抗作假人格迫选测验中瑟斯顿 IRT 模型的影响因素*

李 辉[1)] 肖 悦[1)] 刘红云[1,2)†]
(1) 北京师范大学心理学部,100875,北京; 2)未来教育高精尖创新中心,100875,北京)

摘要 使用蒙特卡洛(Monte Carlo)模拟研究的方法,探讨应用瑟斯顿 IRT 模型对抗作假迫选测验计分时需满足的编制条件,考察了测验所测特质个数(2 或 5 个)、每维度包含陈述数量(10 或 20 个)、单维配对题目比例(0 或 20%)和正负向陈述间配对题目比例(0、20%或 40%)对模型估计的影响. 结果如下:1) 特质个数对模型估计有显著影响,测验所测特质个数越多,模型估计越准确;2) 陈述数量影响模型估计,测验包含陈述数量越多,模型估计越精确,且当特质个数较少时,陈述数量的影响更大; 3)测验中单维配对题目的比例基本不影响瑟斯顿 IRT 模型的估计精度;4) 测验中加入一定比例(约 20%)的正负向陈述间配对题目可提高模型的估计精度,且特质个数较少时,该因素的影响更大. 最后,在研究结果的基础上给出了开发抗作假人格迫选测验的建议.

关键词 迫选测验; 瑟斯顿 IRT 模型; 作假; 人格测验; 自模式数据
中图分类号 B841.7 **DOI**: 10.16360/j.cnki.jbnuns.2017.05.019

在人才测评领域常使用人格测验进行人才选拔. 但目前绝大多数人格测验多采用多维 Likert 形式,这种测验形式与高利害人才选拔情境的结合,容易导致各种反应偏差(response bias)[1-2],使得被试很容易"作假(faking)",即被试很容易通过伪装或故意歪曲他们的反应而使自己表现出一个有利的自我表征[3]. 人格测验中的作假问题会对人格测验的效度和录用决策产生一定的影响[4-6].

很多研究都表明迫选式人格测验可以有效地应对反应偏差和作假[1,7-11]. 所谓迫选测验(forced-choice format test),代表一种测验编制技术,它迫使被试在 2 个及以上吸引程度相当的答案或陈述间做出选择或对其排序,从而最小化被试作假的可能性[12]. 为了尽可能降低被试作假的可能,在编制迫选式人格测验时,一般会将社会称许性相当的陈述匹配成题组.

迫选测验在人才选拔实践中比较常见,但研究者却对其提出了一些质疑[13-17],主要针对其传统计分方式带来的自模式数据(ipsative data)的问题. 以题组中包含 2 个陈述的迫选测验为例,使用传统计分方式时,被试在某特质上的得分等于其选择测量该特质的陈述的个数;无论被试怎么反应,所有被试在该测验上的总分都是定值. 当一系列反应的总分是一个常数时,这种数据被称为自模式数据. 大量研究表明,自模式数据会带来一系列测量学的问题,包括使分数解释更困难、测验信度估计不准确等[18-19].

为了从根本上解决与自模式数据相关的问题,有测量学家指出应摒弃传统计分方法并转而使用新的 IRT 计分方法. 近年来,测量学家建构了一些针对迫选测验的 IRT 计分模型,Brown[20] 对这些模型进行了总结. 其中,Brown 等[21-23] 的瑟斯顿 IRT 模型(Thurstonian IRT model)与 Stark 等[24] 提出的 MUPP 模型(multi-unidimensional pairwise-preference model)影响较大. 二者关于被试对单个人格陈述如何反应的假设是不同的. 瑟斯顿 IRT 模型假设被试对单个陈述的反应是优势模型(dominance model),即被试的特质水平越高,被试越容易同意陈述. MUPP 模型则假设被试对单个陈述的反应是展开模型 (unfolding model),即被试特质水平距离陈述的位置参数越近,越容易同意该陈述,反之则越容易不同意. 由于展开模型最初提出时针对的是态度、兴趣测验,现有的绝大部分人格测验项目也更适合优势模型,例如有研究发现 16PF 量表中绝大多数项目与优势模型更拟合[25]. 因此,本文关注瑟斯顿 IRT 模型.

