

# 补偿多维 IRT 模型的 Q 矩阵设计\*

潘世权<sup>1</sup>, 赵守盈<sup>1,2</sup>

(1. 贵州师范大学心理学院, 贵阳 550025; 2. 凯里学院, 凯里 556011)

**摘要:** Q 矩阵不仅可用于表示项目与各维度之间的关系, 还会影响测量模型的可识别性和参数估计精度。本文通过模拟研究, 考察了不同 Q 矩阵设计如何影响补偿多维项目反应理论模型的参数估计精度。研究结果表明: (1) 项目间多维设计仅在项目参数估计上占优势; (2) 在绝大多数条件下, 由模型识别和测量两个维度的项目组成的 Q 矩阵能提供比其他设计更精确的能力参数向量估计值, 同时项目参数估计的误差也在可接受范围内。

**关键词:** 多维项目反应理论; Q 矩阵; 项目内多维; 项目间多维

**中图分类号:** B841.2

**文献标志码:** A

**文章编号:** 1003-5184(2024)03-0273-08

## 1 引言

Q 矩阵的概念源于认知诊断评估 (Cognitive Diagnostic Assessment, CDA), 可用于表示项目与属性之间的关系 (Tatsuoka, 1985), 即测验结构。一般而言, Q 矩阵是一个  $J \times K$  的二分变量矩阵, 矩阵的每一列代表一个属性, 每一行代表一个项目。若 Q 矩阵中的元素  $q_{jk} = 1$ , 表示项目  $j$  考察了属性  $k$ , 否则  $q_{jk} = 0$ 。实际上, Q 矩阵不仅可用于 CDA, 也可以用于多维项目反应理论 (Multidimensional Item Response Theory, MIRT)。在 MIRT 中, Q 矩阵的每一列代表一个维度能力,  $q_{jk} = 1$  表示项目  $j$  测量了维度  $k$ , 否则  $q_{jk} = 0$  (Embretson & Yang, 2013; Wang & Nydick, 2015)。

在 MIRT 中, 一个 Q 矩阵设计的好坏不仅关系到模型是否可识别, 还会影响模型的参数估计精度。若一个模型不可识别, 则会影响参数估计的一致性 (频率统计方法) 或可解释性和收敛性 (贝叶斯统计方法) (Xu & Zhang, 2016)。因此, 在设计一个 Q 矩阵时, 应当首先考虑模型识别问题。

目前, MIRT 的研究者 (Chalmers, 2020; Fu et al., 2021; Man et al., 2019; Sun et al., 2016; Wang & Nydick, 2015; Xu et al., 2022) 采用的模型识别方案通常与 Béguin 和 Glas (2001) 以及 Holman 和 Glas (2005) 的两个方案类似或一致: (1) Q 矩阵至少包含一个  $(K-1) \times K$  的子矩阵  $Q^*$ , 其元素  $q_{jk}^* = 0$ , 其余元素为 1, 其中  $j = 1, \dots, K-1, k = j+1, \dots, K$ , 同时能力参数向量服从均值向量  $\mu$  为  $\mathbf{0}$ , 协方差矩阵  $\Sigma$  为  $K$  阶单位阵  $I_K$  的多元正态分布; (2) Q 矩阵至

少包含一个  $I_K$  (这些项目在所测维度上的区分度固定为 1), 且能力参数向量服从  $\mu$  为  $\mathbf{0}$ ,  $\Sigma$  可自由估计的多元正态分布。尽管上述两种方案可相互转换 (Holman & Glas, 2005), 但  $\Sigma$  可自由估计意味着在能力参数向量估计过程中, 可借用更多的信息, 从而提高其估计精度 (Wang, 2004)。因此, 相较于第一种方案, 第二种方案可能更有优势。

当 Q 矩阵满足模型识别的要求后, 接下来应当考虑其对参数估计精度的影响。无论项目参数还是能力参数向量, 现有研究一致发现: 在保证模型识别的前提下, Q 矩阵包含只测量一个维度的项目越多, 则参数估计的精度越高 (Adams et al., 1997; Babcock, 2011; Cho et al., 2021; Wang, 2015; Wang & Nydick, 2015), 相应的解释是: 对于项目参数, Q 矩阵越复杂则项目参数估计的抽样变异性越大, 进而精度也越差 (Wang & Nydick, 2015); 对于能力参数向量, Q 矩阵越复杂则参数估计算法的偏差会随之上升, 进而精度也越差 (Wang, 2015)。另外, 如前所述, 复杂的 Q 矩阵会降低项目参数估计的精度, 而精度不佳的项目参数会影响能力参数向量估计的精度 (Feuerstahler, 2018; Wang, 2015)。

若根据上述研究的结论, 则会认为项目间多维设计 (每个项目仅测量一个维度) 是最优的 Q 矩阵设计。然而, 编制纯粹只测量一个维度的项目难度较大 (Wang & Nydick, 2015), 这种 Q 矩阵设计很可能会加重测验编制者的负担。不仅如此, 项目间多维设计的 Q 矩阵还面临多维区分度 (Multidimen-

\* 基金项目: 国家教育考试科研规划 2021 年度课题“新高考背景下高中生学业分数报告的研究” (GJK2021017)。

通信作者: 赵守盈, E-mail: zhaoshouying@126.com。

sional Discrimination, MDISC) 不高所带来的负面影响。Reckase 和 McKinley (1991) 以及 Yao 和 Schwarz (2006) 均指出, 项目的 MDISC 越高则能提供的信息量也越高, 这有助于提高能力参数向量估计的精度。

