

基于展开反应机制的人格量表编制： 以大学生性格倾向量表为例

邓远平¹, 戴海琦², 罗晓¹

(1.集美大学教师教育学院, 厦门 361021; 2.江西师范大学心理学院, 南昌 330022)

【摘要】 目的:基于拓广等级展开模型(简称GGUM)编制大学生性格倾向量表(简称CTS-GGUM),并与基于因素分析方法构建的量表(简称CTS-FA)进行比较,探索基于展开反应机制编制人格量表的可行性。**方法:**根据三所大学1209名有效被试的作答数据,分别基于GGUM与因素分析方法进行量表编制并比较。**结果:**CTS-GGUM保留了更多的中段项目,内容效度为0.718,效标关联效度为0.699,同质性信度为0.740,其中同质信度与效标关联效度略低于CTS-FA。**结论:**基于展开反应机制编制人格量表具有可行性,它有更好的被试区分能力,并能提高特质中段被试的测量精度。

【关键词】 展开反应机制; 拓广等级展开反应模型; 因素分析; 性格倾向

中图分类号: R395.1

DOI: 10.16128/j.cnki.1005-3611.2017.03.013

Development of Personality Scale Based on Unfolding Response Mechanism: Taking Character Tendency Scale for Undergraduates as An Example

DENG Yuan-ping¹, DAI Hai-qi², LUO Xiao¹

¹College of Teacher education, Jimei University, Xiamen 361021, China;

²College of Psychology, Jiangxi Normal University, Nanchang 330022, China

【Abstract】 Objective: Developing character tendency scale for undergraduates based on GGUM(CTS-GGUM) and comparing it with scale based on factor analysis method(CTS-FA) in order to explore the feasibility of developing personality scale based on unfolding response mechanism. **Methods:** Developing CTS-FA and CTS-GGUM based on 1209 valid subjects and comparing their performance metrics. **Results:** CTS-GGUM retained items located in the middle range of trait, its coefficient was 0.718 for content validity, 0.699 for criterion related validity and 0.740 for homogeneity reliability, the latter two were slightly lower than CTS-FA. **Conclusion:** Developing personality scale based on unfolding response mechanism is feasible, it has a better distinguish ability and improves measurement accuracy of subjects in the middle section of the trait.

【Key words】 Unfolding response mechanism; GGUM; Factor analysis; Character tendency

在心理测量中,个体对一组刺激的反应有两种机制:累积反应机制(Cumulative Response Mechanism)和展开反应机制(Unfolding Response Mechanism)^[1]。累积反应机制也称优势过程(Dominance Process)^[2],其假设是:项目与被试在同一量表连续体上,被试在项目上的得分会随着其特质水平的提高而增加。累积反应机制常用于认知能力测验或成就测验中,特质水平越高的被试在一个项目上的表现会更好。展开反应机制最早的思想来源于瑟斯顿的态度测量理念,当要对测量态度的项目作出反应时(选项从“完全不同意”到“完全同意”按程度分成若干类),被试往往倾向于同意“在问题的关键时刻,更能代表自己态度”的某一个选项,而不会同意与自

己态度相差甚远的选项。为更好说明展开反应机制,举一例:“对于有些国家的鞭刑,我不清楚它是否有存在的必要”,选项分“反对”和“同意”。显然,此项目表明了对鞭刑一种中立的态度,当被试持中立态度时,其选择“同意”选项的概率最高,而当被试位置逐渐高于或低于项目位置时(即被试对鞭刑态度逐渐倾向同意或反对),其选择“同意”选项的概率均会降低,选择“反对”选项的概率上升。两种项目反应机制显然不同,在展开反应机制中,同一反应可能来自不同潜在原因,即被试可能高于项目位置,也可能低于项目位置;而累积反应机制是一种单向控制过程。

许多研究发现个体对人格项目的作答更符合展开反应机制^[3-9],这些研究是在项目反应理论范畴内,从项目的拟合性的角度对展开反应模型和累积反应模型进行比较,发现前者要优于后者,但目前的

【基金项目】 集美大学博士启动基金项目:基于项目反应理论(IRT)展开模型的人格量表编制研究(Q201408);福建省教育厅社科类项目:人格量表编制的三种方法比较(JAS14166)