Brown 等[21] 通过模拟研究发现测验编制方式会

* 中央高校基本科研业务费专项资金资助
† 通信作者, e-mail: hyliu@bnu.edu.cn
收稿日期:2016-12-09

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

影响瑟斯顿 IRT 模型的估计精度. 为了达到良好估计，应尽量满足：1）测验测量的特质个数较多；2）测量每个特质的陈述数量较多；3）最好有一半题组是正向陈述和负向陈述配对. 但其在模拟中未考虑测验是否可以抗作假，也没有考虑所提出的测验编制建议是否适合于抗作假人格迫选测验的编制. 具体来说，首先，其研究没有考虑陈述的社会称许性，对陈述进行配对时也没有对陈述的社会称许性进行匹配，因而大大降低了测验的抗作假能力. 其次，“最好有一半题组是正向陈述和负向陈述配对”这一测验编制要求是不适合抗作假人格迫选测验的. 因为负向陈述的社会称许性往往比较高，被试若有意作假，一般不会选择负向陈述，因此应尽可能降低测验中该类题组的比例. 以往的研究对正负向陈述间配对比例这一因素的探讨不充分，只比较了无正负向陈述配对和有一半题组为正负向陈述间配对的情况，有必要对正负向陈述间配对比例这一因素进行更充分地探讨，确定能使模型良好估计的最低正负向陈述间配对题组的比例.

另外，Stark 等[24]提出，增加单维配对（即配对的陈述测量同一维度）的题目有助于识别和定位出特质在量尺上的位置，因而有利于迫选测验分数估计精度的提高. 而 Brown 等[21]在其研究中只考虑了多维配对的情况，因此本研究将在配对时加入一定数量的单维配对题目并考察该因素的影响.

综上，本研究通过蒙特卡洛（Monte Carlo）模拟研究的方法，在匹配陈述社会称许性的前提下，探讨特质个数、陈述数量、单维配对比例、正负向陈述间配对比例等几个因素对瑟斯顿 IRT 模型估计的影响，以期为抗作假迫选测验的编制和开发提供一定的参考.

1　瑟斯顿 IRT 模型

1.1　瑟斯顿比较判断法则　模型的提出基于瑟斯顿比较判断法则（Thurston’s law of comparative judgment）[26]. 该法则认为被试在对配对陈述进行比较后会对每个陈述产生一个潜在效应值，被试在比较时选择拥有最大效应值的项目. 以两两配对的迫选测验为例，某题由陈述 i 和陈述 k 配对组成，被试将二者进行比较时对 2 个陈述各自产生一个效应值 t_i 和 t_k，t_i 和 t_k 的相对大小决定被试最后的作答反应. 假设 y_l^* 为 t_i 和 t_k 的差值，那么，被试的作答反应将由 y_l^* 决定，即

$$y_l = \begin{cases} 1, & \text{如果 } y_l^* \geqslant 0, \\ 0, & \text{如果 } y_l^* < 0, \end{cases} \tag{1}$$

式中：y_l 是被试的作答反应，为（0，1）的二分变量；1 表示选择了陈述 i；0 表示选择了陈述 k；y_l^* 是一个连续变量，决定了被试的选择.

Maydeu-Olivares 等[27]指出可以假设效应值 t 为潜在特质分数 η 的线性函数：

$$t = \mu_t + \lambda\eta + e, \tag{2}$$

式中 μ_t 为截距，λ 为因子载荷，η 为潜在特质，e 为特殊因子. 假设陈述 i 和陈述 k 分别测量特质 η_a 和 η_b，则

$$y_l^* = t_i - t_k = \mu_{t_i} - \mu_{t_k} + \lambda_i\eta_a - \lambda_k\eta_b + e_i - e_k \tag{3}$$

1.2　瑟斯顿 IRT 模型　基于瑟斯顿比较判断法则，Brown 等[21]在 2011 年提出了瑟斯顿 IRT 模型. 对两两配对的迫选测验，Brown 为结果变量 y_l 定义了项目特征函数，这是一个针对 0～1 二分数据的标准多维 IRT 模型. 被试选择陈述 i 而非陈述 k 的条件概率为

$$P(y_l = 1 \mid \eta_a, \eta_b) = \Phi\left(\frac{\mu_{t_i} - \mu_{t_k} + \lambda_i\eta_a - \lambda_k\eta_b}{\sqrt{e_i^2 + e_k^2}}\right), \tag{4}$$

也可以表述为

$$P(y_l = 1 \mid \eta_a, \eta_b) = \Phi(\alpha_l + \beta_i\eta_a - \beta_k\eta_b), \tag{5}$$

式中 α 和 β 分别是模型中陈述截距、载荷和特殊因子方差的一个参数转换.

$$\alpha_l = \frac{\mu_{t_i} - \mu_{t_k}}{\sqrt{e_i^2 + e_k^2}}\beta_i = \frac{\lambda_i}{\sqrt{e_i^2 + e_k^2}}\beta_k = \frac{\lambda_k}{\sqrt{e_i^2 + e_k^2}}, \tag{6}$$

式中：$\mu_{t_k} - \mu_{t_j}$ 为模型中的估计参数阈值；λ_i 和 λ_k 为因子载荷；η_a 和 η_b 为被试的潜在特质分数，服从多元正态分布.

