

《心理学报》审稿意见与作者回应

题目：认知诊断模型 Q 矩阵修正：完整信息矩阵的作用

作者：刘彦楼，吴琼琼

第一轮

审稿人 1 意见：

论文《认知诊断模型 Q 矩阵修正：完整信息矩阵的作用》提出了基于完整经验交叉相乘信息矩阵(XPD)构建的 Wald-XPD 方法来修正认知诊断模型中的 Q，模拟研究表明，相对于其它两种 Q 矩阵修正方法，Wald-XPD 方法在大多数条件下可以获得更好的修正结果。实证研究也发现，基于 Wald-XPD 方法修正的 Q 矩阵可以更好地拟合测验数据。论文内容丰富、逻辑较为清晰，具有一定的理论和实践价值。但研究还存在以下值得进一步探讨的地方：

回应：感谢审稿专家对本文的理论和实践价值的肯定，以及在审稿过程中所付出的时间与精力。以下是对审稿专家意见的回应。稿件正文中的修改部分用蓝色字体标注。

意见 1：目前，既存在参数类 Q 矩阵修正方法，也存在非参数类的 Q 矩阵修正方法，作者为何只关注参数类的 Q 矩阵修正方法？

回应：感谢审稿专家的意见。参数化和非参数化 Q 矩阵修正方法的区别在于是否采用参数化的 CDM。参数化的 Q 矩阵修正方法需要参数化的 CDM 的参与；非参数化方法使用统计分析比较理想反应与观察作答反应之间的拟合，从而达到修正 Q 矩阵的目的。然而，非参数化方法中理想反应大多都是在限制条件较为严格情景下获得的，例如，限定所有项目只适用于某个或某几个特殊的(亦称，简化的)CDM，非参数化方法严格的使用条件限制了这些方法的拓展性及实用性。鉴于此，本研究选择关注灵活性更高、拓展性更强的参数化的 Q 矩阵修正方法。

意见 2：文献综述的全面性有待加强，当前除了作者列出的关于 Q 矩阵修正方法的文献外，还有一些作者尚未收集到的文献，如 Terzi 和 de la Torre (2018) 提出的迭代序列搜索算法 (sequential search algorithm)，Kang、Yang 和 Zeng (2019) 提出的基于 RMSEA 的方法，Feng (2013) 提出的两阶段序列搜索算法；Terzi (2017) 提出的迭代 JSD (iterative

Jensen-Shannon Divergence, iJSD) 指标和迭代 GDI 方法; 以及 Lim 和 Drasgow (2017) 提出的非参数 Q 矩阵校准方法等。李佳、毛秀玲和张雪琴 (2021) 对 Q 矩阵修正 (估计) 方法进行了系统介绍, 作者也可参考李佳等人的相关文献。

FENG, Y. (2013). Estimation and Q-matrix validation for diagnostic classification models. (Master's thesis).

Retrieved from <http://scholarcommons.sc.edu/etd/2611>

Terzi, R. (2017). New Q-matrix validation procedures (Doctoral dissertation, Rutgers University-School of Graduate Studies).

Terzi, R., & De la Torre, J. (2018). An iterative method for empirically-based Q-matrix validation. *International Journal of Assessment Tools in Education*, 5(2), 248-262.

Lim, Y. S., & Drasgow, F. (2017). Nonparametric calibration of item-by-attribute matrix in cognitive diagnosis. *Multivariate behavioral research*, 52(5), 562-575.

Kang, C., Yang, Y., & Zeng, P. (2019). Q-matrix refinement based on item fit statistic RMSEA. *Applied psychological measurement*, 43(7), 527-542.

回应: 感谢审稿专家的建设性意见。修改稿中已对“引言”的文献综述部分进行重新梳理和阐述。

结合意见 1, 首先区分了非参数化和参数化的 Q 矩阵修正方法。修改稿中内容如下(第 1 页):

“根据是否采用参数化的 CDM 描述 Q 矩阵与观察作答数据之间的关系, 可以将 Q 矩阵修正方法分为两类: 参数化与非参数化的修正方法。前者需要参数化 CDM 的参与, 后者不需要。例如, 欧氏距离法(Chiu, 2013)、海明距离(汪大勋, 高旭亮, 韩雨婷 等, 2018)、交差方法(intersection and difference; Wang et al., 2018)等属于非参数化的修正方法。一般而言, 非参数化方法比较的是理想反应与观察作答反应之间的拟合, 从而达到修正 Q 矩阵的目的。非参数化方法中, 理想反应大多都是在限制条件较为严格情景下获得的, 例如, 限定所有项目只适用于某个或某几个特殊的(亦称, 简化的)CDM。也就是, 尽管非参数化的 Q 矩阵修正方法具有样本量要求小、易实现等优点, 但是严格的前提条件限制了这些方法的拓展性及实用性。”

其次, 根据参数化 Q 矩阵修正方法的适用范围, 将其分为适用于简化 CDM 的修正方法和饱和 CDM 的修正方法。修改稿中的内容如下(第 2 页):

“参数化 Q 矩阵修正方法是在参数化模型框架下, 使用各种统计量估计出最能拟合观察数据

的 \mathbf{Q} 矩阵。在特殊的 CDM 框架下, 如 DINA (Haertel, 1984)、DINO (Templin & Henson, 2006)、R-RUM (Hartz & Roussos, 2008) 等, 研究者开发的参数化修正方法主要有: δ 法(de la Torre, 2008)、 γ 法(涂冬波 等, 2012)、S 统计量方法(Liu et al., 2012)、迭代修正序列搜索(iterative modified sequential search; Terzi & de la Torre, 2018)、RMSEA 统计量(Kang et al., 2019)、加权残差 R 法(Yu & Cheng, 2020)、最优反应分布纯度方法(李佳 等, 2022)等。饱和 CDM 框架下的参数化 \mathbf{Q} 矩阵修正方法主要包括: GDI(G-DINA discrimination index)方法(de la Torre & Chiu, 2016)、残差方法(Chen, 2017)、iJSD (iterative Jensen-Shannon divergence)方法以及 iGDI (iterative GDI)方法(Terzi, 2017)、TLP (truncated L_1 penalty function)方法(Xu & Shang, 2018)、相对拟合统计量方法(汪大勋 等, 2020)、Ma 和 de la Torre (2020)提出的 GDI 和基于不完整信息矩阵(incomplete information matrix)的 Wald 检验相结合的 Stepwise 方法(为了便于理解且与本文中提出的新方法加以区分, 将 Stepwise 方法称为 Wald-IC 方法)、以及 Hull 方法(Na jera et al., 2021a)等。尽管一些参数化的修正方法可能存在运算量大、速度慢等问题, 但是, 这类修正方法尤其是在饱和的 CDM 框架下开发的方法的优点在于灵活性高、不需要非参数化方法那样严格的前提假设。因为饱和模型包含多类特殊模型作为特例, 且在 \mathbf{Q} 矩阵没有错误设定或存在少量错误时, 可以较为容易地通过模型比较的方法获得恰当的特殊模型。”

第三, 在稿件的“6.2 研究展望”部分, 增加了如下内容(第 22 页):

“Wang 等人(2020)评估了在 \mathbf{Q} 矩阵部分已知的情况下, GDI 和 Wald-IC 方法在估计新项目的 \mathbf{q} 向量中的表现。基于此, 未来研究者可以在 \mathbf{Q} 矩阵部分已知的情况下进一步评估 Wald-XPD 方法估计 \mathbf{Q} 矩阵的表现, 并与已有的 \mathbf{Q} 矩阵估计方法, 如 ICC-IR 方法(汪大勋, 高旭亮, 蔡艳 等, 2018)、似然比 D^2 方法(喻晓锋 等, 2015)、非参数 \mathbf{Q} 矩阵校准(Lim & Drasgow, 2017)、两阶段搜索算法(Feng, 2013)、似然比检验(Wang et al., 2020)等方法进行比较。”

意见 3: 建议作者加强对问题提出的论述, 审稿人了解作者所研究的内容, 只是对于论文的问题提出部分 (如以往研究的不足), 审稿人觉得还不是非常清晰。

回应: 感谢审稿专家的宝贵意见。首先, 我们发现以往研究者(Ma et al., 2019; Na jera et al., 2021a)提出的 Wald-IC 及 Hull 方法用于 \mathbf{Q} 矩阵修正时有较好的表现, 但 Wald-IC 以及 Hull 方法在正确修正错误标定的属性方面的表现较差。其次, 在梳理完整和不完整信息矩阵的类型和应用时, 发现不完整信息矩阵构建的统计量用于后续的统计检验时, 会导致一些错误。例如, 低估模型参数标准误、以及不完整信息矩阵构建的 Wald 统计量用于项目水平模型比较和项目功能差异检验时导致一类错误控制率膨胀。故综合这两方面的内容, 本研究认为

Wald-IC 在正确修正错误标定的属性方面的较差表现可能是因为在构建 Wald 统计量时采用的是不完整信息矩阵。在修改稿中我们加强了对问题提出部分的论述，修改稿中的内容如下(第 3 页):

“研究者(Najera et al., 2021a)通过模拟研究比较了 GDI、Wald-IC 以及 Hull 方法，发现在大多数条件下 Hull 的表现最好、Wald-IC 的表现稍逊于 Hull。但是，Hull 和 Wald-IC 在正确修正错误标定的属性方面的表现较差，尤其是 Q 矩阵中存在较多错误设定时。可以发现，Wald-IC 统计量是使用不完整信息矩阵计算的。先前研究表明，采用不完整信息矩阵构建的统计量在后续研究中会导致一些错误，如低估模型参数标准误(Philipp et al., 2018)、用于项目功能差异检验及项目水平模型比较时导致一类错误控制率膨胀(Liu, Andersson et al., 2019; Liu, Yin et al., 2019; 刘彦楼 等, 2016)等。基于此，本研究认为 Wald-IC 方法在正确修正错误标定属性方面表现较差的主要原因可能是在 Wald 统计量的计算中采用了不完整的信息矩阵。”

意见 4: 在和以往 Q 矩阵修正方法进行比较时，作者只比较参数方法的原因是什么？

回应: 感谢审稿专家的宝贵问题，对我们行文有重要启发。首先，通过对以往研究者提出的 Q 矩阵修正方法的梳理，发现参数化 Q 矩阵修正方法中，包括适用于简化 CDM 的方法和饱和 CDM 框架下提出的 Q 矩阵修正方法。因为饱和模型中包含多类简化模型作为特例，且在 Q 矩阵没有错误设定或存在少量错误时，可以较为容易地通过模型比较的方法获得恰当的简化模型。故，本研究选择拓展性更强的饱和模型框架下提出的 Q 矩阵修正方法。

其次，饱和模型框架下提出的 Q 矩阵修正方法中，各有优缺点。根据审稿人的意见，本研究增加了对 8 种饱和模型下提出的 Q 矩阵修正方法的评述。修改稿中的内容如下(第 2 页):

“在以上八种饱和 CDM 框架下开发的参数化 Q 矩阵修正方法中，残差方法对于属性过度设定不敏感且在测验长度较短时统计检验力可能会偏低；当样本量较小时，TLP 方法会高估错误设定项目的数量且其用于减少错误报告率的重抽样校正方法(bootstrap bagging method)的耗时可能会特别长；模拟研究发现 iGDI 的表现与 iJSD 的表现相当、甚至在一些条件下优于 iJSD (Terzi, 2017)；相对拟合统计量方法需要比较测验的所有项目关于属性所有可能组合的相对拟合值，尽管研究者提出一些减少计算次数的方法，但是在测验长度较长或属性数量较多的情况下，计算耗时仍有可能特别长。GDI 在饱和 CDM 框架下采用单个项目所有可能的属性掌握模式中正确答对概率的方差来衡量 Q 矩阵中相对应的 q 向量的区分能力，选择有

