

文章编号: 1000-5862(2021)01-0032-06

# 基于主成分因素分析法的 展开型数据单维性检验研究

邓远平<sup>1</sup> 戴海琦<sup>2</sup> 稂永明<sup>3</sup>

(1. 集美大学师范学院 福建 厦门 361021; 2. 江西师范大学心理学院 江西 南昌 330022; 3. 井冈山大学网络信息中心 江西 吉安 343009)

**摘要:** 以拓广的等级展开模型(GGUM)模拟 6 种情形下的展开型数据,使用主成分因素分析法检验展开型数据的单维性。研究发现:在项目位置参数服从均匀分布且被试特质同质性较强的情形下,前 2 个因子的特征值比( $E_1/E_2$ )小于 3,其余情形的比值均大于 3;当被试样本特质同质性较弱时,特征值大于 1 的因子有 3 个,其余情形的因子有 2 个;当测验项目位置参数服从两端分布时,项目 2 因素负荷图呈“括号”形;当测验项目位置参数服从均匀分布时,项目 2 因素负荷图呈“马蹄”形。研究结果表明:能力反应数据单维性的检验标准并不适用于展开型数据,当从因子特征值比、项目负荷图判定展开型数据的单维性时,应考虑项目位置参数分布特点和被试样本特质的同质性程度。

**关键词:** 主成分因素分析法; 展开反应机制; 广义等级展开模型; 单维性

**中图分类号:** B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2021.01.05

## 0 引言

心理测量学将个体对项目的反应分为累积反应机制(Cumulative Response Mechanism)和展开反应机制(Unfolding Response Mechanism)<sup>[1]</sup>,累积反应机制是指在特质连续体上,被试特质水平越高,其在项目上的表现也会越好,认知能力测验和成就测验的项目反应符合这一机制;展开反应机制的理念源自瑟斯顿的态度测量模式,即在被试对态度项目作出反应时(注:项目选项的两极分别是“完全不同意”、“完全同意”,中间选项等距分级),被试一般同意与自己态度相符的选项,而不会选择远离自己态度的选项。在项目反应理论(Item Response Theory)中,这 2 种反应机制均有成熟的数学模型,分别被称为累积模型和展开模型。大量研究发现展开模型比累积模型更能拟合人格反应数据<sup>[2-9]</sup>。

目前,使用比较成熟的展开模型属于单维模型,其使用的前提就是必须保证项目反应数据的单维性。检验展开型数据单维性有 2 种做法:(i) 使用主成分因素分析法。若前 2 个因子的特征值比大于 5 (Reckase 标准)或大于 3 (Hambleton 标准),便可认

为影响测验项目反应的特质有一个居绝对主导的因素<sup>[10]</sup>,大多数研究采用这一做法<sup>[8-9,11-16]</sup>。通过前 2 个因子特征值比大小来判定单维性的做法在能力领域中被证明是可行的,但将这一做法用在反应机制完全不同的展开型数据上是否合适仍有待进一步的研究。(ii) 将文献[17-18]对展开型数据的因素分析结论作为判断单维性的标准,这一标准根据特征值大于 1 的因子数和项目负荷图来衡量。若对项目反应数据进行因素分析,产生了 2 个特征值大于 1 的因子,且这些项目的因子负荷图形似“马蹄”,则认为这批数据满足单维性的要求<sup>[3,7,15,18-20]</sup>。但不应该忽略的是,这一结论是在特定的测验情形下产生的,即项目位置参数符合等距分布、被试样本特质符合正态分布。人格量表大多为李克特量表,其编制方法主要基于项目分析和因素分析,并不能保证项目在所测量的特质上等距分布;从被试的角度来看,自然人群的某种人格特质基本呈正态分布,但现实中的样本往往来自特定的群体,不同群体的人格特质同质性程度可能不同。因此,若基于主成分因素分析来判定展开型数据的单维性,应当要探究项目位置参数分布和被试样本特质同质性程度不同组合下产生的各种可能结果。