那么, 当 Q 矩阵满足模型识别的要求后, 该如何设计剩余的部分? 为了回答这一问题, 本文将借助模拟研究进一步探讨 MIRT 模型的 Q 矩阵设计问题。需要注意的是, MIRT 模型可分为补偿和非补偿两类, 本文仅关注补偿 MIRT 模型, 这是因为该类模型更常用, 且非补偿 MIRT 模型的参数估计难度更大。

## 2 模拟研究

### 2.1 模拟设计

测验维度分为  $K=3$  和  $K=5$ , 项目数量  $J$  为  $K$  的整数倍, 分为  $J=3K$ 、 $J=4K$ 、 $J=5K$  和  $J=6K$ , 每种项目数量包含可能的  $I_K$  数量分别为: 1 至 3、1 至 4、1 至 5 和 1 至 6 (彭亚风等, 2016)。在上述设定中, 每个 Q 矩阵均至少包含一个  $I_K$ , 以此保证模型识别。另外, 当 Q 矩阵包含可能的  $I_K$  数量为最大值时, Q 矩阵为项目间多维设计, 否则为项目内多维设计。对于 Q 矩阵除  $I_K$  外的部分, 按如下方式设定: 当  $K=3$ , Q 矩阵除  $I_K$  外的部分分成测量两个维度和三个维度两种水平, 分别记为  $K_3^*=2$  和  $K_3^*=3$ 。若 Q 矩阵全由  $I_K$  组成, 则记为  $K_3^*=1$ 。当  $K=5$ , Q 矩阵除  $I_K$  外的部分分成四种水平, 测量两个维度、三个维度、四个维度和五个维度, 分别记为  $K_5^*=2$ 、 $K_5^*=3$ 、 $K_5^*=4$  和  $K_5^*=5$ 。若 Q 矩阵均由  $I_K$  组成, 则记为  $K_5^*=1$ 。

为了让模拟结果可归因于不同的 Q 矩阵设计, 本研究主要参考 Chen 等人 (2017) 的思路生成 Q 矩阵及其项目参数。第一步, 以多维双参数逻辑斯蒂克 (Multidimensional Two-Parameter Logistic, M2PL) 模型 (Reckase, 2009) 为真模型, 为  $K=3$  和  $K=5$  分别生成一个项目数量  $J=6K$ , 且每个项目均测量  $K$  个维度的临时测验, 其区分度参数  $a_{jk}$  从  $U(1, 2)$  中随机抽取, 截距参数  $d_j$  从  $N(0, 1)$  中随机抽取 (Cho et al., 2021)。这里选用 M2PL 模型主要是因为其较为常用且参数估计简单, M2PL 模型可表示为:

$$P(Y_{ij} = 1 | \theta_i) = \frac{\exp(\sum_{k=1}^K a_{jk} q_{jk} \theta_{ik} + d_j)}{1 + \exp(\sum_{k=1}^K a_{jk} q_{jk} \theta_{ik} + d_j)} \quad (1)$$

其中  $\theta_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iK})'$  为被试  $i$  的  $K$  维能力参数向量。生成临时测验后, 生成每个  $K_3^*$  和  $K_5^*$  条件的基础 Q 矩阵及其项目参数 (基础 Q 矩阵见附

录)。最后, 在基础 Q 矩阵的基础上, 根据项目数量以及 Q 矩阵包含可能的  $I_K$  数量, 生成不同条件下的 Q 矩阵及其项目参数。下文以  $K_5^*=2$  为例说明最后两步:

(1) 由于每个 Q 矩阵至少包含一个  $I_K$ , 因此在  $K_5^*=2$  的基础 Q 矩阵中, 其前  $K$  行由  $I_K$  组成, 而剩余的行由随机测量两个维度的项目组成。此外, 该基础 Q 矩阵的项目参数在临时测验的基础上获得。具体而言, 若项目  $j$  测量了维度  $k$ , 那么其  $a_{jk}$  与临时测验的  $a_{jk}$  相等, 否则限定为 0。对于项目  $j$  的  $d_j$ , 设定它与临时测验的  $d_j$  相等。(2) 生成基础 Q 矩阵后, 若  $J=XK$  ( $X$  为 3、4、5 和 6 中的任意一个数) 且  $I_K$  数量为 1, 此时 Q 矩阵为基础 Q 矩阵的前  $XK$  行; 若项目数量不变且  $I_K$  数量为 2, 则 Q 矩阵的前  $2K$  行由 2 个  $I_K$  组成, 其余部分为基础 Q 矩阵的  $2K+1$  至  $XK$  行; 若项目数量不变且  $I_K$  数量为 3, 则 Q 矩阵的前  $3K$  行由 3 个  $I_K$  组成, 其余部分为基础 Q 矩阵的  $3K+1$  至  $XK$  行, 其余条件以此类推。另外, 对于该步骤所生成的 Q 矩阵, 其项目参数的设定方式与基础 Q 矩阵类似, 不同之处在于这些 Q 矩阵的项目参数在基础 Q 矩阵的基础上获得。