最大区分能力的 \mathbf{q} 向量作为正确设定的 \mathbf{q} 向量。相对于 GDI 而言，iGDI 的估计效果有了一定程度的改善，但是这类方法的主要问题是需要人为地确定一个截止值(Na jera et al., 2021b)。以 GDI 研究为基础，Ma 和 de la Torre (2020)将 \mathbf{Q} 矩阵修正的视角延伸到多级计分模型，在 seq-GDINA 模型(the sequential GDINA model; Ma & de la Torre, 2016)下提出了 GDI 和基于不完整信息矩阵的 Wald 检验相结合的 Wald-IC 方法。Wald-IC 方法首先采用 GDI 方法从单一属性的 \mathbf{q} 向量中确定第一个所需属性，再逐步多次采用 Wald 统计量决定是否增加或删除属性来选择恰当的 \mathbf{q} 向量。也就是，在单个项目上 Wald-IC 仅需执行 $K-1$ 个统计检验即可完成。Hull 方法试图在模型拟合与简约之间找到一种平衡以此选择恰当的 \mathbf{q} 向量，研究者(Na jera et al., 2021a)通过模拟研究比较了 GDI、Wald-IC 以及 Hull 方法，发现在大多数条件下 Hull 的表现最好、Wald-IC 的表现稍逊于 Hull。但是，Hull 和 Wald-IC 在正确修正错误标定的属性方面的表现较差，尤其是 \mathbf{Q} 矩阵中存在较多错误设定时。”

因此，本研究选择考察 GDI、Wald-IC、Hull 及 Wald-XPB 方法在 \mathbf{Q} 矩阵修正中的表现。

意见 5: PVAF 中的分子和分母各代表什么意思?

回应: 感谢审稿专家的意见。我们已对 PVAF 中的分子分母的含义加以解释。

首先，PVAF 的分子为 ζ_j^2 ，在修改稿的“2.1GDI 方法”部分对 ζ_j^2 进行了解释，修改稿中内容如下(第 5 页):

“GDI 方法采用辨别指数 ζ_j^2 (discriminating index)表示正确作答概率的方差，即项目 j 的某个 \mathbf{q} 向量关于所有可能的属性掌握模式的被试正确作答概率的方差:

$$\zeta_j^2 = \sum_{l=1}^{2^K} w(\mathbf{a}_l | \mathbf{x}) [p_j(\mathbf{a}_l) - \bar{p}_j]^2 \quad (2)$$

其次，PVAF 的分母为 $\zeta_{1:K}^2$ ，在修改稿的“2.1GDI 方法”部分对 $\zeta_{1:K}^2$ 进行了解释，修改稿中内容如下(第 6 页):

“故 de la Torre 和 Chiu (2016)计算了 \mathbf{q} 向量的所占方差 PVAF(the proportion of variance accounted for):

$$\text{PVAF} = \frac{\zeta_j^2}{\zeta_{1:K}^2} \quad (4)$$

其中， $\zeta_{1:K}^2$ 表示项目 j 的全为 1 的 \mathbf{q} 向量关于所有可能的属性掌握模式的被试正确作答概率

的方差。”

意见 6: 从 q 向量的修正流程来看, 似乎就是对 q 向量进行估计, 以明确最佳的 q 向量。此时, 审稿人觉得这和 Q 矩阵的估计非常相像。作者可能需要进一步明确 Q 矩阵修正和 Q 矩阵估计之间的本质差异。

回应: 感谢审稿专家的意见。通过对 Q 矩阵修正方法以及估计方法的梳理, 参考 Wang 等人(2020)以及汪大勋等人(2020)对于 Q 矩阵修正和估计的界定, 本研究认为, 若已有关于全部项目的 q 向量的界定(如, 原始 Q 矩阵), 对所有项目的 q 向量进行修正即为 Q 矩阵修正; 若已知部分项目的 q 向量, 对于未知 q 向量的旧项目或者新项目进行估计即为 Q 矩阵估计。本研究是在全部项目的 q 向量已知的前提下(即, 领域专家已经给出原始 Q 矩阵的前提下), 对所有项目的 q 向量进行修正, 故本研究属于 Q 矩阵的修正。

为了进一步明确 Q 矩阵修正和估计之间的差异, 我们在修改稿的“引言”部分, 对 Q 矩阵全部已知前提下提出的 Q 矩阵修正方法进行了综述。此外, 经过适当修改后, Q 矩阵修正方法或估计的原理与方法是可以互相借用的。例如 Wang 等人(2020)的研究中将 Wald-IC 方法用于 Q 矩阵估计, 故本研究认为未来研究者也可以尝试将新开发的 Wald-XPD 方法用于 Q 矩阵估计。所以, 在修改稿的“6.2 研究展望”部分介绍了 Q 矩阵估计的方法, 并提出可以将 Wald-XPD 方法用于 Q 矩阵估计并与现有的 Q 矩阵估计方法进行比较。修改稿中的内容如下(第 22 页):

“Wang 等人(2020)评估了在 Q 矩阵部分已知的情况下, GDI 和 Wald-IC 方法在估计新项目的 q 向量中的表现。基于此, 未来研究者可以在 Q 矩阵部分已知的情况下进一步评估 Wald-XPD 方法估计 Q 矩阵的表现, 并与已有的 Q 矩阵估计方法, 如 ICC-IR 方法(汪大勋, 高旭亮, 蔡艳 等, 2018)、似然比 D^2 方法(喻晓锋 等, 2015)、非参数 Q 矩阵校准(Lim & Drasgow, 2017)、两阶段搜索算法(Feng, 2013)、似然比检验(Wang et al., 2020)等方法进行比较。”

意见 7: Terzi (2017) 提出了迭代 GDI (iGDI) 方法进行 Q 矩阵修正, 结果发现 iGDI 的结果要优于 GDI, 因此, 建议作者在本研究中使用 iGDI 代替 GDI 进行 Q 矩阵修正。

回应: 感谢审稿专家的建设性意见。首先, GDI 方法的迭代分为两种: 第一种是迭代 GDI 方法(iGDI; Terzi, 2017)属于测验水平的迭代, 即在修正完所有项目的 q 向量之后, 更新 Q 矩阵进行迭代。第二种是迭代和动态(Iterative and Dynamic Procedure)的 GDI 方法(Najera et al., 2021b), 属于项目水平的迭代, 即在对某个项目的 q 向量修正之后, 就更新 Q 矩阵, 然

后进行迭代。测验水平的迭代是在所有项目的 \mathbf{q} 向量都修正过的基础上进行的，可能仍然包含较多的错误设定。项目水平的迭代是在只对原始 \mathbf{Q} 矩阵中一个项目的 \mathbf{q} 向量进行修正的基础上进行的，每次迭代都是在相对较优的 \mathbf{Q} 矩阵的基础上进行的，故研究者(Na jera et al., 2021b)认为项目水平的 GDI 迭代优于测验水平的 GDI 迭代。

其次，项目水平的迭代和动态的 GDI 方法在每次修正完一个项目之后，要更新该项目的截止值。因为如果截止值选择不合适，则每次迭代之后选择的该项目的“正确设定的 \mathbf{q} 向量”也是错误的，那么，基于该错误的 \mathbf{q} 向量进行进一步迭代时，就会出现“滚雪球”效应(Thus, each iteration will probably increase the distance between the provisional Q-matrix and the true Q-matrix in a sort of “snowball” effect,), 即错误的 \mathbf{q} 向量迭代之后产生更多错误。故 Na jera 等人(2021b)将 GDI 方法以项目水平迭代的方式进行时，根据更新后的 \mathbf{q} 向量选择了新的截止值。本研究限于研究目的和篇幅，没有对截止值问题进行更新和探讨。

最后，Wald-XPDI 统计量有明确的渐近分布(χ^2 分布)，不需要像 GDI 类方法(GDI、iGDI、迭代和动态 GDI)那样人为地确定一个截止值，故迭代进行时不会因为截止值问题导致“滚雪球”效应。

在修改稿的“引言”部分和“6.2 研究展望”进行了修改。

“引言”部分修改内容如下(第 4 页):“限制 GDI 及 iGDI 方法实践应用的主要原因是这两种方法均需要人为地设置一个截止值。但是 GDI 方法可为每个项目选择具有最大区分能力的单个属性;且与 iGDI 相比固定的截止值对 GDI 方法的影响相对较小(Na jera et al., 2021b), 因此本研究将 GDI 也纳入比较。”

“6.2 研究展望”部分修改内容如下(第 21 页):“虽然 Wald-XPDI 统计量有明确的渐近分布(χ^2 分布)，不需要像 GDI 类方法那样人为地确定一个截止值，但限于研究目的和篇幅本文仅在 0.05 显著性水平上对于 Wald-XPDI 统计量的表现进行了显著性检验。未来研究者可以进一步探讨不同的显著性水平对于 Wald-XPDI 统计量表现的影响。”

意见 8: 建议使用 3 级标题对研究结果进行概括和总结，以增加研究结果的可读性。

回应: 感谢审稿专家的宝贵建议。已对“4.2 研究结果”部分进行概括和总结，并增加了 3 级标题:“4.2.1 4 种方法表现的综合比较”、“4.2.2 Wald-XPDI、Wald-IC 与 HullIP 方法的具体比较”。

为增加研究结果的可读性，使用这两个三级标题将本研究的研究结果分成两个部分。首

先是对 GDI、Hull (HullP、HullR)、Wald-IC 以及 Wald-XPB 方法在各个指标上的综合表现进行分析, 综合比较的结果发现, Wald-XPB 以及 HullP 在各个指标上有相对较好的表现; 鉴于 Wald-XPB 是在 Wald-IC 基础上新提出的方法, 故又对 Wald-XPB、Wald-IC 以及 HullP 方法在 QRR、TPR 以及 TNR 这三个主要指标上的表现进行更详细的分析。相关修改内容如下(第 15 页):

“4.2 研究结果

4.2.1 4 种方法表现的综合比较

表 2 呈现了 GDI、Hull (HullP、HullR)、Wald-IC 以及 Wald-XPB 方法在各个因素水平上的 QRR、TPR、TNR、OS 和 US 值, 表中加粗数据是相同条件下的最优结果。

首先, 比较的是各实验条件的综合影响。Q 矩阵错误设定的比例、项目质量、样本量以及属性分布对于 GDI、Wald-IC、Hull (HullP、HullR)以及 Wald-XPB 方法在 QRR、TPR、TNR 指标上的表现有明显影响。除 Hull (HullP、HullR)方法的 TPR 指标受项目质量的影响较小外, 在 Q 矩阵错误设定比例小、项目质量高、样本量大以及属性均匀分布水平下, 这些 Q 矩阵修正方法的表现优于其他水平。就 JK 因素而言, JK 对于 GDI、Wald-IC 和 Wald-XPB 在 QRR 指标上的表现, 以及所有的修正方法在 TNR 指标上的表现影响明显; 以上指标在 JK=8 水平下的结果优于 JK=4。