收稿日期: 2019-11-05

基金项目: 国家自然科学基金(31660278)、福建省教育厅社科类项目(JAS180204)和李尚大基金(SD201917)资助项目。

作者简介: 邓远平(1979-),男,江西兴国人,副教授,博士,主要从事心理与教育测量研究。E-mail: dyp304@126.com

本文针对以上问题,设计项目位置参数分布和被试样本特质同质性不同水平的组合情形,模拟在各情形下的展开型数据,并比较主成分因素分析结果,从而检验现行的单维展开型数据检验标准是否准确,期望研究结果能为展开型数据的单维性检验提供合理的参考标准。

## 1 方法

本文采用 Montecarlo 方法,使用 R 语言来编写程序,利用 SPSS21 软件进行主成分因素分析。

### 1.1 展开型数据的模型 GGUM

广义等级展开模型(Generalized Graded Unfolding Model, GGUM)<sup>[21]</sup>是目前人格测量中广泛采用的展开模型,它可同时估计2级和多级计分的项目参数。GGUM假定主观反应遵循E. Muraki 拓展的分部评分模型(Generalized Partial Credit Model, GPCM)<sup>[22]</sup>,GGUM使用E. Muraki的模型是鉴于它的一般性。当应用于主观反应时,GPCM表述为

$$P(Y_i = y | \theta_j) = \exp(\alpha_i(y(\theta_j - \delta_i) - \sum_{k=0}^y \tau_{ik})) / \sum_{\omega=0}^M (\exp(\alpha_i(\omega(\theta_j - \delta_i) - \sum_{k=0}^{\omega} \tau_{ik}))), \quad (1)$$

(1)式的限制条件为 $\sum_{k=0}^M \tau_{ik} = 0$ ,  $Y_i$ 是对态度项目*i*的一个主观反应; $y = 0, 1, 2, \dots, M$ ,  $y = 0$ 相当于项目以下反对的最高水平选项; $y = M$ 相当于项目以上反对的最高水平的选项; $\alpha_i$ 是项目*i*的区分度; $\tau_{ik}$ 是相对于项目*i*而言第*k*个主观反应曲线阈限值在特质连续体上的位置; $\tau_{i0}$ 定义为0,它可定义为任何常量,并不影响结果的概率<sup>[22]</sup>。

(1)式定义了一个主观反应水平上的项目反应

模型,然而,模型必须以观察反应来定义。一个观察反应对应2个主观反应,2个主观反应互为排斥,所以,被试对某个观察反应的概率为2个主观反应项概率之和,观察反应的概率函数为

$$P(Z_i = z | \theta_j) = (\exp(\alpha_i(z(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik})) + \exp(\alpha_i((M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}))) / (\sum_{\omega=0}^M (\exp(\alpha_i(\omega(\theta_j - \delta_i) - \sum_{k=0}^{\omega} \tau_{ik}))) + (\exp(\alpha_i((M-\omega)(\theta_j - \delta_i) - \sum_{k=0}^{\omega} \tau_{ik}))))). \quad (2)$$

(2)式的 $\tau_{ik}$ 参数不必连续有序排在潜在特质连续体上,这些值表明连续的主观反应类别概率的交汇点。然而,主观反应概率最大值对应的特质必须是有序排列的,这一特征由运用于主观反应类别的累积模型所保证。

### 1.2 测验项目与被试样本生成

1.2.1 测验项目参数模拟方法 基于项目的位置参数( $\delta$ )分布特点,测验有2种情形: $A_1$ (两端分布,即项目均匀分布在特质连续体的两端): $\delta \sim U(-3, -1) \cup (1, 3)$ 和 $A_2$ (均匀分布,即项目等可能地散布在整个特质连续体上): $\delta \sim U(-3, 3)$ 。项目区分度( $\alpha$ )服从均匀分布,即 $\alpha \sim U(0.5, 2)$ 。所有项目有6个选项等级,项目阈值参数( $\tau_c$ )单独生成, $\tau_c \sim U(-1.4, -0.4)$ ( $c$ 为项目选项等级数减1),阈值参数使用递归式生成:

$$\tau_{k-1} = \tau_k - 0.25 + e_{k-1}, \quad k = 2, 3, \dots, C. \quad (3)$$

(3)式中的 $e_{k-1}$ 为随机误差且 $e_{k-1} \sim N(0, 0.04)$ ,增加这一误差是为了让测验项目阈值参数间隔不会完全相等,更符合现实情境中的作答反应估计结果;这2种情形的测验项目数均设定为20个,项目参数的描述性统计如表1所示。