被试数  $N=1000$ , 能力参数向量真值从  $MVN(\mathbf{0}, \Sigma)$  中随机抽取, 其中  $\Sigma$  是对角线元素为 1, 非对角线元素按如下方式生成的协方差矩阵: 当各维度之间为低相关,  $\Sigma$  的非对角线元素从  $U(0.1, 0.3)$  中随机抽取; 当各维度之间为高相关,  $\Sigma$  的非对角线元素从  $U(0.5, 0.7)$  中随机抽取 (Cho et al., 2021)。待得到所有参数真值后, 根据 M2PL 模型生成作答矩阵。

考虑到项目参数估计的误差是一个可排除的因素, 因此除了上述条件之外, 本研究还设定有无项目参数估计误差两种条件。在有项目参数估计误差的条件中, 项目参数估计采用目前常用的 MH-RM (Metropolis-Hastings Robbins-Monro) 算法 (Cai, 2010), 该过程借助 mirt 包 (Chalmers, 2012) 完成, 然后将估计值用于能力参数向量估计。在无项目参数估计误差的条件中, 直接将项目参数真值用于能力参数向量估计。能力参数向量的估计采用期望后验估计 (Expected A Posteriori, EAP; Segall, 1996)。EAP 本质上是将参数的后验分布均值作为其估计值, 一般采用积分求解 (Wang, 2015), 但也可直接通过 MCMC 算法求解 (Ulitzsch et al., 2022)。由于 mirt 包无法用于所有的模拟条件, 因而采用 JAGS (Plummer, 2017) 的 MCMC 算法求解, 具体设定如下: 设定两条随机起点的马尔科夫链, 迭代 10000 次

且前 4000 次为预热值。设定能力参数向量的先验分布为多元正态分布,其均值向量为所有元素为 0 的向量。另外,参考 mirt 包的 fscores 函数,当有项目参数估计误差时,将 MH - RM 算法估计的协方差矩阵作为多元正态分布的协方差矩阵,否则直接将协方差矩阵的真值作为多元正态分布的协方差矩阵。上述设定可保证参数估计的结果达到收敛标准,即潜在量尺缩减因子小于 1.2 (Gelman & Rubin, 1992)。有关代码可参阅 Man 等人 (2019) 的附录。

除了 Q 矩阵外,上述过程重复 20 次以减小随机误差。

## 2.2 评价指标

采用均方误差 (Mean Square Error, MSE) 来评价项目参数的估计精度:

$$MSE = \frac{\sum_{r=1}^R (\hat{x}_r - x_r)^2}{R} \quad (2)$$

采用平均欧式距离 (Mean Euclidean Distance, Mean ED) (Chen et al., 2017) 来评价能力参数向量的估计精度:

$$MeanED = \frac{\sum_{r=1}^R \sum_{i=1}^N [\sum_{k=1}^K \hat{\theta}_{ikr} - \theta_{ikr}]^2}{N \times R}^{1/2} \quad (3)$$

其中  $r$  为第  $r$  次重复且  $R = 20$ ;  $\hat{x}_r$  和  $x_r$  分别为第  $r$  次重复的估计值与真值 ( $\hat{\theta}_{ikr}$  和  $\theta_{ikr}$  同理)。上述指标越小越好。

## 2.3 模拟结果

图 1 呈现了  $K = 3$  在有项目参数估计误差条件下的能力参数向量估计精度结果 (需要注意的是,由于项目间多维设计不包含  $I_K$  数量的变化,因而其结果在图中与横坐标平行,图 2 至图 4 同理)。当 Q 矩阵包含的  $I_K$  数量为 1 时,所有项目内多维设计之间的差异最大。随着  $I_K$  数量上升,它们之间的差异随之减小,且与  $K_3^* = 1$  的表现接近。综合来看,  $K_3^* = 2$  的表现最优,随后依次为  $K_3^* = 1$  和  $K_3^* = 3$ 。需要注意的是,对于  $K_3^* = 3$ ,其表现虽在多数条件下优于或接近  $K_3^* = 1$ ,但在部分条件下明显劣于  $K_3^* = 1$ ,因此不能认为  $K_3^* = 3$  在总体上优于  $K_3^* = 1$ 。图 2 呈现了  $K = 3$  在无项目参数估计误差条件下的结果。总体上,各 Q 矩阵设计的表现略优于有项目参数估计误差的条件,但是图 2 的变化趋势与图 1 一致。

图 3 呈现了  $K = 5$  在有项目参数估计误差条件下的能力参数向量估计精度结果。与  $K = 3$  的结果类似,当 Q 矩阵包含的  $I_K$  数量为 1 时,所有项目内多维设计之间的差异最大。随着  $I_K$  数量上升,它们

之间的差异随之减小,且与  $K_5^* = 1$  的表现接近。综合来看,  $K_5^* = 2$  的表现最优,随后依次为  $K_5^* = 3$ 、 $K_5^* = 1$ 、 $K_5^* = 4$  和  $K_5^* = 5$ 。需要注意的是,  $K_5^* = 4$  的表现虽在多数条件下接近或略优于  $K_5^* = 1$ ,但在部分条件下明显劣于  $K_5^* = 1$ ,因此不能认为  $K_5^* = 4$  在总体上优于  $K_5^* = 1$ 。图 4 呈现了  $K = 5$  在无项目参数估计误差条件下的结果。总体上,各 Q 矩阵设计的表现优于有项目参数估计误差的条件,这种现象在复杂 Q 矩阵上更明显 (如  $K_5^* = 5$ )。不过,图 4 的变化趋势依旧与图 3 一致。