进一步得到二元结果变量在相应特质上所提供的信息量函数. 假设一个二元比较结果中的 2 个陈述分别测量特质 a 与 b，则该题目在 a、b 这 2 个特质上所提供的信息量函数可分别用

$$I_l^a(\eta_a, \eta_b) = \frac{[\beta_i - \beta_k \operatorname{corr}(\eta_a, \eta_b)]^2 [\Phi(\alpha_l + \beta_i\eta_a - \beta_k\eta_b)]^2}{P_l(\eta_a, \eta_b)[1 - P_l(\eta_a, \eta_b)]}, \tag{7}$$

$$I_l^b(\eta_a, \eta_b) = \frac{[-\beta_k + \beta_i \operatorname{corr}(\eta_a, \eta_b)]^2 [\Phi(\alpha_l + \beta_i\eta_a - \beta_k\eta_b)]^2}{P_l(\eta_a, \eta_b)[1 - P_l(\eta_a, \eta_b)]} \tag{8}$$

表示. 从式（7）和（8）可知，一个题对在某一特质上的信息量，与题对中 2 个陈述的截距和载荷及特质间的相关都有关. 一定程度上可以从信息量的角度解释诸如正负向配对等因素对模型估计的影响.

2　研究方法

2.1　模拟研究设计　使用蒙特卡洛模拟研究的方法，用最简单常见的陈述两两配对的方式形成题组，称为一个题对. 为了改进 Brown 等[21]研究中未考虑测验抗作假性的不足，首先限定两两配对陈述的社会称许性相近以保证测验的抗作假效果. 在匹配了称许性的

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

前提下,考虑了4个因素的影响:

1)特质个数,与Brown等的研究一致,取2和5这2种水平;

2)陈述数量,与Brown等的研究类似,分别取每维度10和20个陈述2种水平,这2种水平也比较接近实际常用的量表;

3)单维配对比例,指由2个来自同一维度的陈述配对组成的题对数在总题对数中所占的比例,这是以往基于瑟斯顿IRT模型的模拟研究中均未考虑的因素,参考Stark等[24,28]的研究,取0和20%2种水平,测验中剩余题对中的陈述则来自不同的维度;

4)正负向陈述间配对比例,指由一个正向陈述和一个负向陈述配对组成的题对数在总题对数中所占的比例,Brown等[21]研究中仅考虑0和50%2种水平,为了更细致地探讨该因素的影响,本文设定0、20%和40%3个水平,剩余的题对中配对的2个陈述则均为同向.

各因素水平组合共有24种不同的测验编制条件.

2.2 模拟数据的生成和模型估计 首先,模拟生成陈述的参数,包括截距、因子载荷和社会称许性.

1)截距:在区间[-0.8,0.8]上服从均匀分布.

2)因子载荷:服从均匀分布,正向陈述的载荷为正,范围为[0.45,0.9];负向陈述的载荷为负,范围为[-0.9,-0.45],其中载荷正值与负值比例为2∶1.由于被试的社会称许性作答反应倾向与题目对特质的描述程度存在关联,称许性接近的陈述往往载荷值也相对比较接近[29],因此结合实践经验,设定因子载荷和社会称许性之间的相关系数为0.6.

3)社会称许性:均匀分布,范围为[1,5],其值越高,表示该陈述的社会称许性越高.除此之外,由于负向陈述的社会称许性往往高于正向陈述,因此,将正向陈述的社会称许性范围设置为[1.0,3.8],负向陈述的社会称许性范围则为[2.2,5.0].

然后,将陈述两两配对组成题对,为了保证测验的抗作假效果,配对陈述之间的社会称许性之差要小于或等于1.在配对时,还需按研究设计控制单维配对比例(0、20%)和正负向陈述间配对的比例(0、20%、40%).由于特质个数为2或5,陈述数量为每维度10或20个,最终迫选测验的题目(或题对)数量为10、20、25或50.