人格量表编制方法仍以因素分析(Factor Analysis, FA)为主,基于展开反应机制的编制方法相对而言有何优劣,需要在心理测量学各个指标上进行系统的比较。本研究分别用项目反应理论中的拓广等级展开模型(Generalized Graded Unfolding Model, GGUM)^[10]和因素分析两种方法编制大学生性格倾向量表(分别简称CTS-GGUM, CTS-FA),通过实证研究探索展开模型在人格量表编制中的可行性及其优劣,从而丰富人格量表编制的理论及方法。

1 研究方法

1.1 项目编写及评定

根据荣格的内外向理论,参考现有的内外向量表,结合大学生的学习、生活情况,自编项目。编写项目时尽量做到:语言精练不拖沓;语意明了无歧义;项目内容尽量避免社会赞许性;从内容角度考虑,尽量让项目分布在内外向特质区间;项目内容为大学生所熟悉。初稿共有项目80道,均为六级计分,选项从“完全不赞同”到“完全赞同”。

初稿形成后,邀请三位具有博士学位和高级职称的心理学工作者(研究方向分别为人格心理学、人才测评、心理测量)对题目内容代表的内外向程度进行评定:1-极内向、2-较内向、3-中间、4-较外向、5-极外向,若题目内容未测量到内外向这一特质,请专家用0标识,若有2个及以上专家同时判定为0,则此题删除。

1.2 问卷施测

将初稿印刷成纸笔问卷,在三所本科院校随机抽取1317人,使用统一的指导语进行团体施测,问卷作答完成后现场收回。数据经整理,剔除无效问卷108人,有效被试1209人,其中文史类431,理工类778人。

1.3 工具

使用SPSS21软件进行数据整理及描述统计,使用GGUM2004软件进行参数估计、拟合性计算等。

2 结果

2.1 数据单维性检验

采用SPSS21对大学生性格倾向量表初稿的作答数据进行因素分析,Bartlett检验值为33741.58($P<0.000$),说明变量之间有共享因素的可能性;KMO值为0.938,表明数据样本适合于因素分析;得到特征值大于1的因子5个,因子特征值及解释的方差见表1。

表1表明,第一特征值(15.097)是第二特征值(4.643)的3.25倍,符合汉密尔顿大于3的标准^[11],有理由认为这批数据只有一个因素占统治地位,问卷符合单维性的假设。

表1 量表项目数据因素分析结果

成份	合计	方差的 %	累积百分比
1	15.097	18.871	18.871
2	4.643	5.804	24.675
3	3.446	4.307	28.982
4	2.605	3.256	32.238
5	2.385	2.981	35.219

2.2 项目参数及分析

将纸笔问卷收集到的1209份数据采用SPSS21软件进行初步整理,然后用GGUM2004估计项目参数及拟合指标,软件设置迭代次数为外循环200次、内循环50次、迭代精度为0.001,一次性估计出项目参数、被试参数及项目拟合指标。Infit和Outfit统计量是GGUM2004系统中的常用的模型-数据拟合统计量,软件推荐Infit和Outfit的临界值为1.3,相应的卡方值用于项目拟合检验时为0.01^[12]。将拟合不良的项目删除,剩余题目的作答数据再用GGUM软件重新估计,直至所有题目符合项目保留要求,最后保留42道题目,共删除38题,保留项目的具体拟合指标见表2。

为了阐述项目反应机制,选取位置参数在中间的一道题(项目6)进行分析,项目内容为:即使我口才好,我也不是很喜欢参加辩论赛。

大多外向者喜欢在公共场合展现自己的特长,而内向者可能由于缺乏自信相对而言不太外露。越外向的人对题目6(即使我口才好,我也不是很喜欢参加辩论赛)越倾向于选择“不赞同”(选项1、2、3),因为口才好,他(她)对“参加辩论赛”的态度不是“不是很喜欢”,而是“很喜欢”的态度;越内向的人对题目6也倾向于选择“不赞同”(选项1、2、3),因为越内向越缺乏自信、越敏感、怕被关注,对“辩论赛”这类活动就可能越排斥,哪怕口才好,也怕表现不佳受到不良评价,对“参加辩论赛”的态度是“很不喜欢”,而不是题目表述的“不是很喜欢”,倾向选择“不赞同”,而对于内外向特质中间左右的被试对这题则态度倾向于“赞同”(选项4、5、6)。