其次, 比较的是四种修正方法的综合表现。所有方法在 QRR 以及 TPR 指标上没有表现出明显优劣; 本研究中新提出的 Wald-XPB 在 TNR 指标上的表现明显优于其他方法; GDI 在 OS 指标上的表现较优, 但是在 US 指标上表现相对较差; HullR 在 OS 指标上的表现较差, 但是在 US 指标上表现相对较优; Wald-IC 在 US 指标上表现相对较差。

根据以上综合比较可知, Wald-XPB 以及 HullP 在各个指标上有相对较好的表现; 另外, 鉴于 Wald-XPB 是在 Wald-IC 基础上新提出的方法, 因此, 有必要对于 Wald-XPB、Wald-IC 以及 HullP 方法在 QRR、TPR 以及 TNR 这三个主要指标上的表现进行更详细的分析。

4.2.2 Wald-XPB、Wald-IC 与 HullP 方法的具体比较

图 3 中呈现的是 Wald-XPB、Wald-IC 以及 HullP 方法在 48 种具体的模拟条件下的 QRR 指标的值。可以发现, 项目质量对于这三种方法的表现影响最为明显, 随着项目质量的提高, QRR 的值也在提高。另外, 样本量、Q 矩阵错误设定的比例以及属性分布对于这三个方法在 QRR 指标上的表现稍有影响, 且趋势一致。就 QRR 指标而言, Wald-XPB、Wald-IC 以及 HullP 方法的表现仅有细微差异, 即当 $IQ=0.4$ 时 Wald-XPB 的表现略微低于另外两种方法。

图 4 中呈现的是以上三种方法在 TPR 指标上的表现。可以发现，在所有条件下 Wald-IC 以及 HullIP 方法均能获得较高的 TPR 值。项目质量对于 Wald-XPD 方法的表现有一定的影响：项目质量较低时，Wald-XPD 在 TPR 指标上的表现不如 Wald-IC 以及 HullIP 方法；随着项目质量的提高，三种方法在 TPR 指标上的表现相当。

图 5 呈现的是三种方法在 TNR 指标上的表现，由图可知 Wald-XPD 方法的 TNR 指标在所有条件下均优于 HullIP 和 Wald-IC 方法。随着 Q 矩阵错误设定比例降低、项目质量提高、测验长度增加，这三种方法的 TNR 的值也在升高。其中，项目质量对于各个方法在 TNR 指标上的表现影响最大，例如，项目质量低时 HullIP 方法的 TNR 值在 JK=4 水平下明显低于另外两种方法。均匀分布下，各个方法在 TNR 指标上的表现均优于高阶分布。”

意见 9：图 3 到图 5 中的变量标签建议使用直接使用中文，而不使用英文；如 IQ 直接写成题目质量，QM 直接写成 Q 误设率。

回应：感谢审稿专家的宝贵意见。已将图 3 到图 5 中的变量标签改成中文。修改稿中的图如下(第 17 页)：

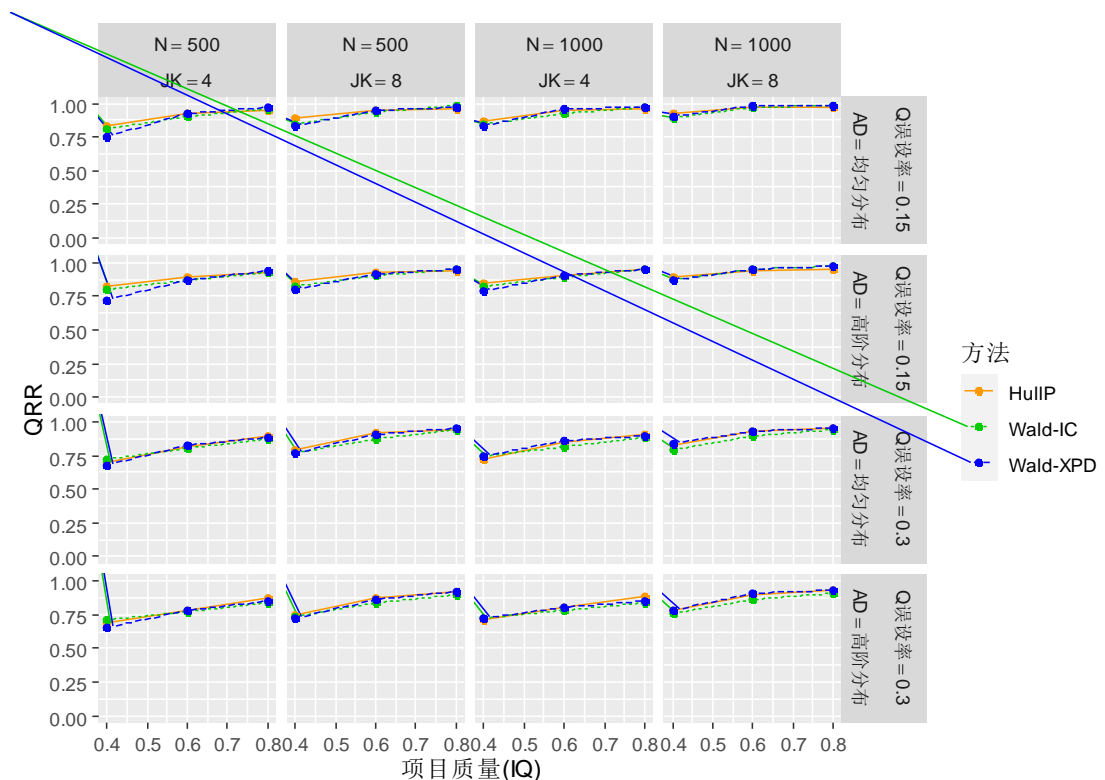


图 3 Wald-XPD、Wald-IC 与 HullIP 方法在 QRR 指标上的表现

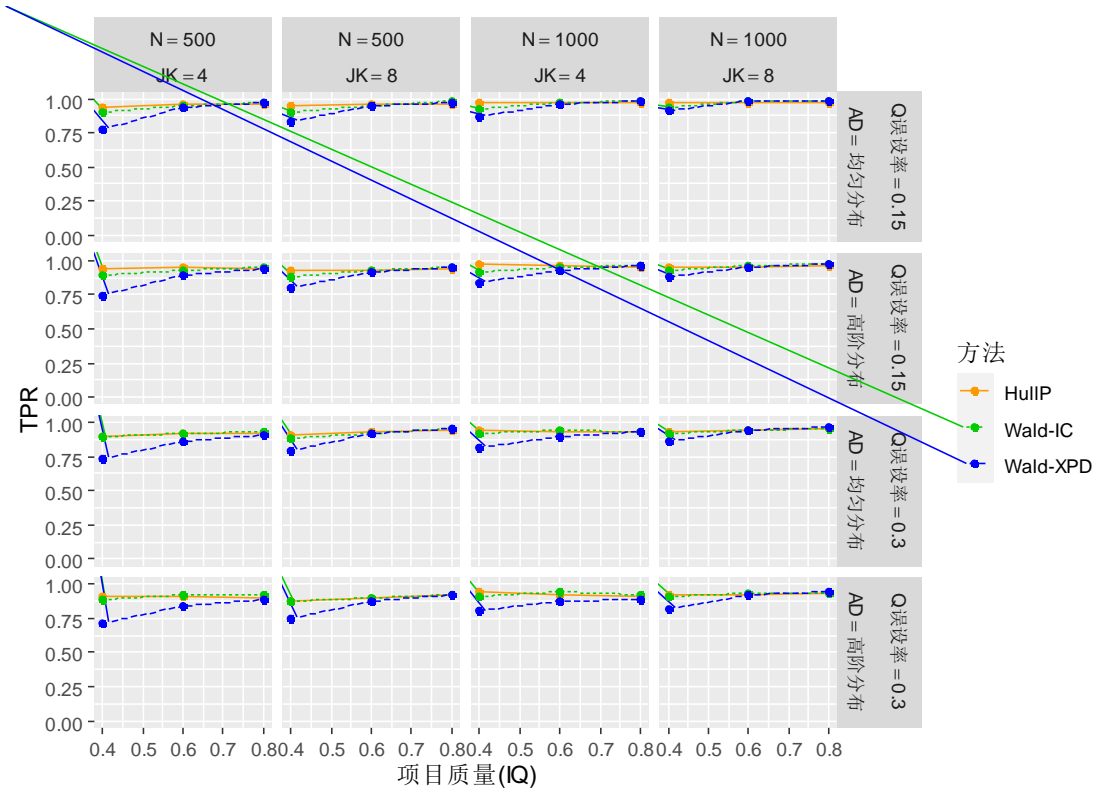


图4 Wald-XPDP、Wald-IC 与 HullIP 方法在 TPR 指标上的表现

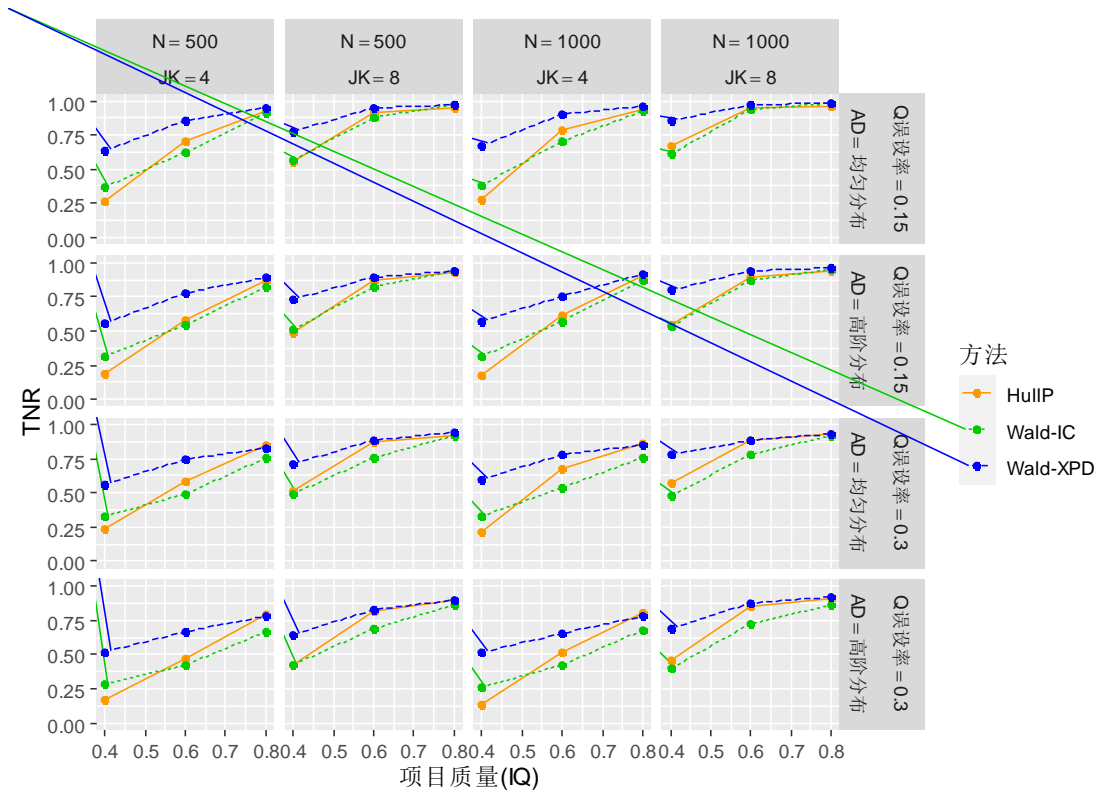


图5 Wald-XPDP、Wald-IC 与 HullIP 方法在 TNR 指标上的表现

意见 10:实证研究中,作者只考察了 Wald-XPDP 的表现,由于 Wald-XPDP 是 Wald-IC 的拓展,因此审稿人觉得可以增加 Wald-IC 的修正结果,然后再比较 Wald-XPDP 和 Wald-IC 两种方法的结果。