表1 测验项目参数统计量

测验情形		项目参数							
		$\alpha$	$\delta$	$ \delta $	$\tau_1$	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$
$A_1$ 测验	最小值	0.59	-2.72	1.20	-2.48	-2.23	-1.97	-1.69	-1.37
	最大值	1.97	2.99	2.99	-1.37	-1.09	-0.90	-0.65	-0.44
	平均值	1.21	0.12	2.09	-1.93	-1.69	-1.44	-1.18	-0.94
	标准差	0.40	2.21	0.57	0.32	0.32	0.33	0.32	0.30
$A_2$ 测验	最小值	0.59	-2.81	0.04	-2.43	-2.17	-1.90	-1.64	-1.32
	最大值	1.97	1.96	2.81	-1.29	-1.07	-0.83	-0.64	-0.41
	平均值	1.20	-0.28	1.39	-1.86	-1.62	-1.37	-1.13	-0.88
	标准差	0.42	1.64	0.87	0.37	0.35	0.34	0.34	0.32

1.2.2 被试样本的模拟方法 考虑到现实情境中的被试样本特质的同质性可能不同,本文将被试样

本分为3种: $B_1: \theta \sim N(0, 0.5)$ ,样本特质同质性强; $B_2: \theta \sim N(0, 1)$ ,样本特质同质性中等; $B_3: \theta \sim$

$N(0, 2)$  样本特质同质性弱,被试特质分数限定在  $[-4, 4]$  内.

### 1.3 测验项目作答数据的生成

项目的位置参数分布有 2 个水平,被试样本特质的同质性有 3 个水平,在 2 维度组合下便有 6 种测验情形,具体如表 2 所示.

根据每种测验情形的项目和被试特点,选用 GGUM 模拟各情形下被试的项目反应数据,具体方

法如下: (i) 将项目参数、被试特质水平代入 GGUM,并计算出被试在项目每个选项中的选择概率; (ii) 用 6 个选项的选择概率将  $[0, 1]$  分为 6 个对应区间,其中第 6 个区间设定为闭区间,其余 5 个区间均为左闭右开; (iii) 在区间  $[0, 1]$  上随机生成一个数字,数字必位于 6 个区间之一,所处区间对应的选项为被试在此项目上的作答<sup>[23]</sup>. 为了减少随机误差,每一种测验情形的作答数据重复生成 30 次.

表 2 6 种测验情形

测验情形	$A_1B_1$	$A_1B_2$	$A_1B_3$	$A_2B_1$	$A_2B_2$	$A_2B_3$
被试样本特质的同质性	强	中	弱	强	中	弱
项目位置参数分布	两端分布	两端分布	两端分布	均匀分布	均匀分布	均匀分布

## 2 研究结果

### 2.1 KMO 检验与 Bartlett 球形检验

在主成分因素分析前,应对生成数据的因素分析条件进行检验,一般用 KMO 检验和 Bartlett 球形检验 2 种方法. KMO 检验统计量是用于比较项目间

的简单相关系数和偏相关系数的指标. KMO 值越接近 1 说明项目间的相关性越强,所生成的数据适合因素分析. Bartlett 球形检验是对项目的相关阵是否为单位阵进行检验,若原假设被拒绝,则可认为项目相关阵与单位阵有显著差异,项目数据适合因素分析. 各测验情形下的模拟数据有 30 批次,取 KMO 和 Bartlett 球形检验的平均值、最大值和最小值,结果如表 3 所示.

表 3 KMO 值与 Bartlett 球形检验结果

实验情形	$A_1B_1$	$A_1B_2$	$A_1B_3$	$A_2B_1$	$A_2B_2$	$A_2B_3$
KMO 值	平均值	0.95	0.98	0.98	0.93	0.96
	最大值	0.96	0.98	0.98	0.93	0.96
	最小值	0.95	0.98	0.97	0.92	0.96
Bartlett 球形检验	平均值	4 852.47**	13 564.13**	19 657.39**	3 832.77**	10 909.11**
	最大值	5 143.67**	13 896.54**	20 222.09**	4 053.47**	11 212.38**
	最小值	4 672.08**	13 284.61**	19 448.27**	364.18**	1 052.08**

注:  $A_1, A_2$  分别代表项目位置参数服从均匀分布、两端分布;  $B_1, B_2, B_3$  分别代表被试特质同质性强、中、弱. \* 表示  $0.01 \leq p < 0.05$ , \*\* 表示  $p < 0.01$ ,下同.