模拟生成被试样本后,潜在特质分数服从标准正态分布,特质真值方差限定为1,设定特质之间存在较低程度的正相关,相关值为0.2.被试对每个陈述反应的独立因子方差固定为0.5.对24种不同的条件均生成样本量为1 000,每种条件重复1 000次.接着基于潜在特质分数及陈述参数,按照瑟斯顿比较判断法则,模拟生成二分的被试反应数据.最后使用瑟斯顿IRT模型进行估计,参数估计方法为ULSMV(该方法通过更多计算,对未加权最小二乘法(unweighted least squares,ULS)的拟合函数进行了校正);潜在特质水平的估计则采用极大后验估计法(maximum a posteriori,MAP).

3 研究结果

3.1 模型收敛 计算成功收敛的样本次数及比例,收敛比例越高,说明模型应用效果越好.表1中呈现了各条件下模型的收敛比例.特质个数较多或陈述数量较多时,样本成功收敛比例更高.增加一定比例的单维配对题对,对模型收敛起到微弱的正向作用.增加正负向陈述间配对的题对,有助于模型收敛,但比例为20%和40%时基本无差异.

3.2 陈述参数估计精度 计算各条件下各参数(阈值、载荷、特质间相关)估计值误差均方根(root mean square error,RMSE).其值越小表示参数估计值与真值的偏差越小,估计精度越高.表1列出了所有条件下各参数估计的RMSE.

表1 各模拟条件下收敛率及参数估计误差均方根(RMSE)

特质个数	陈述数量	单维配对比例	正负向陈述间配对比例	收敛率	参数估计RMSE		
					阈值	因子载荷	特质间相关
个		%					
2	10	0	0	74.6	0.654	1.115	0.243
2	10	0	20	99.8	0.149	0.224	0.196
2	10	0	40	100	0.071	0.133	0.207
2	10	20	0	93.3	0.589	0.945	0.266
2	10	20	20	98.9	0.158	0.268	0.161
2	10	20	40	100	0.067	0.120	0.140
2	20	0	0	94.4	0.219	0.302	0.161
2	20	0	20	100	0.060	0.094	0.128
2	20	0	40	98.7	0.060	0.189	0.256
2	20	20	0	95.3	0.167	0.376	0.209
2	20	20	20	100	0.057	0.095	0.141
2	20	20	40	100	0.061	0.082	0.081
5	10	0	0	94.8	0.063	0.168	0.232
5	10	0	20	100	0.070	0.103	0.065
5	10	0	40	100	0.066	0.096	0.057
5	10	20	0	93.9	0.064	0.159	0.257
5	10	20	20	100	0.066	0.109	0.070
5	10	20	40	100	0.065	0.111	0.096
5	20	0	0	99.8	0.061	0.103	0.164
5	20	0	20	100	0.063	0.080	0.046
5	20	0	40	100	0.065	0.081	0.044
5	20	20	0	100	0.059	0.094	0.123
5	20	20	20	100	0.058	0.078	0.048
5	20	20	40	100	0.063	0.081	0.043

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

首先，分析各因素对阈值估计精度的影响. 图 1-a、b、c 分别描述了不同条件下阈值估计的 RMSE. 方差分析结果如下：特质个数越多，阈值估计 RMSE 越小（$F(1,9)=17.80$，$p=0.002$，$\eta_p^2=0.66$）；每维度包含陈述数量越多，阈值估计 RMSE 越小（$F(1,9)=8.80$，$p=0.016$，$\eta_p^2=0.49$）；单维配对比例主效应不显著（$F(1,9)=0.12$，$p=0.737$）；正负向陈述间配对比例主效应显著（$F(2,9)=12.23$，$p=0.003$，$\eta_p^2=0.73$），正负向陈述间配对比例为 0 时估计 RMSE 最大，但正负向陈述间配对比例为 20%和 40% 这 2 种条件下估计精度无显著差异（Bofferroni 事后检验）. 另外，特质个数与陈述数量间交互作用显著（$F(1,9)=8.00$，$p=0.020$，$\eta_p^2=0.47$），特质个数较少时，增加陈述数量对阈值估计精度的影响更大（详见图 1-a）；特质个数与正负向陈述间配对比例间交互作用显著（$F(2,9)=12.66$，$p=0.002$，$\eta_p^2=0.74$），特质个数较少时增加正负向配对比例对提升阈值估计精度的效果更明显（图 1-c）.