2.3 CTS-GGUM信度

42题的量表信息量曲线见图2,在整个特质区间内,量表信息曲线不均匀,内向端对应的信息量总体要高于外向端,原因在于位置参数小于0的题目

占了整个题库的69.05%。IRT中的量表信息量代表了测量误差,一般认为量表信息量达到25或以上时(对应标准误为0.2及其以下),被试特质能得到准确

的估计,图2中可以看出本量表对特质水平位于 $[-1.8, 0]$ 的被试测量精度较好。

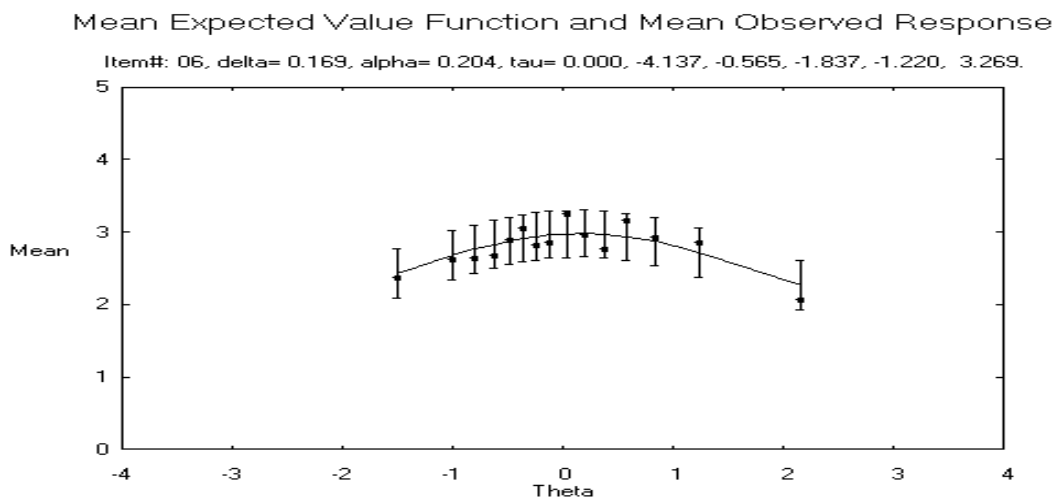
表2 项目参数及拟合指标

ID	α	δ	tau1	tau2	tau3	tau4	tau5	Infit	Outfit
1	0.309	1.822	-1.748	-0.529	-0.95	0.715	1.926	1	0.99
2	0.722	2.745	-4.959	-3.441	-3.273	-2.694	-1.628	0.97	0.99
3	0.543	2.771	-6.775	-4.3	-4.7	-4.95	-3.411	1.02	1
4	1.297	2.185	-3.865	-3.891	-3.002	-2.938	-1.984	0.94	0.95
5	0.089	-3.351	-6.906	-5.237	-3.221	-4.341	-0.448	1	1
6	0.204	0.169	-4.137	-0.565	-1.837	-1.22	3.269	1	1
7	0.655	-2.917	-4.534	-3.042	-3.356	-2.762	-0.896	0.96	1.03
8	1.009	-3.172	-3.358	-2.782	-2.691	-1.672	-1.328	0.96	0.94
9	0.609	-2.538	-3.071	-1.706	-1.013	-1.202	0.143	1.01	1.02
10	0.226	-1.697	-2.24	0.671	2.183	-0.99	2.531	0.99	1.04
11	0.769	2.606	-5.096	-4.569	-4.043	-3.977	-2.448	1.01	1.07
12	0.423	2.014	-5.205	-5.066	-4.403	-4.173	-2.087	1.01	1.08
13	0.621	-3.016	-5.421	-3.325	-3.261	-2.901	-0.929	0.97	0.98
14	0.329	-3.232	-6.724	-3.818	-6.164	-3.505	-0.384	0.99	0.96
15	0.656	2.054	-4.989	-3.078	-1.876	-1.209	0.663	0.99	1
16	0.724	-1.803	-1.506	-1.101	-1.922	0.021	0.601	0.98	0.97
17	1.36	-3.189	-3.386	-2.719	-2.604	-1.839	-0.669	0.92	0.96
18	1.086	2.034	-4.015	-3.109	-2.595	-1.886	-0.631	0.95	0.96
19	0.183	-0.5	-3.867	1.235	0.042	1.868	5.55	1	1.01
20	1.475	-3.218	-3.315	-2.594	-2.222	-1.612	-1.086	0.95	1.01
21	0.401	-2.109	-3.649	-1.347	-2.953	-0.166	1.602	0.99	1
22	0.86	-2.599	-4.029	-2.381	-3.022	-1.665	-0.163	0.96	0.98
23	0.63	-2.399	-4.134	-2.361	-4.079	-1.01	0.161	0.97	0.97