附表 1 至附表 3 呈现了项目参数估计精度结果。结果显示 Q 矩阵越复杂,参数的估计精度呈下降趋势,且相较于截距参数,区分度参数受到的影响更大。另外,相关越高,参数的估计精度呈下降趋势,这些结果与 Cho 等人 (2021) 的研究结果一致。不过,能力参数向量的估计精度依旧会随着相关变高而上升,表明高相关带来的信息依旧有利于能力参数向量的估计。

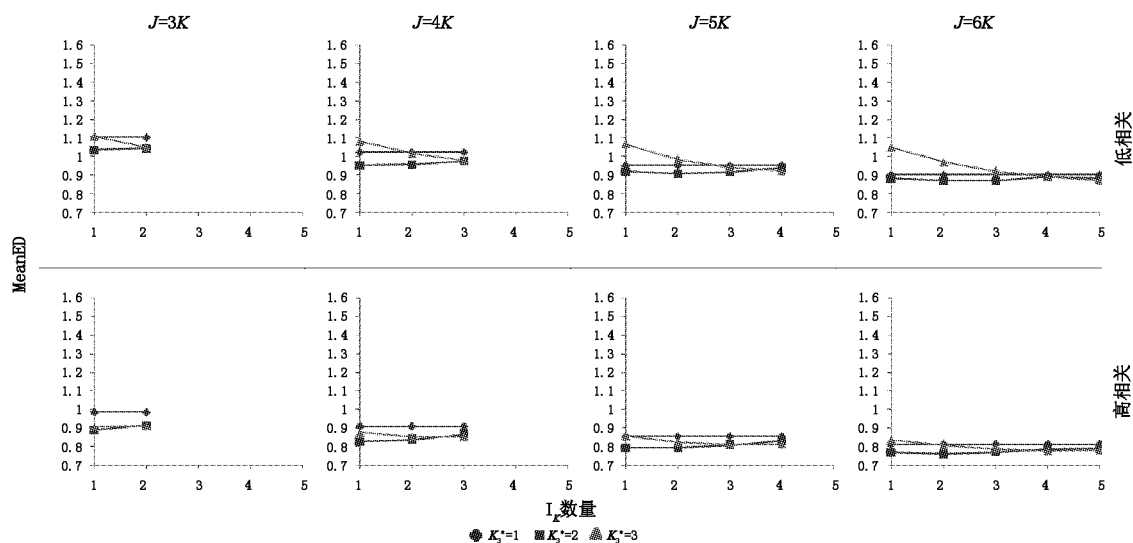
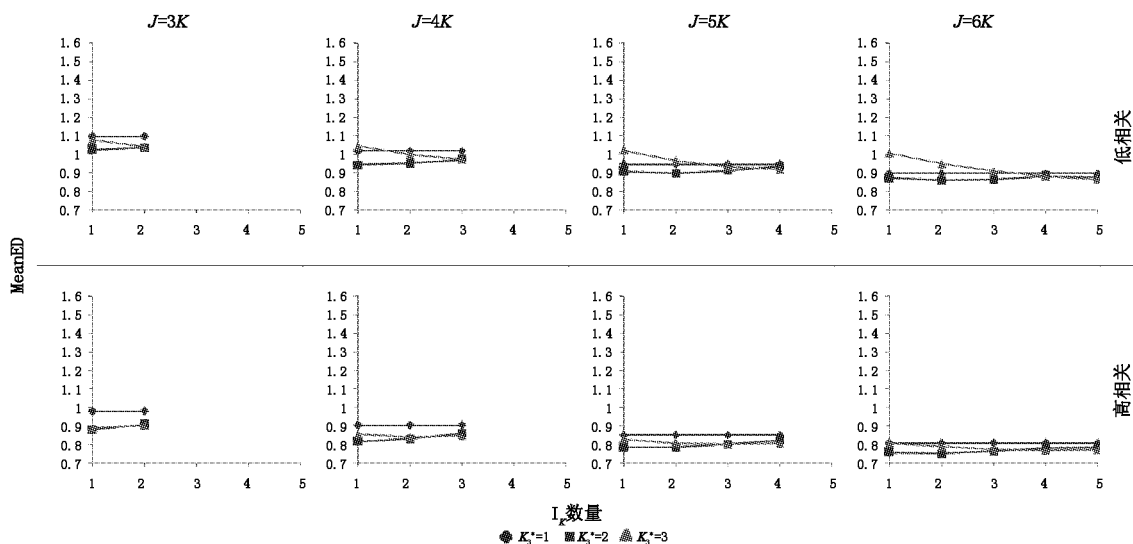
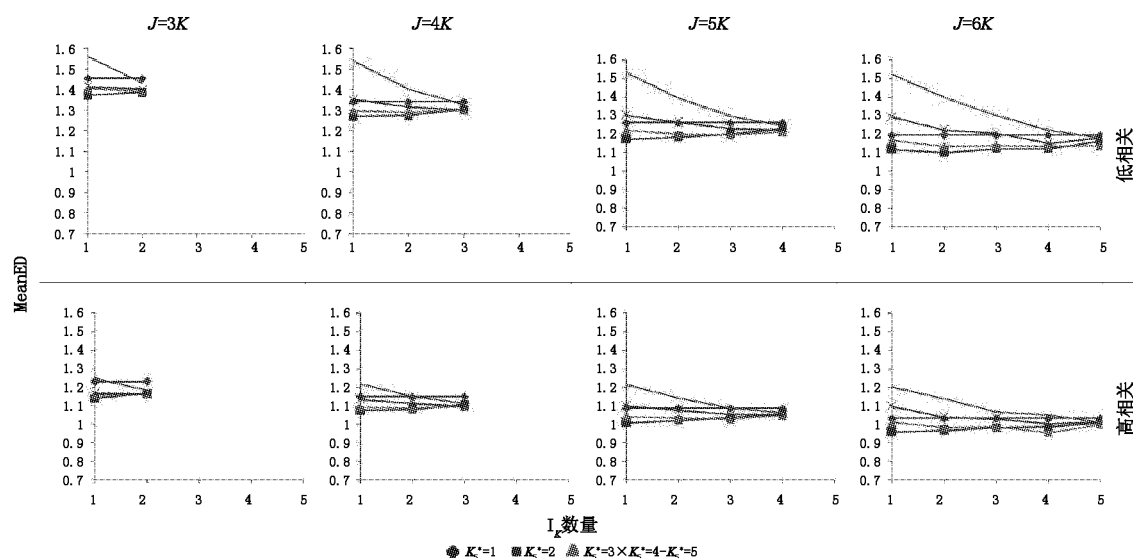
此外,从能力参数向量估计来看,项目间多维设计并不是最优设计,而复杂的项目内多维设计 ( $K_3^* = 3$ 、 $K_5^* = 4$  和  $K_5^* = 5$ ) 虽能提供更高的 MDISC,但无论是否存在项目参数估计误差,其在能力参数估计上的表现通常不及更简单的项目内多维设计,这表明一个 Q 矩阵想要在能力参数向量估计上取得优势,需要兼顾 MDISC 和参数估计算法的偏差。