其次，分析各因素对载荷估计精度的影响. 图 1-d、e、f 分别描述了不同条件下载荷估计的 RMSE. 方差分析结果表明：特质个数越多，载荷估计的 RMSE 越小（$F(1,9)=19.61$，$p=0.002$，$\eta_p^2=0.69$）；陈述数量越多，载荷估计的 RMSE 越小（$F(1,9)=9.81$，$p=0.012$，$\eta_p^2=0.52$）；单维配对比例主效应不显著（$F(1,9)=0.08$，$p=0.785$）；正负向陈述间配对比例主效应显著（$F(1,9)=14.37$，$p=0.002$，$\eta_p^2=0.76$），正负向陈述间配对比例为 0 时估计 RMSE 最大，而 20%和 40% 这 2 种条件下基本无差别；特质个数与陈述数量间交互作用显著（$F(1,9)=5.65$，$p=0.042$，$\eta_p^2=0.39$），特质个数为 2 时比个数为 5 时更易受陈述数量的影响；特质个数与正负向陈述间配对比例二者交互作用显著（$F(2,9)=10.76$，$p=0.004$，$\eta_p^2=0.71$），特质个数为 2 时比个数为 5 时更易受正负向陈述间配对比例的影响.

最后，分析各因素对特质间相关估计 RMSE 的影响. 从图 1-g、h、i 来看，特质间相关估计的 RMSE 是可以接受的. 方差分析结果表明：特质个数较多时，该参数 RMSE 较小（$F(1,9)=23.83$，$p=0.001$，$\eta_p^2=0.73$）；陈述数量较多时，该参数 RMSE 较小（$F(1,9)=7.97$，$p=0.020$，$\eta_p^2=0.47$）；单维配对比例主效应不显著（$F(1,9)=0.72$，$p=0.418$）；正负向陈述间配对比例主效应显著（$F(2,9)=15.77$，$p=0.001$，$\eta_p^2=0.78$），但比例为 20%和 40%时无显著差异. 另外，两两因素间的交互作用均不显著.

3.3　潜在特质估计精度　所有条件下各特质信度范围为 0.19～0.81，75%的在 0.40～0.78 之间（见表 2）. 图 2 呈现了不同特质个数时各因素对所有特质平均真实信度的影响. 方差分析结果表明：特质个数为 5 时，平均测验信度比特质个数为 2 时更高（$F(1,9)=56.27$，$p=0$，$\eta_p^2=0.86$）；陈述数量为每维度 20 个时，测验信度比每维度仅 10 题时更高（$F(19)=110.97$，$p=0$，$\eta_p^2=0.93$），单维配对题目比例对测验信度基本没有影响（$F(1,9)=2.09$，$p=0.182$，$\eta_p^2=0.19$）；正负向陈述间配对比例越高，测验信度越高（$F(2,9)=160.61$，$p=0$，$\eta_p^2=0.97$）；特质个数与正负向陈述配对比例间交互作用显著（$F(2,9)=13.93$，$p=0.002$，