24	0.479	-2.509	-4.875	-2.443	-4.205	-0.778	1.437	0.99	0.98
25	0.75	-2.368	-2.784	-1.167	-1.282	0.132	1.536	1.02	1
26	0.544	-2.223	-3.055	-1.054	-1.861	0.372	1.485	1	1.01
27	0.472	-2.364	-2.786	-1.181	-2.795	-0.43	1.17	0.99	1.02
28	0.792	2.09	-4.657	-3.427	-3.005	-2.251	-0.704	0.99	1.03
29	0.205	-1.847	-4.408	-1.303	-4.176	-1.019	3.23	1	0.97
30	0.157	-1.386	-4.542	-3.08	-4.407	0.269	6.669	1	0.98
31	0.713	-3.368	-3.887	-2.647	-2.942	-2.104	-0.938	0.98	0.98
32	0.143	-1.115	-3.667	2.58	0.246	0.246	4.269	1	1.01
33	0.578	-2.23	-4.046	-1.909	-1.132	-0.202	1.544	0.98	1.02
34	0.379	1.681	-4.75	-2.891	-1	-2.339	-0.633	0.99	1.01
35	0.569	-2.282	-4.14	-1.977	-3.821	-1.641	0.372	0.98	0.99
36	0.535	-3.452	-3.563	-2.759	-3.061	-0.894	0.084	1	1
37	1.643	-2.756	-3.08	-2.24	-2.206	-1.261	-0.715	0.91	0.89
38	0.889	-3.615	-3.737	-2.63	-2.782	-1.744	-0.911	0.99	1.09
39	0.596	2.01	-4.828	-4.08	-2.972	-2.408	-0.683	0.99	1.04
40	0.773	-1.999	-3.476	-1.791	-2.19	0.018	0.349	0.98	1
41	0.11	-1.047	-3.381	1.888	-3.662	2.069	4.943	1	1
42	0.741	1.976	-5.081	-3.955	-2.941	-2.048	-0.8	0.99	1

2.4 CTS-GGUM 内容效度

采用专家判定法,对项目内容所代表的内外向等级作出的评定来证明量表在测量人格内外向的有效性,将结果与项目本身的位置参数作相关分析,结

果表明项目的位置倾向(内外向)与内容能直观联系起来,专家判定的内外向等级与位置参数的基本一致,专家对项目的评定结果平均值与项目位置的斯皮尔曼等级相关系数0.718($P < 0.01$)。



注:横轴为内外向特质,越往右端,表示被试越外向。

图1 第6题的期望得分与观察得分示意图

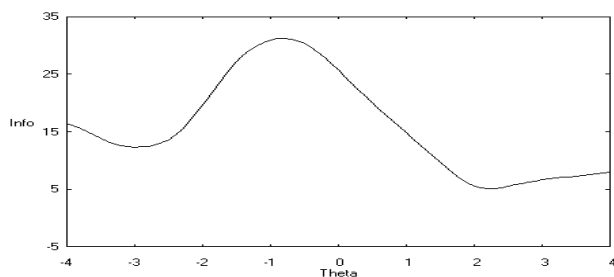


图2 CTS-GGUM的量表信息量

2.5 CTS-GGUM与CTS-FA比较

2.5.1 项目位置比较 采用因素分析来编制大学生性格倾向量表(CTS-FA),将初始问卷项目数据进行因素分析,满足单维性的要求(见表1),项目保留的规则为:①项目在第一因子的载荷最高,且要大于0.5;②项目在众因子上最高的两个载荷绝对值之差大于0.1;最终保留26个项目,这些项目也符合展开反应模型的检验。两种方法下保留项目位置(专家判定的平均等级)如图3所示。

