 2015, 23(4): 576-579

(收稿日期:2015-10-14)