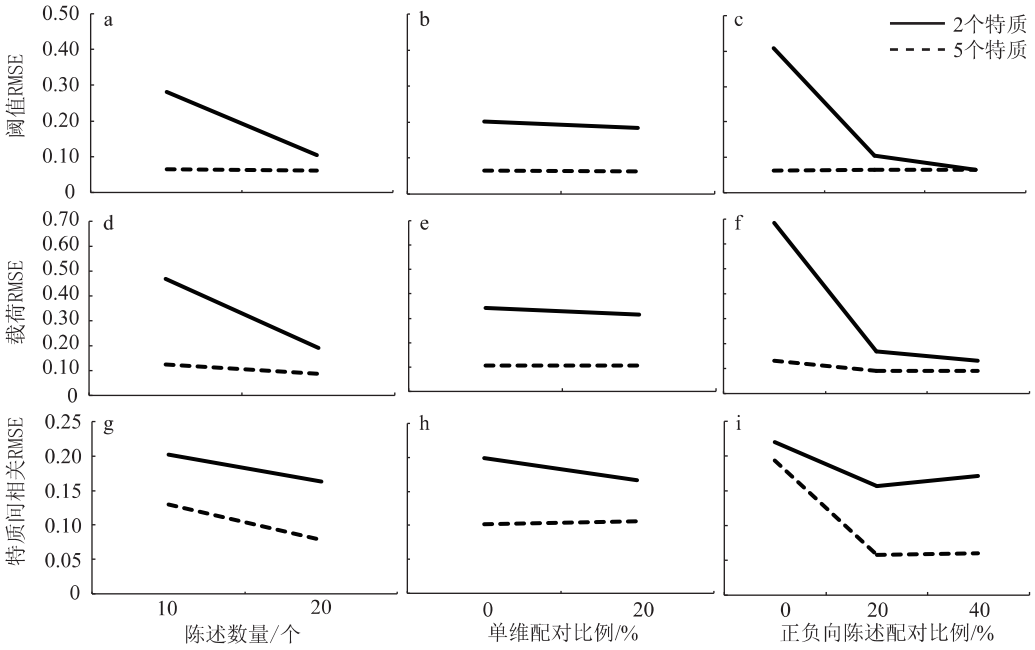

图 1　各因素对参数估计误差均方根的影响

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$\eta_p^2=0.76$),特质个数为2时比个数为5时更易受正负向陈述间配对比例的影响.除此之外,由图2-c可见,当特质个数为5时,正反向陈述间配对比例为40%时的测验信度与该比例仅为20%时基本没有差别,当仅有2个特质时,将正负向陈述间配对比例由20%增加至40%对提升测验信度有一定作用.

表2 各模拟条件下测验真实信度

特质个数	陈述数量	单维配对	正负向陈述间	真实信度					各特质
个		比例/%	配对比例/%	特质1	特质2	特质3	特质4	特质5	信度均值
2	10	0	0	0.35	0.21				0.28
2	10	0	20	0.48	0.51				0.50
2	10	0	40	0.52	0.58				0.55
2	10	20	0	0.31	0.19				0.25
2	10	20	20	0.44	0.53				0.48
2	10	20	40	0.63	0.57				0.60
2	20	0	0	0.38	0.25				0.31
2	20	0	20	0.71	0.67				0.69
2	20	0	40	0.78	0.73				0.75
2	20	20	0	0.34	0.22				0.28
2	20	20	20	0.69	0.64				0.67
2	20	20	40	0.75	0.76				0.75
5	10	0	0	0.47	0.46	0.42	0.42	0.40	0.43
5	10	0	20	0.62	0.56	0.61	0.58	0.60	0.60
5	10	0	40	0.69	0.64	0.63	0.65	0.63	0.65
5	10	20	0	0.38	0.40	0.42	0.43	0.36	0.40
5	10	20	20	0.62	0.52	0.68	0.56	0.51	0.58
5	10	20	40	0.49	0.48	0.54	0.62	0.59	0.54
5	20	0	0	0.62	0.59	0.56	0.50	0.55	0.56
5	20	0	20	0.77	0.79	0.75	0.74	0.77	0.76
5	20	0	40	0.81	0.78	0.79	0.76	0.78	0.79
5	20	20	0	0.62	0.56	0.56	0.52	0.55	0.56
5	20	20	20	0.76	0.78	0.71	0.66	0.76	0.74
5	20	20	40	0.79	0.81	0.80	0.77	0.78	0.79

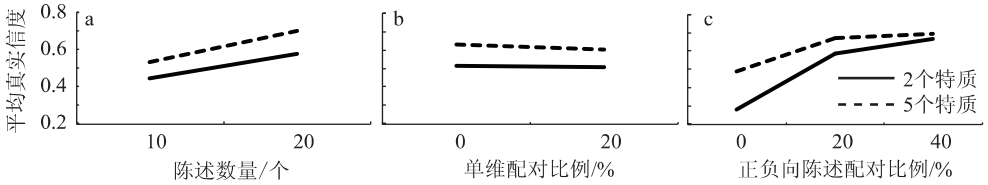

图2 各因素对测验所有特质平均真实信度的影响

4 讨论

为了使得迫选测验具有抗作假能力,对陈述社会称许性进行匹配是关键的步骤,可降低甚至消除被试的社会称许性反应[30].匹配了社会称许性的抗作假迫选测验在实际中应用最为广泛.但对比Brown等[21]的研究结果可以发现,相近条件下本研究中测验信度更低.可见匹配社会称许性会降低测验信度.这主要因为社会称许性相近的题目其载荷往往也相近,因此匹配社会称许性会导致配对陈述之间的载荷差异较小,此时题目提供的信息量较小,降低了模型对潜在特质分数的估计精度[22].由此可见,在开发抗作假迫选测验时,为了确保瑟斯顿IRT模型的精确估计,应特别注意在编制测验阶段考虑各种因素的影响.结果显示,各因素的影响如下.