表 3 表明,在各情形下的 KMO 值最大为 0.98,最小为 0.92, Bartlett 球形检验卡方值均在 0.01 的显著性水平上拒绝原假设,2 种检验方法的结果均支持数据进行因素分析.

### 2.2 主成分因素分析结果

采用 SPSS21 软件依次对各情形下的 30 批作答数据进行主成分因素分析,保留特征值大于 1 的因子,并计算前 2 个因子的特征值比.

2.2.1 前 2 个因子的特征值比 ( $E_1/E_2$ ) 的差异检验 在各种测验情形下均生成了 30 批次的数据, Levene 检验发现在各测验情形下的  $E_1/E_2$  方差齐性 ( $L_{\text{Levene Statistic}} = 1.532$ ,  $\mu_{f_1} = 5$ ,  $\mu_{f_2} = 174$ ,  $p > 0.05$ ), 以前 2 个因子特征值为因变量、测验情形为分组变量进行方差分析,多重比较采用 LSD 法,分析结果如表 4 所示.

表 4 各测验情形下  $E_1/E_2$  的方差分析结果

实验情形	$A_1B_1$	$A_1B_2$	$A_1B_3$	$A_2B_1$	$A_2B_2$	$A_2B_3$
平均值	3.13	5.37	6.22	2.60	4.08	4.30
标准差	0.04	0.03	0.02	0.04	0.04	0.03
F 值	47 771.38**					
LSD	在各情形间均有显著差异					
偏 $\eta^2$	0.99					

表 4 表明,特征值比在各情形间均有显著差异,在  $A_2B_1$  情形下的特征值比的平均值最小(2.60),在  $A_1B_3$  情形下有最大值(6.22);项目位置参数无论服从两端分布还是服从均匀分布,随被试样本特质的同质性减弱特征值比均会增大. 效应量偏  $\eta^2$  值为 0.99,这表明特征值比的变异基本可由自变量解释.

2.2.2 特征值大于 1 的因子特征值统计 对各测验情形下因子(特征值大于 1)特征值进行简单描述性统计,分析结果如表 5 所示.

表 5 各测验情形下因子特征值的描述性统计

实验情形	第 1 因子			第 2 因子			第 3 因子		
	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值
$A_1B_1$	6.25	6.46	6.13	1.23	1.33	1.13			
$A_1B_2$	10.73	10.91	10.59	1.69	1.79	1.59			
$A_1B_3$	12.44	12.53	12.32	1.57	1.64	1.51	1.18	1.24	1.15
$A_2B_1$	5.20	5.31	5.03	1.66	1.86	1.53			
$A_2B_2$	8.15	8.34	8.00	3.35	3.47	3.24			
$A_2B_3$	8.59	8.72	8.43	4.99	5.07	4.91	1.15	1.20	1.10

从表 5 可看出,在 2 种项目位置分布情形下,随被试样本特质的同质性减弱,第 1 因子特征值会上升;第 2 因子特征值在项目位置参数服从两端分布时略有变化,但在均匀分布下会上升;在  $A_1B_3$  和  $A_2B_3$  这 2 种测验情形下会产生 3 个特征值大于 1 的因子,在其余情形下均产生 2 个特征值大于 1 的因子.

2.3 项目的 2 因子负荷散点图

2.3.1 测验项目位置参数服从 2 种分布下的因子负荷散点图 为了探析项目因素负荷与被试样本特质的同质性、项目位置分布间的关系,将各测验情形下的项目 2 因子负荷图绘出,并在图 1 中标出项目所对应的位置参数秩次,图 1 为测验项目位置参数服从 2 种分布下的因子负荷图.