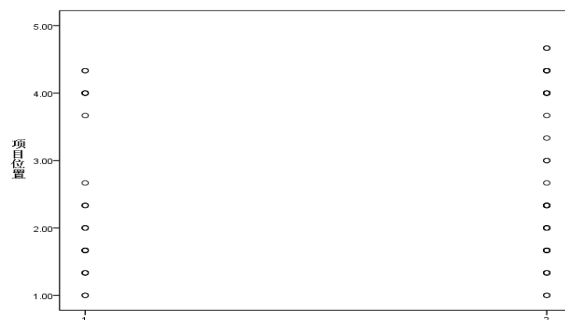
图3表明两种方法编制的量表有一个最大的不同在于因素分析法将中间等级项目基本排除在外,以2.5-3.5作为中间等级间距,因素分析法下仅存1题,其专家评定等级位置为2.667,而GGUM模型中有5题。

2.5.2 量表区分能力比较 CTS-FA量表分数可用总分或均分来表示被试特质水平,基于展开模型构建的量表采用GGUM软件估计被试的特质水平。两种方法下的被试量表分数分布见图4。

图4表明,CTS-FA对被试的区分能力不如CTS-GGUM,在保留三位小数的情况下,CTS-FA量

表分数在大多特质水平点上不止一人,其中在特质为3.038上达到32人,而在CTS-GGUM保留三位小数情况下,被试特质水平未出现相同者。

2.5.3 信度比较 基于因素分析和基于展开模型构建属于不同的测量理论范畴,前者属于经典测验理论,后者属于项目反应理论,两种理论下的信度概念的内涵不一样,直接进行跨理论比较并不合适,故将两者放在同一理论范畴下作比较。在经典测验理论下,因素分析是基于项目间的相关系数,位置位于特质维度中间的项目往往会被删除,分布于两端的项目保留,故其内在一致性系数要高,CTS-FA的Cronbach α 系数为0.895,CTS-GGUM有较多的中等项目,项目间的内在一致性要更低, α 系数为0.740。CTS-FA的题目均包含在CTS-GGUM中,后者多了16道题,项目反应理论的信度观指的是测验信息量,而测验信息量指的是各项目信息量的累加,可知在各特质点上,CTS-GGUM提供的信息量会高于CTS-FA,尤其是在特质水平中段。