1)所测特质个数越多,模型估计越准确.以往研究也发现,当特质数量多时自模式数据的自模性质会弱化[31].

2)每维度包含的陈述数量越多,模型估计越精确,测验信度越高.研究还发现,当测验所测特质个数较少时,增加陈述数量对模型估计精度的正向影响比特质个数较多时更大.因此,若测验所测特质个数较少,可以适当增加陈述数量.

3)增加单维配对题目对瑟斯顿IRT模型的估计基本无影响.当前研究发现该因素的影响不能由MUPP模型推广至瑟斯顿IRT模型中.

4)增加正负向陈述间配对题对可以提高模型估计精度,主要是由于正向陈述和负向陈述的配对能够

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

提供更多关于特质水平的信息（由式（7）和（8）亦可知）. 研究还发现正负向陈述间配对比例为 20％和 40％时模型估计精度差异很小，这对于实际测验编制者非常有价值. Brown 建议要在测验中加入 50％正负向陈述间配对题目，主要是由于其对该因素的探讨不充分，实际上出于保证测验抗作假能力的考虑，加入 20％左右正负向陈述间配对题目更合理. 另外，研究发现当所测特质个数较少时，该因素对模型估计的影响比所测特质较多时更大.

5　总结

基于本研究的结论，对编制具备抗作假能力的两两配对迫选测验的实际工作者提出以下建议：

1）所测特质个数及测量各特质的陈述数量宜较多；

2）在配对前应标定各陈述描述特质的方向（正向陈述或负向陈述），评定各陈述的社会称许性，并在配对时将测量不同特质的陈述按社会称许性相近原则进行匹配；

3）配对时需加入约 20％的正负向陈述间配对题对，当测验所测特质较少时可适当增加该比例，但该比例不宜过高.

本研究结论可指导抗作假人格迫选测验的开发. 后续研究可关注 2 个以上陈述配对的测验. 还可以关注其他迫选测验计分模型，探讨影响其他模型的因素并分析是否与瑟斯顿 IRT 模型的影响因素相同. 比较不同模型在各条件下的拟合和估计情况，据此提出不同条件下模型选择的建议.

6　参考文献

[1] CHEUNG M W, CHAN W. Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis [J]. Structural Equation Modeling, 2002, 9(1): 55

[2] MORRISON E W, BIES R J. Impression management in the feedback-seeking process: a literature review and research agenda [J]. The Academy of Management Review, 1991, 16(3): 522

[3] 骆方，张厚粲. 人格测验中作假的控制方法[J]. 心理学探新，2007，27(4)：78

[4] HONKANIEMI L, TOLVANEN A, FELDT T. Applicant reactions and faking in real-life personnel selection[J]. Scandinavian Journal of Psychology, 2011, 52(4): 376

[5] KOMARS, BROWN D, KOMAR J, et al. Faking and the validity of conscientiousness: a Monte Carlo investigation[J]. Journal of Applied Psychology, 2008, 93(1): 140

[6] MUELLER-HANSON R, HEGGESTAD E D, THORNTON G C. Faking and selection: considering the use of personality from select-in and select-out perspectives[J]. Journal of Applied Psychology, 2003, 88(2): 348

[7] BOWEN C C, MARTIN B A, HUNT S T. A comparison of ipsative and normative approaches for ability to control faking in personality questionnaires[J]. International Journal of Organizational Analysis, 2002, 10(3): 240

[8] BARTRAM D. Increasing validity with forced-choice criterion measurement formats[J]. International Journal of Selection and Assessment, 2007, 15(3): 263

[9] CHRISTIANSEN N D, BURNS G N, MONTGOMERY G E. Reconsidering forced-choice item formats for applicant personality assessment [J]. Human Performance, 2005, 18(3): 267

[10] HOGAN R. In defense of personality measurement: new wine for old whiners [J]. Human Performance, 2005, 18(4): 331

[11] JACKSON D N, WROBLEWSKI V R, ASHTON M C. The impact of faking on employment tests: does forced choice offer a solution? [J]. Human Performance, 2000, 13(4): 371

[12] 王珊，骆方，刘红云. 迫选式人格测验的传统计分与 IRT 计分模型[J]. 心理科学进展，2014，22(3)：549

[13] DUNNETTE M D, MCCARTNEY J, CARLSON H C, et al. A study of faking behavior on a forced-choice self-description checklist[J]. Personnel Psychology, 1962, 15(1): 13