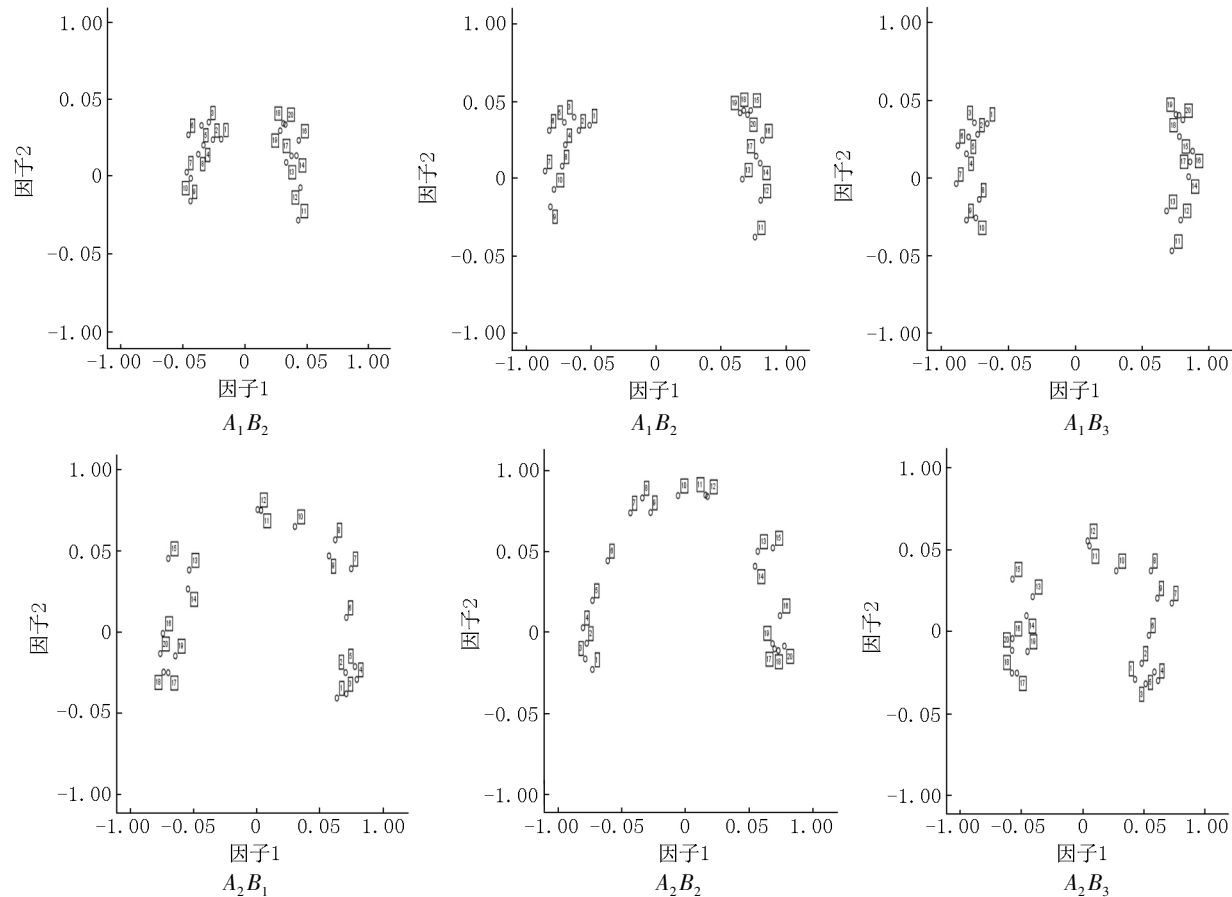


图 1 项目位置参数 2 种分布下的项目因子负荷图

图 1 表明,当项目位置参数服从两端分布情形时,项目的 2 因子负荷分布呈“括号”形,测验项目聚在第 1 因子轴两端,其中位置参数小于 0 的项目在左侧,大于 0 的项目在右侧;位置参数绝对值大的项目聚在“括号”的上端,中等大小的项目聚在“括号”的下端.随着被试样本特质的同质性减弱,第 1

因子负荷的平均绝对值上升,依次为 0.37、0.72、0.79;第 2 因子负荷平均绝对值变化较小,依次为 0.21、0.25、0.25.