回应:感谢审稿专家的意见。我们在“5 实证数据分析”加入了 Wald-IC 的修正结果,计算了拟合指标值,与 Wald-XPDP 方法进行了比较。修改稿中内容如下(第 20 页):

“从拟合指标的角度来看,修正前 Q 矩阵的拟合指标为: $AIC=4979.26$ 、 $BIC=5245.28$; Wald-IC 方法修正后 Q 矩阵的拟合指标为: $AIC=4964.20$ 、 $BIC=5171.11$; Wald-XPDP 方法修正后 Q 矩阵的拟合指标为: $AIC=4962.48$ 、 $BIC=5152.5$ 。两种方法修正后 Q 矩阵的 AIC 和 BIC 指标均明显减小,其中, Wald-XPDP 方法修正后的 Q 矩阵获得更小的拟合指标,说明采用 Wald-XPDP 方法修正后的 Q 矩阵与数据的拟合更优。”

意见 11:研究的讨论部分不够深入,只是对结果的简单描述。

回应:感谢审稿专家的建设性意见。已对本研究的“6.1 结论与讨论”部分进行修改(第 20 页):

“6.1 结论与讨论

CDM 依赖正确设定的 Q 矩阵以获得准确的属性剖面分类(Rupp & Templin, 2008)。以往研究者提出的 GDI、Wald-IC、Hull 方法在多数的应用情景中虽然有较好的表现,但这些方法对 Q 矩阵中错误标定的属性不够敏感。本研究提出使用完整的 XPDP 矩阵计算用于 Q 矩阵修正的方法,即 Wald-XPDP 方法,并系统探讨了不同样本量、测验长度、Q 矩阵错误设定比例、属性分布等因素对 Q 矩阵修正结果的影响。采用实证数据展示了新提出的 Wald-XPDP 方法在实际应用中的表现与价值。

本研究结果表明:(1)整体而言, Wald-XPDP 方法的表现优于 GDI、Wald-IC、Hull 方法。Wald-XPDP 方法能够弥补 GDI、Wald-IC、Hull 方法在一些条件下对于错误标定属性不敏感的问题,且在 Q 矩阵恢复率和保留正确标定属性的比例方面也有较好的表现。(2)GDI、Wald-IC、Hull 和 Wald-XPDP 方法随着项目质量的提高、样本量增大、测验长度增加以及 Q 矩阵错误设定比例的降低,在 Q 矩阵修正中有更好的表现。当属性掌握模式来源于均匀分布时,这些方法在 Q 矩阵修正中的表现优于高阶分布。

需要特别指出的是,尽管 Wald-XPDP 方法的整体上的表现较优,但是 Wald-IC、Hull 以及 Wald-XPDP 方法在 TPR、TNR 指标上表现不同。在 TPR 指标上, Wald-XPDP 受项目质量低的影响明显;在 TNR 指标上, Wald-IC 和 Hull 受项目质量低以及测验长度短的因素的影响明显。TPR 指标数值低,说明 Q 矩阵修正方法倾向于修改正确设定的属性, TNR 数值

低则说明 Q 矩阵修正方法修改错误设定属性的能力弱。综合 TPR、TNR 这两个指标，可以发现尽管 Wald-XPB 方法在项目质量差的条件下能够较为有效地修正错误设定的属性，但是也可能存在过度修改正确设定属性的问题。也就是说，Wald-XPB 方法尽管提高了 Q 矩阵修正的表现，但是在项目质量较低的情境下，有可能会错误地修正了正确设定的 q 元素。Wald-IC 以及 HullP 尽管在项目质量差的条件下几乎不存在过度修改正确设定属性的问题，但却存在无法有效修正错误设定属性的问题，尤其是 HullP 方法。所以，本研究建议使用 Q 矩阵修正方法时，需要注意项目质量。如果发现项目质量较低，可以结合多种修正方法、参考专家意见进而获得准确的 Q 矩阵。

本研究采用采用 C++ 语言编写 XPB 矩阵，在一定程度上能够提高 Q 矩阵修正的速度，但是，由于 Wald-XPB 方法考虑模型中的全部参数且采用迭代的方式进行，在一些条件下可能存在耗时较长的问题。譬如，Wald-XPB 方法最短的平均用时是 12.50s，最长的平均时间需要 746.01s。Wald-XPB 方法在各个模拟条件下的平均运行时间参见附录中的表 A1。”

.....

审稿人 2 意见：

Q 矩阵修正是一个重要的研究问题。本文针对该问题，借鉴 Wald-IC 方法的思想，将该方法中不完整信息量替换为完整信息量，进行 Q 矩阵修订，具有一定的新意和应用价值。文稿行文流畅，方法得当。下面就几个表述问题，提出修改意见。对文中表述不准确或有误的地方，请作者检查和修改。

回应：感谢审稿专家对本文选题、创新性以及应用价值的肯定，以及在审稿过程中所付出的时间与精力。以下是对审稿专家意见的回应。稿件正文中的修改部分用蓝色字体标注。

意见 1：2.1 部分 $W(\alpha_i)$ 表示拥有属性掌握模式为 α_i 的被试的后验概率，是对题目 j 而言，被试群体中具有 α_i 的被试反应计算而来的吗？建议给出计算公式

回应：感谢审稿专家的宝贵意见。为了便于理解，根据审稿专家的意见，我们将初稿中的 $w(\alpha_i)$ 改为了 $w(\alpha_i|\mathbf{x})$ ，并在修改稿中增加了以下内容(第 5 页)：

“ $w(\alpha_i|\mathbf{x})$ 表示在测验项目的观察反应矩阵 \mathbf{x} 中属性掌握模式为 α_i 的被试的后验概率：

$$w(\alpha_l | \mathbf{x}) = \sum_{i=1}^N w(\alpha_l | \mathbf{x}_i) = \frac{\sum_{i=1}^N L(\mathbf{x}_i | \alpha_l) \pi(\alpha_l)}{\sum_{l=1}^{2^K} L(\mathbf{x}_i | \alpha_l) \pi(\alpha_l)} \quad (3)$$

公式(3)中, N 表示样本量; $L(\mathbf{x}_i | \alpha_l)$ 表示属性掌握模式为 α_l 的被试 i 在所有项目上作答反应 \mathbf{x}_i 的条件似然函数; $\pi(\alpha_l)$ 表示拥有第 l 种属性掌握模式的被试在总体中所占的比例, 即第 l 个结构参数。”

意见 2: 2.2 部分“只有增加相关属性会增加这两个指标的值, 忽略增加不相关属性对他们产生的影响”表述不完整, 后者是会产生影响还是降低影响?

回应: 感谢审稿专家的意见。我们已对“2.2 Hull 方法”部分有歧义的表达进行了修改, 修改稿中内容如下(第 6 页):

“对于 Hull 方法的两个拟合指标而言, 添加项目中相关联的属性会显著增加拟合指标的值; 添加不关联的属性也会增加拟合指标的值, 但影响可能较小。故从拟合—简约相平衡的视角出发, 在 Hull 图中选择先使拟合指标显著增加, 然后使拟合指标平缓增加的候选 \mathbf{q} 向量作为正确设定的 \mathbf{q} 向量。”

意见 3: 全文公式没有编号;

回应: 感谢审稿专家的意见。已对全文的公式进行编号。

意见 4: 2.3 部分第二段、第三段和第四段, 需要重大修改。

原因是: 第一, 第一句不是说状态是否必须, 应该说哪个属性是否必须; 第二, 为什么 $R \times P_j(\alpha_l) = 0$, 就是不必须的? 第三, 第三个公式, 两个举证相乘, 顺序写反了; 第四, “ $P_j(\alpha_l)$ 表示拥有属性掌握模式为 α_l 的被试在项目 j 中正确作答概率向量”应该表达的是 “ $P_j(\alpha_l)$ 表示由每一属性掌握模式 $\alpha_l (l = 1, 2, \dots, 2^K)$ 在项目 j 中正确作答概率构成的向量”; 第五, 句末是 $K-1$ 还是 K^*-1 ? ; 第六, 第四段第一句, $R \times P_j(\alpha_l) = 0$ 没有体现出于属性 k 的相关性, 怎么检验属性 k 的必然性呢? 因为 $P_j(\alpha_l)$ 是一个固定的向量。

回应: 感谢审稿专家的宝贵意见。根据专家以上 6 条意见, 已对“2.3 Wald-IC 方法”部分进行重大修改(第 7 页):

“2.3 Wald-IC 方法

用于 \mathbf{Q} 矩阵修正的 Wald 统计量也是在项目水平上进行的。其基本原理是：假设项目 j 所对应的 \mathbf{q} 向量定义了 2 个及以上的属性，如果将某一属性从 \mathbf{q} 向量中移除而没有导致模型—数据拟合变差，那么这个属性就不是必需的。为便于理解，现举例说明。假设一个测验共测量了 2 个属性，即 $K=2$ 。那么，所有可能的属性掌握模式有 4 种，可以表示为：

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

假定待检验的 \mathbf{q} 向量为 $\mathbf{q}=[1,1]$ ，那么相应的项目正确作答概率的向量可以表示为：

$$\mathbf{p}_j(\mathbf{a}) = \begin{bmatrix} p_j(\mathbf{a}_1) \\ p_j(\mathbf{a}_2) \\ p_j(\mathbf{a}_3) \\ p_j(\mathbf{a}_4) \end{bmatrix} = \begin{bmatrix} \delta_{j0} \\ \delta_{j0} + \delta_{j1} \\ \delta_{j0} + \delta_{j2} \\ \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j12} \end{bmatrix}$$

检验属性 1(即 α_1)是否是必需的，首先需要构建 α_1 的 \mathbf{R} 矩阵：

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

本文中，“ \times ”用于矩阵或向量时，表示矩阵相乘。若 α_1 在统计上不是必需的，那么 $\mathbf{R} \times \mathbf{p}_j(\mathbf{a}) = \mathbf{0}$ 。

即：

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} p_j(\mathbf{a}_1) \\ p_j(\mathbf{a}_2) \\ p_j(\mathbf{a}_3) \\ p_j(\mathbf{a}_4) \end{bmatrix} = \begin{bmatrix} -\delta_{j1} \\ -(\delta_{j1} + \delta_{j12}) \end{bmatrix} = \mathbf{0}$$

表明掌握 α_1 不会增加正确答对项目 j 的概率，故 α_1 不是必需的。此外，检验属性 2 (即 α_2)

是否必需的 \mathbf{R} 矩阵为：

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

需要说明的是，对于项目 j 而言，不同的待检验属性所对应的 \mathbf{q} 向量是不同，也就是项目参数估计值是不同的，因此，向量 $\mathbf{p}_j(\mathbf{a})$ 的值不是固定的。

Wald-IC 统计量的形式为：

$$\text{Wald}_{(\text{IC})} = [\mathbf{R} \times \mathbf{p}_j(\boldsymbol{\alpha})]' (\mathbf{R} \times \mathbf{V}_{(\text{IC})j} \times \mathbf{R}')^{-1} [\mathbf{R} \times \mathbf{p}_j(\boldsymbol{\alpha})] \quad (6)$$