[14] GREER T, DUNLAP W P. Analysis of variance with ipsative measures[J]. Psychological Methods, 1997, 2(2): 200

[15] LOO R. Issues in factor analyzing ipsative measures: The learning style inventory (LSI1985) example[J]. Journal of Business and Psychology, 1999, 14(1): 149

[16] MEADE A W. Psychometric problems and issues involved with creating and using ipsative measures for selection[J]. Journal of Occupational and Organizational Psychology, 2004, 77(4): 531

[17] TENOPYR M L. Artifactual reliability of forced-choice scales[J]. Journal of Applied Psychology, 1988, 73(4): 749

[18] DUNLAP W P, CORNWELL J M. Factor analysis of ipsative measures[J]. Multivariate Behavioral Research, 1994, 29(1): 115

[19] HICKS L E. Some properties of ipsative, normative,

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

and forced-choice normative measures[J]. Psychological Bulletin, 1970, 74(3): 167

[20] BROWN A. Item response models for forced-choice questionnaires: a common framework [J]. Psychometrika, 2014, 81(1): 135

[21] BROWN A, MAYDEU-OLIVARES A. Item response modeling of forced-choice questionnaires [J]. Educational and Psychological Measurement, 2011, 71(3): 460

[22] BROWN A, MAYDEUOLIVARES A. Fitting a Thurstonian IRT model to forced-choice data using Mplus [J]. Behavior Research Methods, 2012, 44(4): 1135

[23] BROWN A, MAYDEUOLIVARES A. How IRT can solve problems of ipsative data in forced-choice questionnaires[J]. Psychological Methods, 2013, 18(1): 36

[24] STARK S, CHERNYSHENKO O S, DRASGOW F. An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: the multi-unidimensional pairwise-preference model [J]. Applied Psychological Measurement, 2005, 29(3): 184

[25] 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, 91(1): 25

[26] THURSTONE L L. A law of comparative judgment [J]. Psychological Review, 1927, 34(4): 273

[27] MAYDEU-OLIVARES A, BROWN A. Item response modeling of paired comparison and ranking data [J]. Multivariate Behavioral Research, 2010, 45(6): 935

[28] STARK S, CHERNYSHENKO O S, DRASGOW F, et al. Adaptive testing with multidimensional pairwise preference items improving the efficiency of personality and other noncognitive assessments[J]. Organizational Research Methods, 2012, 15(3): 463

[29] BÄCKSTRÖM M, BJÖRKLUND F, LARSSON M R. Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally [J]. Journal of Research in Personality, 2009, 43(3): 335

[30] MCCLOY R A, HEGGESTAD E D, REEVE C L. A silk purse from the sow's ear: retrieving normative information from multidimensional forced-choice items [J]. Organizational Research Methods, 2005, 8(2): 222

[31] BARON H. Strengths and limitations of ipsative measurement [J]. Journal of Occupational and Organizational Psychology, 1996, 69(1): 49

Influencing factors of Thurstonian IRT model in faking-resisting forced-choice questionnaire

LI Hui[1)] XIAO Yue[1)] LIU Hongyun[1,2)]

(1)Faculty of Psychology, Beijing Normal University, 100875, Beijing, China;
2)Beijing Advanced Innovation Center For Future Education, 100875, Beijing, China)

Abstract A Monte Carlo simulation was used here to compare performance of Thurstonian IRT model across a variety of faking-resisting forced-choice designs. In total, 24 conditions were examined by crossing the following factors: 1) number of trait (2 or 5); 2) items per trait (10 or 20); 3) percentage of pairs composed of items on the same dimension (0 or 20%); 4) percentage of pairs composed of items keyed in opposite directions (0, 20% or 40%). Larger number of traits was found to lead to more accurate estimation of model parameters, and more precise recovery of trait score. Items per trait had a positive impact on application of Thurstonian model, with a larger amount indicating more precise parameter estimation and better true trait score recovery. It had greater influence when there were not enough traits. The percentage of item pairs composed of items on the same dimensions had no significant influence on model estimates. Including a percentage of pairs with items keyed in opposite direction (20%) could greatly improve accuracy of parameter and trait score estimation. The factor impact was greater when the number of traits measured in the questionnaire was small. These results have important implications for how faking-resisting forced-choice questionnaires should be designed in the future.

Keywords forced-choice questionnaire; Thurstonian IRT model; faking; personality test; ipsative data

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net