当项目位置参数服从均匀分布时,因子负荷图形似“马蹄”,位置参数绝对值大的项目在两尾端,并稍向内卷.结论与 M. L. Davsion<sup>[17]</sup> 和 D. Andrich

等<sup>[18]</sup>一致。在因子负荷量方面,随着被试样本特质的同质性减弱,第1因子负荷平均绝对值先增后略减,依次为0.48、0.60、0.57,而第2因子负荷一直在增加,依次为0.25、0.36、0.39。

2.3.2 测验项目位置参数与因子负荷间的相关性 为了探析项目负荷与项目位置参数2者间的关系,将2个因子负荷转为等级,分别计算2者与位置参数秩次的相关系数,分析结果如表6所示。

表6 位置参数与因子负荷的等级相关

实验情形	第1因子	第2因子
$A_1B_1$	0.56**	0.06
$A_1B_2$	0.60**	0.16
$A_1B_3$	0.75**	0.10
$A_2B_1$	-0.82**	-0.23
$A_2B_2$	-0.91**	0.25
$A_2B_3$	0.96**	-0.01

表6表明,当项目位置参数服从两端分布时,项目的第1因子负荷与其位置参数有中等程度相关,随着被试样本特质的同质性减弱,相关系数增大;当项目位置参数服从均匀分布时,项目第1因子负荷与项目位置参数相关紧密,在被试样本特质同质性强与中这2种情形下,相关系数为负,而在同质性弱的情形下2者相关系数为正。无论项目参数服从何种分布,项目第2因子负荷与位置参数均无显著相关。

### 3 讨论

van Schuur 等<sup>[24]</sup>认为符合展开反应机制的项目作答数据不能采用因素分析方法,因为这类数据违背了该方法的使用条件,本文模拟6种不同测验情形下的展开型数据 KMO 值均大于0.9,且 Bartlett 球形检验的卡方值均达到显著水平,结果表明所模拟的单维展开型数据符合因素分析的条件。

当前展开型数据单维性检验主要借用能力反应数据的做法,即要求因素分析结果的第1因子与第2因子的特征值之比至少要大于3,本文发现这一做法并不通用于各情形下的模拟数据,若项目位置参数服从均匀分布,且被试样本特质的同质性强,则因素分析发现,前2个因子特征值之比要小于3。按照能力测验的单维性标准,会误认为数据并不符合单维性,随着样本同质性减弱,第1因子与第2因子的特征值比在增大;M. L. Davsion<sup>[17]</sup>和 D. Andrich

等<sup>[18]</sup>均认为,对遵循展开反应机制的作答数据进行因素分析,结果将出现2个特征值大于1的因子,项目在这2个因子负荷图中将呈现“马蹄”形,许多研究人员将这一结论用来作为判断展开型数据是否符合单维性的一种依据,但这一结论也不通行。若被试样本特质的同质性弱,则在测验项目位置参数服从2种分布情形下,均会产生3个特征值大于1的因子;而项目的2个因子负荷图呈“马蹄”形也只是在项目的位参数服从均匀分布时才产生。若项目的位参数服从两端分布,则项目的2个因子负荷图呈“括号”形,其中位置参数绝对值大的项目聚在“括号”上端,中间位置的项目聚在“括号”下端,无论项目的因子负荷图还是项目分布点均与 M. L. Davsion<sup>[17]</sup>的研究结果不一样。另外,本研究对项目的位参数与因子负荷间的相关关系进行了分析,结果表明项目位参数与第1因子负荷相关紧密,这个因子可当作是项目所测的核心特质维度,但与第2因子的负荷相关不具显著性, van Schuur 等<sup>[24]</sup>称其为“额外因子现象”(extra-factor phenomenon),原因可能是观察变量的分值与潜在特质之间没有强的线性关系。

当前,人格量表编制大多基于因素分析,保留的都是具有题-总分高相关的项目,低相关项目则被删除。人格项目遵循展开反应机制,处于特质连续体中间位置的项目往往会被删除,因为处于特质两端的被试距离这类项目位置距离较远,所以往往持否定态度,而特质水平接近题目位置的被试则持同意态度,这一作答情形会导致题-总分的低线性相关,最终组成测验的项目位参数分布特征呈两端分布。因此,在将项目因子负荷图作为量表单维性的证据时,若直接套用前人的“马蹄”形的标准,则易导致错误结论。