注:1为基于因素分析,2为基于展开模型,圆圈粗体表示有多个项目

图3 两种量表构建方法下的项目位置比较

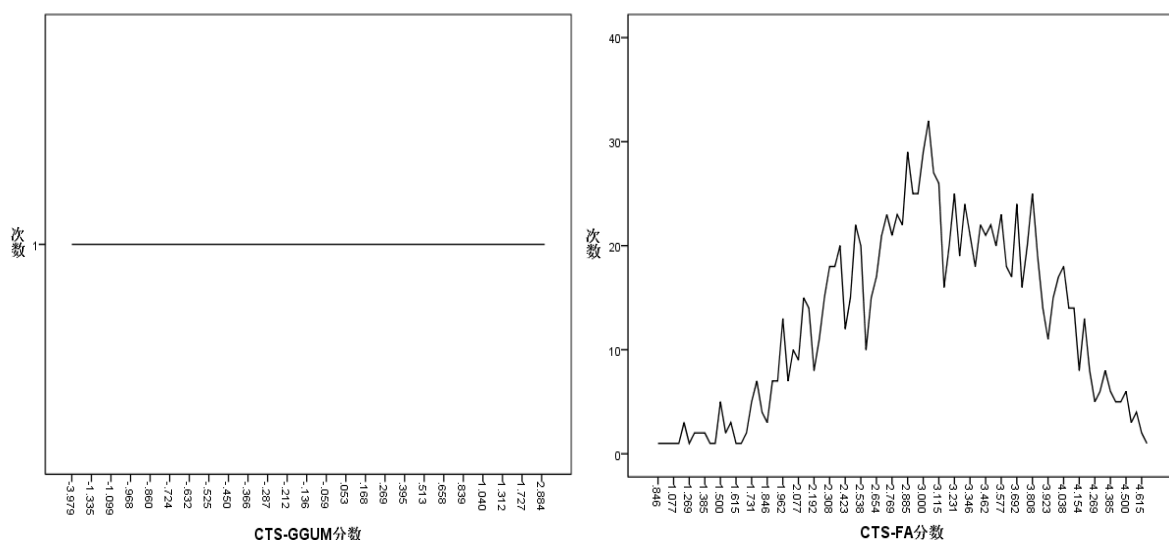


图4 CTS-GGUM与CTS-FA量表分数分布

2.5.4 效标关联效度比较 效标采用EPQ的内外向量表,有效作答757人,CTS-FA与效标的相关系数为0.747($P=0.000$),CTS-GGUM效标相关系数为0.699($P=0.000$),均具有统计学意义。

3 讨 论

本研究基于展开反应机制编制的CTS-GGUM,共有42道拟合良好的项目,项目内容与项目位置参数有着明显的对应关系,内向特质的项目位置参数为负,外向特质的项目位置参数为正,中等项目的位置参数在0附近;CTS-GGUM保留了更多中等位置的项目,从而让中段的被试能找到合适的项目施测;测量学指标方面,CTS-GGUM的内容效度达到了0.718,项目位置参数与专家判定大体一致,效标关联效度为0.699,属中等水平;信度方面,同质性信度为0.740,信息量是项目反应理论的信度指标,CTS-GGUM的测验信息量在整个维度区间内大小不一,[-1.8,0]间的被试测验信息量在25以上,测量精度较高。

因素分析是目前的人格量表编制的主要方法,本研究基于初始问卷80个项目的作答数据编制了CTS-FA,并与CTS-GGUM进行比较。前者的项目数量仅为后者的61.9%,差异最大的地方在维度中段,CTS-FA仅存一题,后者有5题,因为通过因素分析方法构建的量表,其项目都与因子分具有较强的相关(即因子负荷高),中段项目仅有中等特质的被试持赞成态度,而维度两端的被试则持反对意见,这样便无法保证项目得分与被试特质之间有较高的线性相关,而内外向倾向性明显的项目与被试特质往

往有较强的线性相关而得以保留,从计分方向的角度也可理解,一道中等位置项目的计分方向往往不是很明确,不便进行计分,如“我觉得鞭刑很残忍,但它对罪犯有震慑作用”,如果要区分对鞭刑支持还是反对,此题记分方向并不很明确,用因素分析方法构建时,这类项目往往会被删除。当然CTS-GGUM属于项目反应理论范畴,不需考虑项目的计分方向。

在效标关联效度方面,选择EPQ的内外向分量表作为效标,CTS-FA要略高CTS-GGUM,前者为0.747,后者为0.699,原因在于前者与效标的编制方法一样,均为因素分析,故两个量表都较缺乏中间位置项目,而CTS-GGUM保留了相当数量的中等项目,这对被试的结果的排序造成了影响,导致CTS-GGUM的效标分数要略低;信度方面,CTS-FA的同质信度为0.896,远高于CTS-GGUM的0.740,其原因就在于前者基本舍弃了中间项目,被试在维度两端的项目作答要较一致,项目间的相关系数要更高,从项目反应理论的角度理解,CTS-FA提供的信息量则不如CTS-GGUM;在量表区分能力上,CTS-FA对被试的量化只考虑到被试的项目选项,通过项目选项来赋值,对被试的水平区分能力要比CTS-GGUM更低,后者对被试的量化要复杂得多,每个项目有位置参数、区分度和阈值参数,再结合被试的项目选项作答,采用项目反应模型来计分。