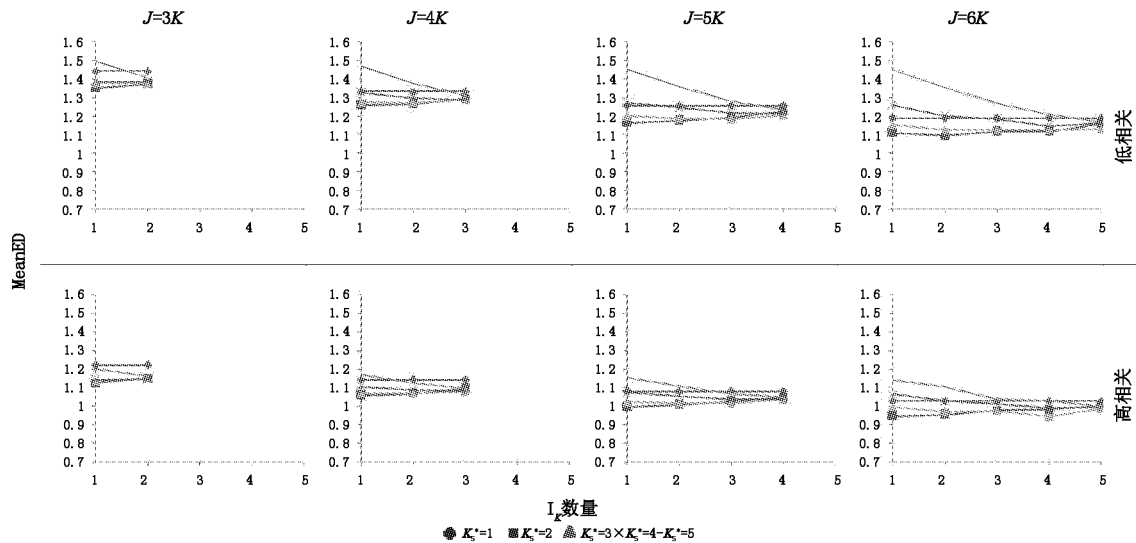
最后,随着项目数量的增加,只有能力参数向量的估计精度会随之上升,而项目参数估计精度的变化无特别规律,这是因为项目数量的增加会降低随机误差对能力参数向量估计的干扰 (Feuerstahler, 2018),而项目参数的估计精度更依赖于被试数量和能力参数向量的分布 (Feuerstahler, 2018; Wang, 2015)。

## 3 总结与展望

### 3.1 总结

不同 Q 矩阵设计不仅反映了不同的测验蓝图,还会对补偿 MIRT 模型的可识别性和参数估计精度造成不同的影响。本文在前人研究的基础上,以模型可识别为前提,通过模拟研究进一步探讨了不同 Q 矩阵设计对该类模型的参数估计精度所造成的影响。研究结果表明:(1)项目间多维设计仅在项目参数估计上占优势;(2)在绝大多数条件下,由模型识别和测量两个维度的项目组成的 Q 矩阵能提供比其他设计更精确的能力参数向量估计值,同时项目参数估计的误差也在可接受范围内。

图1 能力参数向量估计精度( $K=3$ 且有项目参数估计误差)图2 能力参数向量估计精度( $K=3$ 且无项目参数估计误差)图3 能力参数向量估计精度( $K=5$ 且有项目参数估计误差)

图4 能力参数向量估计精度( $K=5$ 且无项目参数估计误差)

根据模拟研究的结果,建议实践者在编制测验的过程中,若以补偿MIRT为测量模型,首先应当保证模型识别,然后剩余的项目应当以测量两个维度的项目为主。当然,在不得不使用复杂Q矩阵的情况下,需考虑增加样本量(包括项目数量和被试数量)来提高参数估计的精度。