### 4 结论

能力领域的单维性检验方法并不能直接适用于展开型数据,展开型数据主成分因素分析结果受项目位参数分布和被试特质同质性的影响。在对展开型数据进行单维性检验时,可先对数据进行模型-资料拟合检验,先对拟合良好的项目进行参数估计,然后对作答数据进行主成分因素分析,将因素分析结果、项目位参数和被试样本特质同质性综合起来,若项目位参数服从均匀分布且被试样本特质

的同质性强,则前2个因子特征值之比要小于3,其余情形下比值要大于3;若项目位置参数服从两端分布,则项目的因子负荷图应呈“括号”形,若项目位置参数服从均匀分布,则负荷图呈“马蹄”形。

## 5 参考文献

- [1] Thurstone L L. The measurement of values [M]. Chicago, IL: University of Chicago Press, 1959.
- [2] Wall J E, Baker H E. The interest-finder: evidence of validity [J]. Journal of Career Assessment, 1997, 5(3): 255-273.
- [3] Stark S, Chernyshenko O S, Drasgow F, et al. Examining assumptions about item responding in personality assessment: should ideal point methods be considered for scale development and scoring? [J]. Journal of Applied Psychology, 2006, 96(1): 25-39.
- [4] Mangos P M, Johnston J H. Applying unfolding item response theory to enhance measurement of cultural norms [EB/OL]. [2019-09-13]. <http://ftp.rta.nato.int/Public/PubFullText/RTO/MP/RTO-MP-HFM-142/MP-HFM-142-04.doc>
- [5] Tay L, Drasgow F, Rounds J, et al. Fitting measurement models to vocational interest data: are dominance models ideal? [J]. Journal of Applied Psychology, 2009, 94(5): 1287-1304.
- [6] 郭庆科, 苗金凤, 王昭. IRT 展开模型及对非累积反应机制的检测 [J]. 心理学探新, 2006, 26(1): 66-69, 78.
- [7] 邓稳根, 戴海琦, 黄海, 等. 展开模型对 EPQ 反应数据的拟合调查 [J]. 心理与行为研究, 2009, 7(4): 300-303, 311.
- [8] 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 [J]. 中国临床心理学杂志, 2009, 17(3): 333-336.
- [9] 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 [J]. 心理学探新, 2012, 32(5): 447-453, 460.
- [10] 漆书青, 戴海琦. 项目反应理论及其应用研究 [M]. 南昌: 江西高校出版社, 1992: 44.
- [11] 吴志勇. 基于 IRT 展开模型的高校教师专业素质测评研究 [D]. 长沙: 湖南师范大学, 2009.
- [12] 李雪江. 试题具有理想点反应特征的自尊量表的编制研究 [D]. 南昌: 江西师范大学, 2010.
- [13] 屈型杰. 小学生学习态度测验的初步编制 [D]. 长沙: 湖南师范大学, 2011.
- [14] 高士乘. 基于 IRT 展开模型的中学生学习倦怠量表的编制 [D]. 长沙: 湖南师范大学, 2014.
- [15] 刘国庆, 何壮, 赵守盈. 艾森克人格问卷的展开模型分析 [J]. 贵州师范大学学报: 自然科学版, 2015, 33(6): 36-41.
- [16] Ling Ying, Zhang Minqiang, Kenneth D Locke, et al. Examining the process of responding to circumplex scales of interpersonal values items: should ideal point scoring methods be considered? [J]. Journal of Personality Assessment, 2016, 98(3): 310-318.
- [17] Davison M L. On a metric unidimensional unfolding model for attitudinal and developmental data [J]. Psychometrika, 1977, 42(4): 774-789.
- [18] Andrich D, Styles I M. The structural relationship between attitude and behavior statements from the unfolding perspective [J]. Psychological Methods, 1998, 3(4): 454-469.
- [19] Rost J, Luo Guanzhong. An application of a Rasch-based unfolding model to a questionnaire on adolescent centrism [EB/OL]. [2019-10-16]. [https://www.researchgate.net/publication/268425150\\_Chapter\\_26\\_An\\_Application\\_of\\_a\\_Rasch-Based\\_Unfolding\\_Model\\_to\\_a\\_Questionnaire\\_on\\_Adolescent\\_Centrism](https://www.researchgate.net/publication/268425150_Chapter_26_An_Application_of_a_Rasch-Based_Unfolding_Model_to_a_Questionnaire_on_Adolescent_Centrism).
- [20] Sgammato, Adrienne N. An application of unfolding and cumulative item response theory models for noncognitive scaling: examining the assumptions and applicability of the generalized graded unfolding mode [J]. Dissertation Abstracts International Section A: Humanities and Social Sciences, 2010, 70: 2482.
- [21] Roberts J, Donoghue J, Laughlin J. A general item response theory model for unfolding unidimensional polytomous responses [J]. Applied Psychological Measurement, 2002, 24(1): 27-32.
- [22] Muraki E. A generalized partial credit model: application of an EM algorithm [J]. Applied Psychological Measurement, 1992, 16(1): 159-176.
- [23] Roberts J, Yan Lin, Laughlin J E. Computerized adaptive testing with the generalized graded unfolding model [J]. Applied Psychological Measurement, 2001, 25(2): 177-196.
- [24] van Schuur, Henk A L Kiers. Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead [J]. Applied Psychological Measurement, 1994, 18(2): 97-110.