其中， $\mathbf{V}_{(\text{IC})j}$ 是基于不完整信息矩阵计算的项目 j 正确作答概率的方差—协方差矩阵。

Wald-IC 方法修正 \mathbf{Q} 矩阵的步骤为：首先，需要构建一个 $2^{K_j^*} \times 2^{K_j^*}$ 的 \mathbf{R} 矩阵， K_j^* 表示待检验的 \mathbf{q} 向量中定义的项目 j 需要的属性数量。在零假设下，即属性 k 在统计上不是必需的，那么 $\mathbf{R} \times \mathbf{p}_j(\boldsymbol{\alpha}) = \mathbf{0}$ 。其次，需要对不完整信息矩阵求逆获得项目正确作答概率的方差—协方差矩阵 $\mathbf{V}_{(\text{IC})j}$ 来构建 Wald 统计量。Ma 和 de la Torre (2020) 采用的是 de la Torre (2009) 提出的考虑全部项目正确作答概率的不完整信息矩阵估计方法 \mathbf{I}_{D09} ：

$$\mathbf{I}_{D09} = \frac{\partial \mathbf{l}(\mathbf{x})}{\partial [\mathbf{p}_j(\boldsymbol{\alpha})]} \times \frac{\partial \mathbf{l}(\mathbf{x})}{\partial [\mathbf{p}_j(\boldsymbol{\alpha})]'} \quad (7)$$

其中， $\mathbf{l}(\mathbf{x})$ 表示观察数据的对数似然函数。理论上，用于 \mathbf{Q} 矩阵修正的 Wald 统计量渐近 χ^2 分布，自由度是 $2^{K_j^*} - 1$ 。但是 Wald-IC 统计量中方差—协方差矩阵的计算存在不准确的问题，因此在 \mathbf{Q} 矩阵修正时可能导致效果不理想。”

意见 5： 3.1 部分第二段，应该补充完整信息量有哪些方法，进而介绍 XPD 是怎么算的。另外，完整信息量矩阵的作用是什么，在此处采用信息矩阵能解决的问题的依据是什么？应补充说明

回应：感谢审稿专家的意见。首先，对“3.1 使用 XPD 矩阵构建 Wald-XPB 统计量”部分进行修改，补充了以往研究者提出的完整信息矩阵的类型，然后介绍了 XPD 矩阵的计算公式。修改稿中内容如下(第 9 页)：

“研究者们提出了很多完整信息矩阵的估计方法，主要有：完整的经验交叉相乘信息矩阵(Liu et al., 2021; Philipp et al., 2018; 刘彦楼 等, 2016)、完整的观察信息矩阵(observed information matrix, Obs; Liu et al., 2021; 刘彦楼 等, 2016)、完整的三明治信息矩阵(sandwich-type information matrix, Sw; Liu, Xin et al., 2019; Liu et al., 2021)等。由于考虑所有模型参数，完整信息矩阵的计算量较大，尤其是观察信息矩阵以及三明治信息矩阵涉及观察数据的对数似然函数关于所有模型参数的二阶偏导，计算量非常大。鉴于 \mathbf{Q} 矩阵修正过程中的运算速度问题，本文采用观察数据对数似然函数关于项目参数 $\boldsymbol{\delta}$ 和结构参数 $\boldsymbol{\pi}$ 的一阶导向量交叉相乘而

计算的 XPD 矩阵:

$$I_{XPD} = \begin{bmatrix} \frac{\partial l(\mathbf{x})}{\partial \delta_1} \times \frac{\partial l(\mathbf{x})}{\partial \delta_1} & K & \frac{\partial l(\mathbf{x})}{\partial \delta_1} \times \frac{\partial l(\mathbf{x})}{\partial \pi_{L-1}} \\ M & O & M \\ \frac{\partial l(\mathbf{x})}{\partial \pi_{L-1}} \times \frac{\partial l(\mathbf{x})}{\partial \delta_1} & L & \frac{\partial l(\mathbf{x})}{\partial \pi_{L-1}} \times \frac{\partial l(\mathbf{x})}{\partial \pi_{L-1}} \end{bmatrix} \quad (8)$$

其次,在“引言”部分增加了对于问题提出以及完整信息矩阵的作用的论述。修改稿中内容如下(第3页):

“研究者(Na jera et al., 2021a)通过模拟研究比较了 GDI、Wald-IC 以及 Hull 方法,发现在大多数条件下 Hull 的表现最好、Wald-IC 的表现稍逊于 Hull。但是, Hull 和 Wald-IC 在正确修正错误标定的属性方面的表现较差,尤其是 Q 矩阵中存在较多错误设定时。可以发现, Wald-IC 统计量是使用不完整信息矩阵计算的。先前研究表明,采用不完整信息矩阵构建的统计量在后续研究中会导致一些错误,如低估模型参数标准误(Philipp et al., 2018)、用于项目功能差异检验及项目水平模型比较时导致一类错误控制率膨胀(Liu, Andersson et al., 2019; Liu, Yin et al., 2019; 刘彦楼 等, 2016)等。基于此,本研究认为 Wald-IC 方法在正确修正错误标定属性方面表现较差的主要原因可能是在 Wald 统计量的计算中采用了不完整的信息矩阵。

研究者(Liu et al., 2016; Liu, Xin et al., 2019; Liu et al., 2021; Philipp et al., 2018; 刘彦楼 等, 2016)认为 CDM 中同时存在两种类型的模型参数:项目参数和结构参数。不完整信息矩阵(de la Torre, 2009, 2011)忽略了结构参数。尽管这类方法计算量较小,但是在计算方差—协方差矩阵时有较大可能会造成不可忽略的错误,由此可能导致 Q 矩阵修正结果出现问题。以往研究者提出了多种完整信息矩阵估计方法(Liu et al., 2016; Liu et al., 2019; Liu et al., 2021; Philipp et al., 2018; 刘彦楼 等, 2016),但是这些关于模型参数的信息矩阵无法直接用于 Q 矩阵修正中 Wald 统计量的计算,因为此类 Wald 统计量中使用的是关于模型参数的方差—协方差矩阵。此外,与其他完整信息矩阵相比,经验交叉相乘信息矩阵(empirical cross-product information matrix, XPD; Liu et al., 2021; Philipp et al., 2018; 刘彦楼 等, 2016)计算量较小,故本研究在包含全部模型参数的 XPD 矩阵的基础上,经过转换获得关于项目正确作答概率的方差—协方差矩阵,以此构建用于 Q 矩阵修正的 Wald 统计量(记为 Wald-XPD)。”

意见 6: 3.1 部分第三段(2)的这句话,如何解释?怎么从参数的方差协方差矩阵就能转换

为作答概率的方差协方差矩阵？

回应：感谢审稿专家的宝贵意见。为了便于理解，对“3.1 使用 XPD 矩阵构建 Wald-XPD 统计量”部分的 \mathbf{M}_j 矩阵进行了详细说明。修改稿中的内容如下(第 9 页)：

“采用 \mathbf{M}_j 矩阵(de la Torre, 2011)通过矩阵乘法将项目参数的方差—协方差矩阵 Σ_j ，转换为项目正确作答概率的方差—协方差矩阵 $\mathbf{V}_{(\text{XPD})_j}$ ，即： $\mathbf{V}_{(\text{XPD})_j} = \mathbf{M}_j \times \Sigma_j$ 。 \mathbf{M}_j 矩阵是 $2^{K_j^*} \times 2^{K_j^*}$ 维的矩阵，表示项目 j 中各个属性掌握模式与项目参数之间的对应关系，可以将项目参数转换为各个属性掌握模式下的正确作答概率。例如，假设项目 j 中 $K_j^*=2$ ，则对于饱和 G-DINA 模型而言该项目的 \mathbf{M}_j 矩阵可以表示为：

$$\mathbf{M}_{j[4 \times 4]} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

通过 \mathbf{M}_j 矩阵，可以获得该项目中各个属性掌握模式下的正确作答概率向量 $\mathbf{p}_j(\boldsymbol{\alpha})$ ：

$$\mathbf{p}_j(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} \delta_{j0} \\ \delta_{j1} \\ \delta_{j2} \\ \delta_{j12} \end{bmatrix} = \begin{bmatrix} \delta_{j0} \\ \delta_{j0} + \delta_{j1} \\ \delta_{j0} + \delta_{j2} \\ \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j12} \end{bmatrix}$$

根据统计学中模型参数方差—协方差矩阵的性质(或参考 Li & Wang, 2015)，可以通过 \mathbf{M}_j 矩阵将项目参数的方差—协方差矩阵 Σ_j 转换为项目正确作答概率的方差—协方差矩阵 $\mathbf{V}_{(\text{XPD})_j}$ 。”

意见 7： 3.1 部分倒数第二个公式，代表的矩阵乘法怎么乘，表达式不符合矩阵乘积的计算规则？

回应：非常感谢审稿专家的指出本文表达存在的问题。已经对这个公式进行修改，修改稿中内容如下(第 10 页)：

$$\mathbf{p}_j(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} \delta_{j0} \\ \delta_{j1} \\ \delta_{j2} \\ \delta_{j12} \end{bmatrix} = \begin{bmatrix} \delta_{j0} \\ \delta_{j0} + \delta_{j1} \\ \delta_{j0} + \delta_{j2} \\ \delta_{j0} + \delta_{j1} + \delta_{j2} + \delta_{j12} \end{bmatrix}$$

意见 8: 3.2 部分步骤三, 需要重大修改。我的理解是, 修正的过程中是添加、移除循环进行, 添加了一个属性就要对该属性的必要性进行检验, 必要的条件在对其他属性是否有必要删除进行检验。这里需要指明检验统计量 Wald 的判断标准。对这部分表述和图 2 需要进行修订。图 2 中应体现 Wald 统计量的判断过程, 还有原图缺少了一个箭头和一个“否”字。

回应: 感谢审稿专家专业、严谨、详细的意见。已对“3.2 Wald-XPB 方法的具体实施步骤”部分 Wald-XPB 方法的具体实施步骤以及图 2 进行修改。修改稿中的内容如下(第 11 页):

“步骤(3): 更新集合 **A**、**B**。选出具有较大 PVAF 的 **q** 向量进行修正, 将该 **q** 向量中各属性使用 Wald-XPB 统计量进行显著性检验, 确定该 **q** 向量对应的集合 **A** 和集合 **B** 中的属性是否应该被移除或添加; 然后判断 **q** 向量的 PVAF 是否大于 0.95, 大于 0.95 说明这个 **q** 向量是合适的, 停止修正。否则继续修正。

步骤(4): 重复步骤(3), 直到某个 **q** 向量的 PAVF 值大于 0.95, 或者没有属性被移除或添加则停止修正。”

修改后的图 2 如下所示(第 12 页):