### 3.2 展望

目前,有研究者探讨了补偿MIRT模型在多维计算机化自适应测验(Multidimensional Computerized Adaptive Testing, MCAT)中的Q矩阵或题库设计问题,但未能得到一致的结论:Chen和Wang(2016)的研究发现,能力参数向量估计值在项目间多维题库中更精确,他们认为这可能是因为项目间多维题库有足够多的项目单独测量了每个维度。相反,任赫和陈平(2021)以及Chen等人(2017)的研究发现,由于项目内多维题库能提供更高的MDISC,因而能力参数向量的估计精度在项目间多维设计的题库中更低。然而,由于本文的模拟研究在非自适应的测验情境下进行,没有考虑MCAT的特点,因而本文的结论无法适用于该类测验情境,未来可深入探讨如何设计补偿MIRT模型在MCAT中的Q矩阵或题库。

另外,如引言所述,由于非补偿MIRT模型的参数估计比较复杂,本文仅关注了补偿MIRT模型的Q矩阵设计问题,未来可深入探讨非补偿MIRT模型的Q矩阵设计问题。

### 参考文献

彭亚风,罗照盛,喻晓峰,高椿雷,李喻骏.(2016). 认知诊断评价中测验结构的优化设计. *心理学报*, 48(12), 1600 -

1611.

任赫,陈平.(2021). 两种新的多维计算机化分类测验终止规则. *心理学报*, 53(9), 1044 - 1058.

Adams, R. J., Wilson, M., & Wang, W. - C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1 - 23.

Babcock, B. (2011). Estimating a noncompensatory IRT model using Metropolis within Gibbs sampling. *Applied Psychological Measurement*, 35(4), 317 - 329.

Béguin, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model - fit analysis of multidimensional IRT models. *Psychometrika*, 66(4), 541 - 561.

Cai, L. (2010). Metropolis - Hastings Robbins - Monro algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics*, 35(3), 307 - 335.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1 - 29.

Chalmers, R. P. (2020). Partially and fully noncompensatory response models for dichotomous and polytomous items. *Applied Psychological Measurement*, 44(6), 415 - 430.

Chen, P., & Wang, C. (2016). A new online calibration method for multidimensional computerized adaptive testing. *Psychometrika*, 81(3), 674 - 701.

Chen, P., Wang, C., Xin, T., & Chang, H. - H. (2017). Developing new online calibration methods for multidimensional computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 70(1), 81 - 117.

Cho, A. E., Wang, C., Zhang, X., & Xu, G. (2021). Gaussian variational estimation for multidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*, 74(S1), 52 - 85.

Embretson, S. E., & Yang, X. (2013). A multicomponent latent

- trait model for diagnosis. *Psychometrika*, 78(1), 14 – 36.
- Feuerstahler, L. M. (2018). Sources of error in IRT trait estimation. *Applied Psychological Measurement*, 42(5), 359 – 375.
- Fu, Z., Zhang, S., Su, Y. H., Shi, N., & Tao, J. (2021). A Gibbs sampler for the multidimensional four - parameter logistic item response model via a data augmentation scheme. *British Journal of Mathematical and Statistical Psychology*, 74(3), 427 – 464.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457 – 472.
- Holman, R., & Glas, C. A. (2005). Modelling non - ignorable missing - data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology*, 58(1), 1 – 17.
- Man, K., Harring, J. R., Jiao, H., & Zhan, P. (2019). Joint modeling of compensatory multidimensional item responses and response times. *Applied Psychological Measurement*, 43(8), 639 – 654.
- Plummer, M. (2017). *JAGS version 4.3.0 user manual* [Computer software manual]. [https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/jags\\_user\\_manual.pdf](https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/jags_user_manual.pdf)
- Reckase, M. D. (2009). *Multidimensional item response theory*. Springer.
- Reckase, M. D., & McKinley, R. L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15(4), 361 – 373.
- Sun, J., Chen, Y., Liu, J., Ying, Z., & Xin, T. (2016). Latent variable selection for multidimensional item response theory models via  $L_1$  regularization. *Psychometrika*, 81(4), 921 – 939.
- Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika*, 61(2), 331 – 354.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, 10(1), 55 – 73.
- Ulitzsch, E., Yildirim - Erbasli, S. N., Gorgun, G., & Bulut, O. (2022). An explanatory mixture IRT model for careless and insufficient effort responding in self - report measures. *British Journal of Mathematical and Statistical Psychology*, 75(3), 668 – 698.
- Wang, C. (2015). On latent trait estimation in multidimensional compensatory item response models. *Psychometrika*, 80(2), 428 – 449.
- Wang, C., & Nydick, S. W. (2015). Comparing two algorithms for calibrating the restricted non - compensatory multidimensional IRT model. *Applied Psychological Measurement*, 39(2), 119 – 134.
- Wang, W. - C., Chen, P. - H., & Cheng, Y. - Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9(1), 116 – 136.
- Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, 81(3), 625 – 649.
- Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed - format tests. *Applied Psychological Measurement*, 30(6), 469 – 492.