## The Evaluation Model and Index Weighting of Ability to Propose Mathematical Problems

HONG Qingyu<sup>1</sup>, KANG Chunhua<sup>1,2\*</sup>, ZENG Pingfei<sup>1</sup>, YU Xiangjun<sup>1</sup>

(1. College of Teacher Education, Zhejiang Normal University, Jinhua, Zhejiang 321004, China; 2. Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, Jinhua, Zhejiang 321004, China)

**Abstract:** The measurement and evaluation of the ability to propose mathematical problems has become a hot topic in mathematical teaching research, among which the rationality and scientificity of the evaluation model and index weight have become the primary concern of researchers. On the basis of sorting out the definition of mathematical problems, the existing evaluation contents and methods, the evaluation model proposed by mathematical problems is constructed from the three dimensions of the essential characteristics of the problems, the mathematical characteristics of the problems and the linguistic characteristics of the problems, and two levels of weights are assigned to the indexes in the model. The results show that the indicators of confirmatory factor analysis are good, the evaluation model has good structural validity, and the internal reliability of each dimension is high. The consistency index  $C_I$  and the consistency ratio  $C_R$  calculated based on the maximum characteristic root show that the expert weighting based on the matrix discriminant table has a high consistency, and the weighting is reasonable and scientific. The determination of evaluation model and index weight provides a scientific and reasonable way of thinking on how to measure and evaluate students' ability to propose problems.

**Key words:** ability to propose mathematical problems; evaluation indicators; evaluation tools; weighting

(责任编辑: 冉小晓)

(上接第37页)

## The Study on Test of Unfolding Data's Unidimensionality Based on the Principal Factor Analysis

DENG Yuanping<sup>1</sup>, DAI Haiqi<sup>2</sup>, LANG Yongming<sup>3</sup>

(1. Teachers College, Jimei University, Xiamen, Fujian 361021, China; 2. College of Psychology, Jiangxi Normal University, Nanchang, Jiangxi 330022, China; 3. Network Information Center, Jinggangshan University, Jian, Jiangxi 343009, China)

**Abstract:** The six different types of unfolding data are simulated based on the generalized graded unfolding model (GGUM) and the results of factor analysis (FA) based on using component analysis are investigated. The results show that ratio of the first eigenvalue to the second eigenvalue is less than 3 when items' locative parameters distribute uniformly and the homogeneity of sample is strong, the ratio is more than 3 in other cases. When the heterogeneity of sample is strong, the number of factors whose eigenvalues are more than 1 is 3, the number is 2 in other cases. When items' locative parameters distribute in both ends, the shape of items' loading in two factors is a parenthesis. In another case, the shape is a horseshoe. The result shows that the current testing standards about unidimensionality of unfolding data by the factor analysis (FA) are not comprehensive, the results of FA, including eigenvalue ratio of the first two factors and item loading's figure, shall be related with item's locative parameter and the homogeneity of subjects.

**Key words:** principal factor analysis; unfolding response mechanism; generalized graded unfolding model; unidimensionality

(责任编辑: 冉小晓)