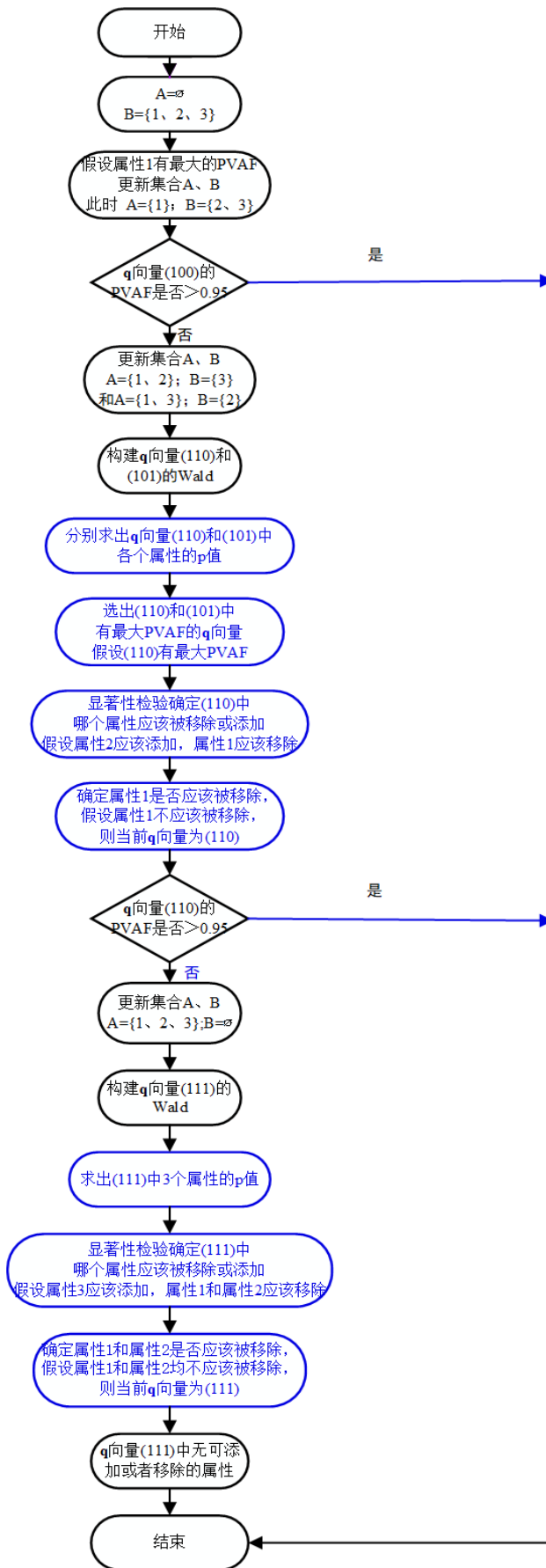


图 2 Wald-XPD 方法用于 q_i 向量的修正流程图

意见 9: Q 矩阵是循环修正还是一次修正?

回应: 感谢审稿专家的意见。为了避免误解, 首先对稿件“3.2 Wald-XPB 方法的具体实施步骤”进行了如下修改(第 10 页):

“步骤(3): 更新集合 A、B。选出具有较大 PVAF 的 q 向量进行修正, 将该 q 向量中各属性使用 Wald-XPB 统计量进行显著性检验, 确定该 q 向量对应的集合 A 和集合 B 中的属性是否应该被移除或添加; 然后判断 q 向量的 PVAF 是否大于 0.95, 大于 0.95 说明这个 q 向量是合适的, 停止修正。否则继续修正。

步骤(4): 重复步骤(3), 直到某个 q 向量的 PAVF 值大于 0.95, 或者没有属性被移除或添加则停止修正。

步骤(5): 在单个项目修正结束后, 重新计算 PVAF 以及 Wald-XPB 统计量, 直到达到最大迭代或者某次迭代结束后的 q 向量与前一次迭代的 q 向量完全相等则停止修正。”

其次, 在“6.1 结论与讨论”部分对 Wald-XPB 方法进行了说明。修改稿中内容如下(第 21 页):

“但是, 由于 Wald-XPB 方法考虑模型中的全部参数且采用迭代的方式进行, 在一些条件下可能存在耗时较长的问题。”

意见 10: 结果没有在 QRP,TPR 没有表现出优势, 方法之间的结果也不具有一致性, 希望能对该方法的优势和各个方法的不一致性进行解释和说明

回应: 感谢审稿专家的意见。我们在“6.1 结论与讨论”部分对 Wald-XPB、Wald-IC 以及 Hull 方法之间的不一致性问题进行了解释和说明。修改稿中内容如下(第 21 页):

“需要特别指出的是, 尽管 Wald-XPB 方法的整体上的表现较优, 但是 Wald-IC、HullP 以及 Wald-XPB 方法在 TPR、TNR 指标上表现不同。在 TPR 指标上, Wald-XPB 受项目质量低的影响明显; 在 TNR 指标上, Wald-IC 和 HullP 受项目质量低以及测验长度短的因素的影响明显。TPR 指标数值低, 说明 Q 矩阵修正方法倾向于修改正确设定的属性, TNR 数值低则说明 Q 矩阵修正方法修改错误设定属性的能力弱。综合 TPR、TNR 这两个指标, 可以发现尽管 Wald-XPB 方法在项目质量差的条件下能够较为有效地修正错误设定的属性, 但是也可能存在过度修改正确设定属性的问题。也就是说, Wald-XPB 方法尽管提高了 Q 矩阵修正的表现, 但是在项目质量较低的情境下, 有可能会错误地修正了正确设定的 q 元素。Wald-IC 以及 HullP 尽管在项目质量差的条件下几乎不存在过度修改正确设定属性的问题, 但却存在无法有效修正错误设定属性的问题, 尤其是 HullP 方法。所以, 本研究建议使用 Q 矩阵修正

方法时，需要注意项目质量。如果发现项目质量较低，可以结合多种修正方法、参考专家意见进而获得准确的 Q 矩阵。”

意见 11：该方法运用完整信息量，计算复杂，建议补充运行时间。

回应：感谢审稿专家的意见。我们在“6.1 结论与讨论”部分介绍了 Wald-XPB 方法的运行最长时间与最短时间，并在附录中呈现了 Wald-XPB 方法在 48 种模拟条件下的运行时间。修改稿中内容如下(第 21 页)：

“但是，由于 Wald-XPB 方法考虑模型中的全部参数且采用迭代的方式进行，在一些条件下可能存在耗时较长的问题。譬如，Wald-XPB 方法最短的平均用时是 12.50s，最长的平均时间需要 746.01s。Wald-XPB 方法在各个模拟条件下的平均运行时间参见附录中的表 A1。”

附录：Wald-XPB 方法在各模拟条件下的平均运行时间

表 A1 Wald-XPB 方法在各模拟条件下的平均运行时间

模拟条件	AD	QM	IQ	N	JK	时间
1	均匀分布	0.15	0.4	500	4	476.16
2	高阶分布	0.15	0.4	500	4	195.68
3	均匀分布	0.15	0.4	500	8	706.40
4	高阶分布	0.15	0.4	500	8	654.93
5	均匀分布	0.15	0.4	1000	4	302.90
6	高阶分布	0.15	0.4	1000	4	746.01*
7	均匀分布	0.15	0.4	1000	8	505.79
8	高阶分布	0.15	0.4	1000	8	320.67
9	均匀分布	0.15	0.6	500	4	68.17
10	高阶分布	0.15	0.6	500	4	67.66
11	均匀分布	0.15	0.6	500	8	54.11
12	高阶分布	0.15	0.6	500	8	81.36
13	均匀分布	0.15	0.6	1000	4	21.35
14	高阶分布	0.15	0.6	1000	4	90.22
15	均匀分布	0.15	0.6	1000	8	56.11
16	高阶分布	0.15	0.6	1000	8	113.40
17	均匀分布	0.15	0.8	500	4	12.93
18	高阶分布	0.15	0.8	500	4	21.20
19	均匀分布	0.15	0.8	500	8	23.63
20	高阶分布	0.15	0.8	500	8	46.23
21	均匀分布	0.15	0.8	1000	4	12.97
22	高阶分布	0.15	0.8	1000	4	12.50#
23	均匀分布	0.15	0.8	1000	8	48.36
24	高阶分布	0.15	0.8	1000	8	32.42
25	均匀分布	0.3	0.4	500	4	114.85
26	高阶分布	0.3	0.4	500	4	223.68
27	均匀分布	0.3	0.4	500	8	750.26
28	高阶分布	0.3	0.4	500	8	310.86
29	均匀分布	0.3	0.4	1000	4	163.41

30	高阶分布	0.3	0.4	1000	4	226.47
31	均匀分布	0.3	0.4	1000	8	510.00
32	高阶分布	0.3	0.4	1000	8	696.73
33	均匀分布	0.3	0.6	500	4	63.20
34	高阶分布	0.3	0.6	500	4	111.59
35	均匀分布	0.3	0.6	500	8	64.36
36	高阶分布	0.3	0.6	500	8	111.91
37	均匀分布	0.3	0.6	1000	4	61.61
38	高阶分布	0.3	0.6	1000	4	77.84
39	均匀分布	0.3	0.6	1000	8	95.48
40	高阶分布	0.3	0.6	1000	8	152.57
41	均匀分布	0.3	0.8	500	4	45.05
42	高阶分布	0.3	0.8	500	4	12.75
43	均匀分布	0.3	0.8	500	8	22.22
44	高阶分布	0.3	0.8	500	8	72.12
45	均匀分布	0.3	0.8	1000	4	15.34
46	高阶分布	0.3	0.8	1000	4	25.56
47	均匀分布	0.3	0.8	1000	8	54.40
48	高阶分布	0.3	0.8	1000	8	190.39

注：*为 Wald-XPB 方法在模拟条件下的最长运行时间，#为 Wald-XPB 方法在模拟条件下的最短运行时间

第二轮

审稿人 1 意见：

经过前一轮的修改，《认知诊断模型 Q 矩阵修正：完整信息矩阵的作用》的质量有了较大的提升，作者较好地回答了提出的问题。不过还存在以下几个问题：

回应：感谢审稿专家的肯定，以及在审稿过程中所付出的时间与精力。以下是对审稿专家意见的回应。稿件正文中的修改部分用**橙色字体**标注。

意见 1：摘要部分“新开发的 Wald-XPB 方法在各个指标下...”中的“各个指标”指代不是很明晰，建议将主要指标的具体名字给予呈现。

回应：感谢审稿专家的意见。修改稿中已对“摘要”部分“各个指标”用具体名字给予呈现。修改稿中内容如下(第 1 页)：

“发现：新开发的 Wald-XPB 方法在**Q 矩阵恢复率、保留正确标定属性的比例以及修正错误标定属性的比例**这三个主要指标上均有较好的表现，且整体上优于其他方法，……。”

意见 2：引言第一段中的“作为新一代心理测量理论，认知诊断模型...”，理论是理论，模型是模型，此处直接把模型等同于理论，这是不合理的。

回应：感谢审稿专家细致严谨的意见。修改稿中已对“引言”部分第一段进行修改。修改稿中内容如下(第 1 页)：

“作为新一代心理测量理论，认知诊断(cognitive diagnosis)的主要目的是提供关于被试的多维、细粒度潜在特质(如知识、认识过程、技能、策略、人格特质或心理障碍等，统称为属性)的诊断性评价信息；认知诊断模型(cognitive diagnostic model, CDM)是研究者为了实现以上主要目的而提出的一类离散潜变量模型(Rupp et al., 2010)。”

意见 3：引言第二段和第三段之间的衔接不够紧密，第二段结尾部分讲 Q 矩阵误设情况，第三段开头直接讲 Q 矩阵修正方法；这二者之间缺乏相应的过渡语句。

回应：感谢审稿专家的意见。修改稿中已增加“引言”部分第二段和第三段之间的衔接语句。修改稿中内容如下(第 1 页)：

“Q 矩阵的正确性关系到认知诊断结果的准确性，如何修正原始 Q 矩阵中可能存在的错误设定是研究者面临的一个重要理论及现实问题。为了获得正确设定的 Q 矩阵，国内外研究者提出了很多 Q 矩阵修正的方法(李佳 等, 2021)。”