## The Q - matrix Design for Compensatory Multidimensional IRT Models

Pan Shiquan<sup>1</sup>, Zhao Shouying<sup>1,2</sup>

(1. School of Psychology, Guizhou Normal University, Guiyang 550025; 2. Kaili University, Kaili 556011)

**Abstract:** The Q - matrix of multidimensional item response theory (MIRT) models is restricted to binary variables and captures the relationship between items and dimensions. Different Q - matrices can impact the model identifiability and accuracy of parameter estimates in MIRT models. Previous studies have found that items measuring one dimension tend to produce more accurate parameter estimates in compensatory MIRT models. However, items measuring more than one dimension may produce more accurate ability vector estimates because they have higher discrimination. In this article, we further investigated the effects of Q - matrix design on parameter estimates through simulation study under non - adaptive testing scenario. The simulation results showed that (1) compared to other Q - matrix designs, between - item multidimensionality design only had an advantage in item parameter estimates; (2) under almost all conditions, the Q - matrix contained both items guaranteeing model identifiability and items measuring two dimensions not only produced more accurate ability vector estimates than other Q - matrix designs, but also provided acceptable item parameters recovery.

**Key words:** multidimensional item response theory; Q - matrix; within - item multidimensionality; between - item multidimensionality

附录

附表 1  $K=3$  的项目参数估计精度

项目数	$I_K$ 数	$K_3^* = 2$				$K_3^* = 3$			
		低相关		高相关		低相关		高相关	
		区分度	截距	区分度	截距	区分度	截距	区分度	截距
3K	1	0.039	0.026	0.077	0.015	0.119	0.030	0.240	0.034
	2	0.032	0.021	0.058	0.018	0.050	0.026	0.130	0.025
	3	<b>0.030</b>	<b>0.019</b>	<b>0.017</b>	<b>0.015</b>	<b>0.030</b>	<b>0.019</b>	<b>0.017</b>	<b>0.015</b>
4K	1	0.035	0.020	0.058	0.019	0.123	0.032	0.212	0.042
	2	0.021	0.016	0.040	0.018	0.060	0.027	0.140	0.026
	3	0.031	0.020	0.023	0.020	0.029	0.022	0.057	0.019
	4	<b>0.016</b>	<b>0.014</b>	<b>0.012</b>	<b>0.011</b>	<b>0.016</b>	<b>0.014</b>	<b>0.012</b>	<b>0.011</b>
5K	1	0.027	0.019	0.061	0.021	0.137	0.251	0.251	0.044
	2	0.020	0.019	0.049	0.018	0.056	0.133	0.133	0.028
	3	0.015	0.015	0.026	0.018	0.037	0.102	0.102	0.022
	4	0.017	0.015	0.021	0.014	0.019	0.047	0.047	0.018
	5	<b>0.012</b>	<b>0.013</b>	<b>0.009</b>	<b>0.011</b>	<b>0.012</b>	<b>0.009</b>	<b>0.009</b>	<b>0.011</b>
6K	1	0.023	0.023	0.059	0.022	0.127	0.032	0.216	0.045
	2	0.023	0.018	0.039	0.017	0.053	0.028	0.137	0.033
	3	0.017	0.016	0.042	0.015	0.036	0.022	0.113	0.025
	4	0.014	0.015	0.028	0.015	0.024	0.016	0.065	0.023
	5	0.010	0.013	0.015	0.013	0.017	0.015	0.032	0.014
	6	<b>0.010</b>	<b>0.011</b>	<b>0.009</b>	<b>0.013</b>	<b>0.010</b>	<b>0.011</b>	<b>0.009</b>	<b>0.013</b>

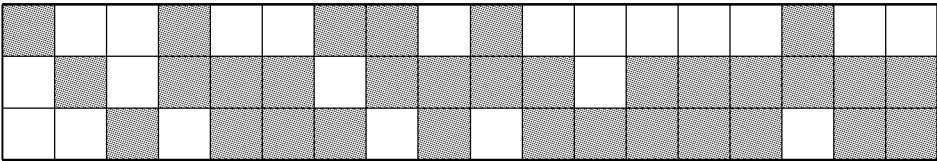
注:加粗为项目间多维设计的结果。下同。

附表 2  $K=5$  的项目参数估计精度(低相关)

项目数	$I_K$ 数	$K_5^* = 2$		$K_5^* = 3$		$K_5^* = 4$		$K_5^* = 5$	
		区分度	截距	区分度	截距	区分度	截距	区分度	截距
3K	1	0.030	0.022	0.056	0.024	0.083	0.045	0.219	0.048
	2	0.019	0.016	0.032	0.018	0.041	0.023	0.071	0.030
	3	<b>0.013</b>	<b>0.014</b>	<b>0.013</b>	<b>0.014</b>	<b>0.013</b>	<b>0.014</b>	<b>0.013</b>	<b>0.014</b>
4K	1	0.019	0.020	0.046	0.024	0.069	0.045	0.217	0.073
	2	0.017	0.017	0.029	0.019	0.040	0.031	0.082	0.038
	3	0.020	0.018	0.015	0.019	0.023	0.017	0.036	0.019
	4	<b>0.010</b>	<b>0.013</b>	<b>0.010</b>	<b>0.013</b>	<b>0.010</b>	<b>0.013</b>	<b>0.010</b>	<b>0.013</b>
5K	1	0.014	0.017	0.032	0.028	0.056	0.051	0.170	0.039
	2	0.011	0.018	0.020	0.020	0.038	0.035	0.079	0.038
	3	0.009	0.012	0.018	0.038	0.027	0.032	0.040	0.025
	4	0.008	0.013	0.012	0.017	0.022	0.035	0.028	0.016
	5	<b>0.006</b>	<b>0.011</b>	<b>0.006</b>	<b>0.011</b>	<b>0.006</b>	<b>0.011</b>	<b>0.006</b>	<b>0.011</b>
6K	1	0.013	0.024	0.023	0.021	0.054	0.033	0.224	0.062
	2	0.013	0.013	0.021	0.025	0.051	0.031	0.090	0.044
	3	0.009	0.018	0.014	0.023	0.031	0.029	0.045	0.032
	4	0.008	0.013	0.013	0.017	0.016	0.008	0.028	0.023
	5	0.007	0.014	0.010	0.015	0.015	0.015	0.017	0.016
	6	<b>0.007</b>	<b>0.016</b>	<b>0.007</b>	<b>0.016</b>	<b>0.007</b>	<b>0.016</b>	<b>0.007</b>	<b>0.016</b>