基于展开反应机制编制的量表有以下几个特点:①能满足测量学指标的要求;②项目位置参数与内容能有机联系,无须反向计分;③对被试的区分能力更强;④保留了中等位置的项目,有利于减少特质中段被试的测量误差,并为自适应测验提供更好的

题库;⑤相对于传统的因素分析方法,基于展开反应机制的量表开发要更复杂且要求更高,如要有足够的被试量来保证项目参数估计准确性、要有合适的展开模型软件作支持。

值得注意的是,本研究选择的效标是EPQ的内外向量表,它采用因素分析方法构建,CTS-FA与效标构建方法一样,而CTS-GGUM是基于展开反应机制,这种比较有失公平性,更合理的效标可采用经验效标法编制的,是经过实践检验的,但目前未有相关的内外向量表。

另外,本研究对数据的维度检验借用了能力测验单维性检验方法,即因素分析结果中的前两个特征值之比,但被试对能力项目的反应属于优势反应过程,而不属展开反应过程,这一标准在展开反应项目数据中的适用性有待进一步的研究,国内外有研究认为因素分析在单维展开反应数据中会得出两个维度^[13,14],但其研究未汇报前两个特征值之比,因素分析法在检验展开反应数据单维性的合理性有待通过模拟研究去证实。

参 考 文 献

- Thurstone LL. The Measurement of Values. Chicago, IL: University of Chicago Press, 1959
- Chernyshenko OS, Stark S, Chan K, et al. Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research*, 2001, 36: 523-562
- Stark S, Chernyshenko OS, Drasgow F, Williams BA. Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology*, 2006, 91(1): 25-39
- Chernyshenko OS, Stark S, Drasgow F, et al. Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment*, 2007, 19(1): 88
- Mangos PM, Johnston JH. Applying unfolding item response theory to enhance measurement of cultural norms. Papers presented at the RTO human factors and medicine panel (HFM) symposium held in Copenhagen, Denmark, 2008. 32-35
- 邓稳根,戴海琦,黄海,戴慧群. 展开模型对EPQ反应数据的拟合调查. *心理与行为研究*, 2009, 4: 300-303, 311
- 郭庆科,陈英敏,孟庆茂. 自陈量表式测验应用IRT的可行性. *心理学报*, 2005, 2: 275-279
- 方平,邓希冯,姜媛. 展开模型和优势模型在职业兴趣测验中的比较研究. *心理学探新*, 2012, 5: 447-453, 460
- 范晓玲,吴志勇,韩喆. IRT展开模型在“大学生学习自主性量表”中的应用. *中国临床心理学杂志*, 2009, 17(3): 333-336
- Roberts J, Donoghue J, Laughlin J. A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 2000, 24: 27-32
- 漆书青,戴海琦. 项目反应理论及其应用研究. 南昌: 江西高校出版社, 1992. 56
- Roberts JS, Shim HS. GGUM2004 technical reference manual. Version, 2004
- Van Schuur, Henk AL Kiers. Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead. *Applied Psychological Measurement*, 1994, 18(2): 97-110
- 邓稳根,戴海琦. 不同IRT模型在生活取向测验修订版中的适用性比较. *江西师范大学学报(自然科学版)*, 2009, 2: 216-220
- Vondracek FW. Vocational identity development in adolescence. In RK Silbereisen, E Todt(Eds.), *Adolescence in context: The interplay of family, school, peers, and work in adjustment*. New York: Springer-Verlag, 1994
- 陈京军,许清华. 大学生自尊与生涯成熟:自我同一性的中介作用. *中国临床心理学杂志*, 2010, 18(3): 372-376
- 司欣芳,刘姝丽,凌辉,张建人,李新利. 职业中专生自我同一性与自我概念的相关研究. *中国临床心理学杂志*, 2012, 20(3): 387-389
- 戴晓阳. 常用心理评估量表手册. 北京: 人民军医出版社, 2015
- Judge TA, Erez A, Bono JE, Thoresen CJ. The core self-evaluations scale: Development of a measure. *Personnel Psychology*, 2003, 56(2): 303-331
- Zung WKK. Rating instrument for anxiety disorders. *Psychosomatics*, 1971, 12: 371-379

(收稿日期:2016-11-08)

(上接第466页)

(收稿日期:2016-09-07)