意见 4：公式 (1) 中的最高交互项参数和下一行的文字解释中的最高交互项参数不一致性 (字母 K 后面的下标 j 是否需要)，需仔细检查。此外，公式 (1) 当中，作者使用的是完整 q 向量中的元素，没有用缩减 q 向量 (reduced q vector)，这个时候，作者应该说明，对于题目 j 未考察的到的属性，其 delta 参数为 0

回应：感谢审稿专家的意见。修改稿中已对公式(1)进行仔细检查和修改。修改稿中内容如下(第 5 页)：

“G-DINA 模型是一般、饱和的 CDM 模型，对其进行适当约束，可以获得多种特殊模型(de la Torre, 2011)。令 α_l 表示第 l 种属性掌握模式， $\mathbf{q}_j = [q_{j1}, \mathbf{K}, q_{jK}]$ 表示项目 j 与测验中 K 个属性之间的对应关系，在饱和的 G-DINA 模型中，正确答对项目 j 的概率可表示为：

$$p_j(\alpha_l) = p_j(\alpha_l, \mathbf{q}_j) = \delta_{j0} + \sum_{k=1}^K \delta_{jk} \alpha_{lk} q_{jk} + \sum_{k=1}^{K-1} \sum_{k'=k+1}^K \delta_{j,2,(k,k')} \alpha_{nk} \alpha_{nk'} q_{jk} q_{jk'} + L \quad (1)$$

其中， δ_{j0} 是项目 j 的截距项参数， δ_{jk} 是 α_{lk} 的主效应参数， $\delta_{j,2,(k,k')}$ 是 α_{lk} 与 $\alpha_{lk'}$ 之间的交互效应参数。需要特别说明的是，在公式错误！未找到引用源。中，如果 α_l 或 \mathbf{q}_j 中的某个元素等于 0，那么对应的项目参数 δ 也等于 0。”

意见 5: 模拟研究结果的三级标题表述上, 建议作者提炼一些更具结论性的语句作为标题, 而不是这种没有提供任何结果信息的标题。

回应: 感谢审稿专家的意见。修改稿中已对“4.2 研究结果”部分做出了进一步的修改, 并提炼出更具结论性的语句作为三级标题。修改稿中内容如下(第 15 页):

“4.2.1 GDI、Hull、Wald-IC 以及 Wald-XPD 在各因素不同水平上的表现

4.2.2 Wald-XPD 在修正错误标定属性时的表现”

意见 6: 评价指标方面, 需要陈述清楚如何看各个指标的结果, 哪些指标越大越好, 哪些指标越小越好。

回应: 感谢审稿专家的意见。修改稿中已在“4.1.3 评价指标”部分增加了如何区分各个指标的结果。修改稿中内容如下(第 14 页):

“以上五个指标从不同方面反映了 Q 矩阵的修正效果。其中, QRR、TPR、TNR 的值越高, 表示该修正方法的 Q 矩阵恢复率以及保留正确标定属性和修正错误标定属性的比例越高, 修正效果越好。OS 和 US 的值越小, 表示该修正方法存在较少过度设定和吝啬设定的趋势, 修正效果越好。”

意见 7: 讨论部分依旧显得单薄, 还需要加强讨论部分的撰写。如作者还可以讨论在 QRR、TPR 或 TNR 等方面, 为何样本量、Q 矩阵错误设定比例以及属性分布对 Wald-XPD、Wald-IC 和 HullIP 三者没什么影响?

回应: 感谢审稿专家的意见。修改稿中已在“6.1 结论与讨论”部分增加了本研究考虑的不同因素水平对于 GDI、Hull、Wald-IC、Wald-XPD 方法的影响。修改稿中内容如下(第 22 页):

“在本研究操纵的 5 种因素中, 项目质量对 GDI、Hull、Wald-IC、Wald-XPD 方法表现的影响较大; 样本量和测验长度也对四种修正方法的表现有一定的影响。出现这种现象的原因可能是, 项目质量越高、样本量越大以及测验长度越长, 被试观察作答反应矩阵中包含的关于 CDM 中未知参数的信息也就越多, 因此, 以上四种方法的表现也就越好。与以往研究类似(Kang et al., 2019; Ma & de la Torre, 2020; Na jera et al., 2021a), 本研究同样发现属性分布对于对 GDI、Hull、Wald-IC、Wald-XPD 方法在 TNR 指标上的表现有细微的影响。出现这种现象的原因可能是, 当属性均匀分布时所有可能属性掌握模式分布的概率是相等的, 也就是被试观察作答反应矩阵中包含的关于结构参数的信息是一样的; 当属性服从高阶分布时, 意味着属性之间存在一定的关联性, 使得某些属性掌握模式分布的概率可能会比较高, 另外

一些属性掌握模式分布的概率会比较低,即被试观察作答反应矩阵中包含的结构参数的信息量会比较少。因此导致了属性均匀分布时,四种方法在各个指标上的表现略优。 Q 矩阵错误设定的比例对 GDI、Wald-IC、Hull 方法表现的影响较大,随着 Q 矩阵错误设定比例的降低,它们能够获得更高的 QRR、TPR 和 TNR 值,这与已有研究结果一致(Ma & de la Torre, 2020; Na jera et al., 2021a); 对 Wald-XPd 方法表现的影响则相对较小。基于 Wald-XPd 在 TNR 指标上的表现,本文认为出现这种结果的原因可能是 Wald-XPd 在迭代结束前的循环中能够有效修正 Q 矩阵错误设定的元素,故最终结果表明其受 Q 矩阵错误设定比例的影响较小。”

意见 8: 建议压缩展望部分的内容,而增加讨论部分的内容。

回应: 感谢审稿专家的意见。修改稿中已增加了“6.1 结论与讨论”部分的内容,压缩了“6.2 研究展望”部分的内容。修改稿中内容如下(第 21-22 页):

“(3)由 HullP、Wald-IC 以及 Wald-XPd 方法的进一步比较的结果可知,三种方法在 Q 矩阵恢复率方面差异较小,HullP、Wald-IC 在保留正确标定的属性方面的表现略优于 Wald-XPd 方法,但在所有模拟条件下,Wald-XPd 方法在修正错误标定的属性方面的表现均优于另外两种方法。(4)实证数据分析的结果表明,Wald-XPd 方法修正后的 Q 矩阵与原始数据有最优的拟合度。

在本研究操纵的 5 种因素中,项目质量对 GDI、Hull、Wald-IC、Wald-XPd 方法表现的影响较大;样本量和测验长度也对四种修正方法的表现有一定的影响。出现这种现象的原因可能是,项目质量越高、样本量越大以及测验长度越长,被试观察作答反应矩阵中包含的关于 CDM 中未知参数的信息也就越多,因此,以上四种方法的表现也就越好。与以往研究类似(Kang et al., 2019; Ma & de la Torre, 2020; Na jera et al., 2021a),本研究同样发现属性分布对于对 GDI、Hull、Wald-IC、Wald-XPd 方法在 TNR 指标上的表现有细微的影响。出现这种现象的原因可能是,当属性均匀分布时所有可能属性掌握模式分布的概率是相等的,也就是被试观察作答反应矩阵中包含的关于结构参数的信息是一样的;当属性服从高阶分布时,意味着属性之间存在一定的关联性,使得某些属性掌握模式分布的概率可能会比较高,另外一些属性掌握模式分布的概率会比较低,即被试观察作答反应矩阵中包含的结构参数的信息量会比较少。因此导致了属性均匀分布时,四种方法在各个指标上的表现略优。 Q 矩阵错误设定的比例对 GDI、Wald-IC、Hull 方法表现的影响较大,随着 Q 矩阵错误设定比例的降低,它们能够获得更高的 QRR、TPR 和 TNR 值,这与已有研究结果一致(Ma & de la Torre, 2020; Na jera et al., 2021a); 对 Wald-XPd 方法表现的影响则相对较小。基于 Wald-XPd 在 TNR 指

标上的表现, 本文认为出现这种结果的原因可能是 Wald-XPDP 在迭代结束前的循环中能够有效修正 Q 矩阵错误设定的元素, 故最终结果表明其受 Q 矩阵错误设定比例的影响较小。”

.....

审稿人 2 意见:

作者针对评审专家的意见进行了非常细致认真的修改, 论文质量得到提升。建议针对以下问题, 对文稿进行进一步修改。

回应: 感谢审稿专家的肯定, 以及在审稿过程中所付出的时间与精力。以下是对审稿专家意见的回应。稿件正文中的修改部分用橙色字体标注。

意见 1: 从三种方法的结果来看, 新方法 Wald-XPDP 对 Q 矩阵的修正率和正确属性的保留率均低于已有的 Hull 方法。如果该方法的修正时间也高于 Hull 方法 (文章仅呈现了新方法的修正时间, 可以看出耗时较长), 那么该方法的优势在哪里? 请在文中适当地地方阐明。

回应: 感谢审稿专家的意见。根据专家的意见, 我们进行了以下修改:

首先, 在“4.2.2 Wald-XPDP 在修正错误标定属性时的表现”部分详细阐述了 Wald-XPDP 相对 Hull 方法的优势。修改稿中内容如下(第 17 页):

“4.2.2 Wald-XPDP 在修正错误标定属性时的表现

.....。

图 5 呈现的是三种方法在 TNR 指标上的表现。所有条件下, Wald-XPDP 方法在 TNR 指标上的表现均是最优的。对比 Wald-XPDP 方法在 TPR 及 TNR 上的综合表现, 可以发现低项目质量条件对这个方法产生了一些不利影响; 但在中等或高项目质量条件下, Wald-XPDP 不仅能有效保留 Q 矩阵中正确设定的元素, 而且也能有效修正 Q 矩阵中错误设定的元素。测验长度较短、项目质量较低及 Q 矩阵错误设定比例较高时 HullP 方法的表现较差; 对比同样条件下 HullP 在 TPR 指标上的表现可知, 尽管 HullP 方法在保留正确标定属性方面略微优于 Wald-XPDP, 但是它较多地保留了错误设定的属性。即, HullP 方法倾向于较少地修正原始 Q 矩阵中的元素。在低项目质量条件下的多数情景中, 尽管 Wald-IC 方法在 TNR 上的表现优于 HullP, 但是在随着项目质量的提高 HullP 在多数情景中的表现优于 Wald-IC。HullP、Wald-IC 以及 Wald-XPDP 方法在 TNR 指标上的表现受样本量、测验长度、项目质量、属性分布及错误设定比例的影响明显。随着 Q 矩阵错误设定比例降低、项目质量提高、测验长度增加, HullP 和 Wald-IC 方法的 TNR 值有所提高, 但仍低于 Wald-XPDP 方法的 TNR 值。”

其次，在“6.1 结论与讨论”部分增加了 Wald-XPDI 方法的优势。修改稿中内容如下(第 21-22 页):

“..... (3)由 HullIP、Wald-IC 以及 Wald-XPDI 方法的进一步比较的结果可知，三种方法在 Q 矩阵恢复率方面差异较小，HullIP、Wald-IC 在保留正确标定的属性方面的表现略优于 Wald-XPDI 方法，但在所有模拟条件下，Wald-XPDI 方法在修正错误标定的属性方面的表现均优于另外两种方法。(4)实证数据分析的结果表明，Wald-XPDI 方法修正后的 Q 矩阵与原始数据有最优的拟合度。