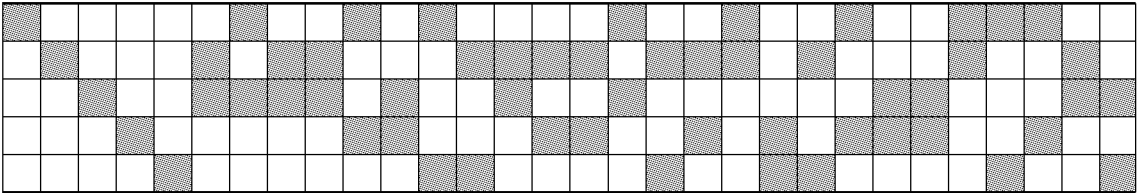
附表 3  $K=5$  的项目参数估计精度 (高相关)

项目数	$I_k$ 数	$K_5^* = 2$		$K_5^* = 3$		$K_5^* = 4$		$K_5^* = 5$	
		区分度	截距	区分度	截距	区分度	截距	区分度	截距
3K	1	0.040	0.019	0.111	0.038	0.193	0.062	0.421	0.098
	2	0.027	0.019	0.064	0.026	0.112	0.036	0.223	0.047
	3	<b>0.007</b>	<b>0.012</b>	<b>0.007</b>	<b>0.012</b>	<b>0.007</b>	<b>0.012</b>	<b>0.007</b>	<b>0.012</b>
4K	1	0.040	0.023	0.105	0.040	0.186	0.065	0.377	0.100
	2	0.021	0.017	0.067	0.034	0.120	0.048	0.250	0.076
	3	0.018	0.014	0.033	0.018	0.065	0.039	0.124	0.042
	4	<b>0.005</b>	<b>0.011</b>	<b>0.005</b>	<b>0.011</b>	<b>0.005</b>	<b>0.011</b>	<b>0.005</b>	<b>0.011</b>
5K	1	0.026	0.020	0.074	0.043	0.166	0.073	0.402	0.096
	2	0.026	0.016	0.055	0.029	0.109	0.055	0.310	0.076
	3	0.016	0.014	0.035	0.020	0.076	0.043	0.169	0.047
	4	0.010	0.015	0.029	0.018	0.045	0.025	0.088	0.031
	5	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>
6K	1	0.025	0.020	0.083	0.043	0.173	0.080	0.469	0.086
	2	0.023	0.017	0.053	0.033	0.134	0.082	0.297	0.083
	3	0.013	0.017	0.042	0.034	0.104	0.046	0.270	0.054
	4	0.014	0.015	0.032	0.020	0.060	0.041	0.144	0.035
	5	0.007	0.011	0.020	0.013	0.030	0.015	0.064	0.019
	6	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>	<b>0.005</b>	<b>0.012</b>

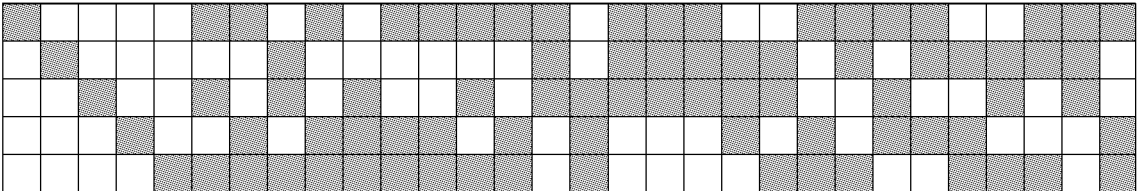


附图 1  $K_5^* = 2$  的基础 Q 矩阵

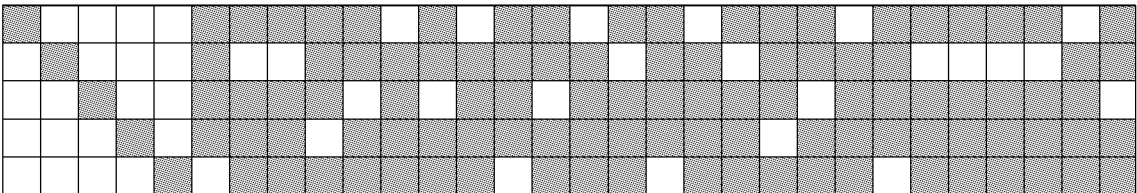
注:行代表维度,列代表项目,灰色代表该项目测量了该维度,下同。



附图 2  $K_5^* = 2$  的基础 Q 矩阵



附图 3  $K_5^* = 3$  的基础 Q 矩阵



附图 4  $K_5^* = 4$  的基础 Q 矩阵