..... Q 矩阵错误设定的比例对 GDI、Wald-IC、Hull 方法表现的影响较大，.....；对 Wald-XPDI 方法表现的影响则相对较小。基于 Wald-XPDI 在 TNR 指标上的表现，本文认为出现这种结果的原因可能是 Wald-XPDI 在迭代结束前的循环中能够有效修正 Q 矩阵错误设定的元素，故最终结果表明其受 Q 矩阵错误设定比例的影响较小。”

意见 2：实证数据分析部分，该研究仅将新方法与已有的 Wald-IC 方法进行了比较，为什么不将三种方法一起比较？

回应：感谢审稿专家的意见。修改稿中已对“5 实证数据分析”部分进行了修改，修改稿中内容如下(第 19-21 页):

“采用实证数据进一步考察 Wald-XPDI 方法的表现，并与 HullIP、Wald-IC 方法进行比较。.....。

表 3 原始 Q 矩阵以及各方法对属性的修正情况

项目	原始 Q 矩阵			
	α_1	α_2	α_3	α_4
1	1	0	0	0
2	0	1*	0*	0
3	0	0	1	0
4	0	0	0	1
5	1*	1	0	0^
6	1*	1	0	0
7	1*	0*	1*	0
8	1*	0*	1	0*
9	1	0	0	1*#^
10	0	1*#^	0	1
11	1*#^	1*#^	0	1
12	1*	0	1*#^	1

注：*为 Wald-XPDI 方法调整的属性，#为 Wald-IC 方法调整的属性，^为 HullIP 方法调整的属性

本研究在饱和 G-DINA 模型框架下，使用 HullP、Wald-IC 以及 Wald-XPD 方法对原始 Q 矩阵进行了修正。表 3 中的结果显示，HullP 方法共修正了 6 个元素，Wald-IC 方法共修正了 5 个元素，Wald-XPD 方法一共修正了 16 个元素，Wald-IC 方法修正的 5 个属性均包括在 Wald-XPD 方法修正的属性之中。……。”

意见 3: 实证数据分析部分，作者没有说明是否先进行了模型比较，看数据更拟合哪种模型，也没有说明使用的是哪种模型进行的比较。

回应：感谢审稿专家的意见。先前研究发现(刘彦楼 等,2019; Liu, Andersson, et al., 2019)，在进行项目水平上的模型比较时，与其他统计量相比(如 LR)，基于完整信息矩阵的 Wald 统计量有明显的优势；且当 Q 矩阵中的一些元素错误设定而导致模型与数据失拟时，基于经验交叉相乘信息矩阵的 Wald 统计量在进行模型比较有较好的一类错误控制率及统计检验力(刘彦楼 等, 2019)。但是，Q 矩阵正确设定条件下用于模型比较的 Wald 统计量的表现要优于模型与数据失拟时的表现(刘彦楼 等, 2019; Liu, Andersson, et al., 2019)。因此，我们认为实践中(当无法有效地指定恰当的特殊 CDM 时)先使用饱和模型进行 Q 矩阵修正，然后再进行项目水平上模型比较可能更合适。

本研究的目的是在一般性的 CDM 框架下，开发具有广泛适应性的 Q 矩阵修正方法。故将重点放在了对可能存在错误设定的原始 Q 矩阵进行修正方面，实证数据部分使用饱和 G-DINA 框架进行模型参数估计及 Q 矩阵修正。为了避免引起误解，修改稿中已对“5 实证数据分析”部分进行了修改。修改稿中内容如下(第 19-20 页)：

“本研究在饱和 G-DINA 模型框架下，根据被试反应数据使用 HullP、Wald-IC 以及 Wald-XPD 方法对原始 Q 矩阵进行了修正。”

“但需要特别说明的是，本研究的目的是在一般性的 CDM 框架下开发具有广泛适应性的 Q 矩阵修正方法。因此，实证数据分析的重点是原始 Q 矩阵的修正，而没有在饱和 G-DINA 模型的基础上进一步在项目水平上进行模型比较(Liu, Andersson, et al., 2019)。……。”

意见 4: 实证数据分析部分，作者只呈现了相对拟合指标的结果，建议增加绝对拟合指标的结果。

回应：感谢审稿专家的意见。修改稿中已对“5 实证数据分析”部分进行了修改，采用实证数据将 HullP、Wald-IC、Wald-XPD 方法进行了比较研究。新增加了表 4 呈现了 3 种方法修

正后的 \mathbf{Q} 矩阵与原始数据的相对拟合指标和绝对拟合指标的结果，并对拟合指标的结果进行了分析。修改稿中内容如下(第 20-21 页)：

“……。使用相对拟合、绝对拟合及近似拟合指标比较原始 \mathbf{Q} 矩阵、HullP、Wald-IC 及 Wald-XPD 方法修正后的 \mathbf{Q} 矩阵的模型—数据拟合表现。拟合指标包括：相对拟合指标 AIC (Akaike information criterion)和 BIC (Bayesian information criterion)、有限信息绝对拟合 (limited-information absolute fit)指标 M_2 及近似拟合指标 RMSEA₂ (root mean square error of approximation; Liu et al., 2016)，结果见表 4。就相对拟合指标而言， $\mathbf{Q}_{\text{HullP}}$ 获得最佳的 AIC 指标， \mathbf{Q}_{XPD} 的 AIC 指标与其接近； \mathbf{Q}_{XPD} 获得最佳的 BIC 指标，其次是 \mathbf{Q}_{IC} ， $\mathbf{Q}_{\text{HullP}}$ 的 BIC 指标最差。即，Wald-XPD 方法修正后的 \mathbf{Q} 矩阵的相对拟合指标更优。在绝对拟合指标 M_2 上， \mathbf{Q}_{IC} 的 $p < 0.01$ ，表明 Wald-IC 方法修正的 \mathbf{Q} 矩阵与数据失拟； $\mathbf{Q}_{\text{HullP}}$ 和 \mathbf{Q}_{XPD} 的 p 值分别为：0.029 和 0.019，表明 HullP 和 Wald-XPD 方法修正后的 \mathbf{Q} 矩阵没有在 0.01 显著性水平上拒绝模型—数据拟合的原假设。对于 RMSEA₂ 指标而言，其值越接近 0 修正效果越好，可以发现 \mathbf{Q}_{XPD} 的 RMSEA₂ 最接近于 0，即 \mathbf{Q}_{XPD} 在 RMSEA₂ 指标上有好的表现(Liu et al., 2016)。综合考虑相对拟合、绝对拟合和近似拟合指标，本研究认为 Wald-XPD 方法修正后的 \mathbf{Q} 矩阵在模型—数据拟合方面表现最优。

表 4 基于三种方法修正前后 \mathbf{Q} 矩阵的拟合指标

Q	相对拟合指标		有限信息拟合指标			
	AIC	BIC	M_2			RMSEA ₂
			M_2	df	p	
$\mathbf{Q}_{\text{original}}$	4979.256	5245.278	23.919	15	0.067	0.0343
\mathbf{Q}_{XPD}	4962.484	5152.500	51.991	33	0.019	0.0338
\mathbf{Q}_{IC}	4964.200	5171.110	50.051	29	0.009	0.0380
$\mathbf{Q}_{\text{HullP}}$	4954.912	5178.709	40.037	25	0.029	0.0345

但需要特别说明的是，……。另外， M_2 统计量在模型参数过度设定时，即模型中冗余参数过多时，可能存在统计检验力不足的问题(参考 Chen et al., 2018)。举例而言，通过对比原始 $\mathbf{Q}_{\text{original}}$ 矩阵及修正后的 \mathbf{Q}_{XPD} 矩阵，可以发现 $\mathbf{Q}_{\text{original}}$ 中可能存在较多过度设定的元素，因此，导致 $\mathbf{Q}_{\text{original}}$ 的 M_2 统计量的 p 值大于 0.01。参考先前研究(Liu et al., 2016)，本文认为在模型—数据拟合评价方面，近似拟合统计量 RMSEA₂ 可能更具参考价值。”

意见 5：第 2 部分，“G-DINA 模型是一般、饱和的 CDM 模型且包含许多特殊模型；对饱和和 G-DINA 模型进行适当约束，可以获得多种特殊模型。”感觉重复了，看起来较累。是否改为“G-DINA 模型是一般、饱和的 CDM 模型，对其进行适当约束，可以获得多种特殊模型。”更好？

回应：感谢审稿专家严谨的意见。修改稿中已对“2 饱和 CDM 框架下的参数化 Q 矩阵修正方法”部分进行了修改。修改稿中内容如下(第 5 页)：

“本研究以 G-DINA 模型为例，考察新提出的 Wald-XPB 方法在 Q 矩阵修正的表现，并与以往研究者提出的 GDI、Wald-IC、Hull 方法进行比较。G-DINA 模型是一般、饱和的 CDM 模型，对其进行适当约束，可以获得多种特殊模型(de la Torre, 2011)。”

正文中未涉及的参考文献：

刘彦楼, 张倩萌, 郑宗军, 尹昊. (2019). 认知诊断模型中项目水平模型比较统计量的健壮性. *心理科学*, 42(5), 1251–1259.

编委复审意见：这篇论文提出基于完整经验交叉相乘信息矩阵(XPB)构建的 Wald-XPB 方法来修正认知诊断模型中的 Q，模拟研究表明，相对于其它两种 Q 矩阵修正方法，Wald-XPB 方法在大多数条件下可以获得更好的修正结果。实证研究也发现，基于 Wald-XPB 方法修正的 Q 矩阵可以更好地拟合测验数据。论文逻辑清晰，具有一定的理论和实践价值。经两位审稿人的评审，以及作者的修改，论文达到发表的要求。建议发表该论文。

主编终审：综合编委和审稿人的意见以及稿件修改情况，同意修改后发表。尚需要在文字表述和标点符号使用方面加以完善，如摘要中有“发现：新开发的 Wald-XPB 方法。。。”，引言中有“可以发现，Wald-IC 统计量是使用不完整信息矩阵计算的。”等不规范表述。有些标点符号使用也不够规范。

回应：非常感谢主编专家的对稿件提出的意见。根据主编的意见，已对稿件中文字表述和标点符号的使用加以反复推敲和修改。具体如下：

首先，将“摘要”部分文字表述不合适的地方修改为：

“研究表明：新开发的 Wald-XPB 方法在 Q 矩阵恢复率、保留正确标定属性的比例以

及修正错误标定属性的比例这三个主要指标上均有较好的表现，且整体上优于其他方法，尤其是在修正错误标定的属性方面。”

其次，“引言”部分文字表述和标点符号使用不合适的地方修改为：

“研究者(Ma & de la Torre, 2020; Na jera et al., 2021)构建的 Wald-IC 统计量是使用不完整信息矩阵计算的。”

“本研究选择 GDI、Hull、Wald-IC 方法与 Wald-XPD 方法进行比较的原因是：首先，Wald-XPD 是在 Wald-IC 方法基础上提出的，新方法与旧方法表现的异同有待探索；其次，先前研究表明在 GDI、Hull、Wald-IC 三种方法中，Hull 的表现是最好的，故有必要比较 Hull 与 Wald-XPD 两种方法的表现；第三，限制 GDI 及 iGDI 方法实践应用的主要原因是这两种方法均需要人为地设置一个截止值，与 iGDI 相比，固定的截止值对 GDI 方法的影响相对较小(Na jera et al., 2020)，因此本研究将 GDI 也纳入比较。”

最后，稿件中其他部分的文字表述和标点符号的使用也进行了修改。