

《心理学报》审稿意见与作者回应

题目：属性多级化的认知诊断模型开发及其 Q 矩阵设计

作者：蔡艳 涂冬波

第一轮

审稿人 1 意见：本文选题具有一定的创新性、前沿性强，研究方法可靠，研究结论可信，可读性强。但是读者还是有两点建议请作者考虑：

意见 1：是不是可以增加例子介绍具有层级结构时，属性可达矩阵的求法；

回应：专家的这条意见非常重要。为了继续沿用 0-1 属性化的可达矩阵这一概念，我们在修改稿中对可达矩阵 R_p 进行了重新定义，同时为了进一步区分 R_p 阵与多级化属性下 Q 阵设计，我们另外增加了一个新矩阵 R_p^* 阵，即测验 Q 阵中应该包含 R_p^* 阵，详见文章第 9 页。同时，根据专家意见，我们补充了线型、收敛型和分支型三种属性层级结构 R_p 阵和 R_p^* 阵的算法及其例子，详见文章第 9、10、11、12 页。并在文章第 21、22 页增加了一项新的实验（即实验 4），专门探讨线型、收敛型和分支型三种属性层级结构下可达矩阵的设计对诊断结果的影响。

意见 2：是不是需要增加参数估计方法的介绍？

回应：非常感谢专家的这条建议，我们在文章第 8、9 页补充了 PA-DINA 和 PA-RRUM 两个模型的 MCMC 参数估计算法。

意见 3：修改文中个别文字的顺序（可能是笔误）。

回应：这是我们在行文时的疏忽，我们又通读了全文，尽力改正了一些词的表达及顺序。详见文章 7-9 页以及 16 页。

审稿人 2 意见：

意见 1：文章的题目是“基于属性多级化的认知诊断模型开发及其 Q 矩阵设计”，但文章中仅研究了属性层级关系为独立型情况下的 Q 矩阵的设计，对线性型、收敛型和分支型等情况没有研究，建议补上。

回应：感谢专家的这条意见非常重要，的确如专家所言，作为 Q 矩阵的设计应该考虑不同属性层级结构下的 Q 矩阵设计，而本文最初版本仅考虑了独立型属性结构下的 Q 矩阵设计。根据专家的建议，我们在文章第 9、10、11、12 页补充了性型、收敛型和分支型属性结构下相关矩阵的求法。并在文章第 21、22 页增加了一项新的实验（即实验 4），专门探讨线型、收敛型和分支型三种属性层级结构下可达矩阵的设计对诊断结果的影响。

意见 2：当属性分为多个水平时，可达矩阵 R_p 的含义是什么？是否还能表达属性之间的层级关系？建议先给出属性分为多个水平时，可达矩阵的定义，然后再给出线性型、收敛型和分支型等情况下的可达矩阵。

回应：非常感谢专家这条重要的意见。为了继续沿用 0-1 属性化的可达矩阵这一概念，我们对可达矩阵 R_p 进行了重新定义，同时为了进一步区分 R_p 阵与多级化属性下 Q 阵设计，我

们另外增加了一个新矩阵 R_p^* 阵，即测验 Q 阵中应该包含 R_p^* 阵，并补充了线型、收敛型和分支型等情况下的 R_p 阵和 R_p^* 阵的求取过程。详见文章第 9、10、11、12 页。

意见 3：在图 1 下面，第 4 行中，根据公式 2.2，应修改为根据公式 2.4。

回应：这是我们行文时的粗心，谢谢专家的指正，我们在文章第 11 页修改过来了。

第二轮

审稿人 1 意见：(1) 文章按照评审专家的建议，补充了对线性型、收敛型和分支型等情况下的 Q 矩阵的设计，并给出了相应的模拟结果。(2) 作者按照评审专家的建议，给出了属性取多个水平时可达矩阵的含义及其构造方法，同时给出了线性型、收敛型和分支型等情况下的可达矩阵的构造方法。

意见 1：文章第 8 页新补充的内容“2.2 属性多级化的 PA-RRUM 模型和 PA-DINA 模型的 MCMC 参数估计”中，在介绍参数估计方法时，被试的属性掌握模式选贝努力分布，即 $\alpha_{ik} \sim Bernoulli(0.5)$ ，这意味着每个属性仅可能取 2 个水平，而不是多个水平，这与文章的内容“属性多级化”不配。实际上，作者新补充的算法就是 DINA 模型和 R-RUM 模型参数估计方法，而不是属性多级化的 PA-RRUM 模型和 PA-DINA 模型的 MCMC 参数估计的方法。

回应：非常抱歉，这时我们行文时的大意。的确如专家所言，在属性多级化认知诊断中 α_{ik} 的取值不再是 0 或 1，而是多水平的，因此它不是 0-1 的贝努力分布。本研究中 α_{ik} 采用了先验信息较少的整数均匀分布，即 $\alpha_{ik} \sim U(0, L_k)$ ， L_k 指属性 k 的水平数有 L 种 ($L \geq 2$)，也即 α_{ik} 服从 0 到 L_k 的整数的均匀分布；同时 MCMC 抽样中， α_{ik} 是从 0 到 L_k 的整数的均匀分布中随机生成，在 Matlab 程序中可以采用函数 $\alpha_{ik} = rand\ int(1, 1, [0, L_k])$ 实现。为此，我们将 α_{ik} 的先验分布及其 MCMC 抽样修正过来了，并补充了 MCMC 算法中使用到的 PA-DINA 和 PA-RRUM 两模型的似然函数公式，详见文章第 9-10 页，非常感谢专家的指正。其实，对于 MCMC 算法过程，属性 0-1 化 CDM 与属性多级化 CDM 基本相似，两者差别在 α_{ik} 的先验分布、 α_{ik} MCMC 抽取过程，以及似然函数（含项目反应函数， α_{ik}^* ， q_{jk}^* ， η_{ij}^* ）不太一样。

意见 2：文章还有几处笔误，比如第 5 页中间“ $q_{ik}1=$ ”应修改为“ $q_{ik}=1$ ”；第 6 页第 1 行“ $\alpha_{jk} =$ ”应修改为“ $\alpha_j =$ ”。

回应：谢谢专家的细致审稿，根据专家的意见我们在文章第 5-8 页将相关错误修改过来了，并进一步修改了其它类似小错误。

第三轮

审稿人 1 意见：作者已经按专家审稿意见修改，建议发表。

回应：感谢专家的肯定

编委复审意见：在论文发表之前，作者要对论文做如下处理：

意见 1：文献综述部分的英文图表要改成中文图表；

回应：我们已将文综述部分的图 1 英文图转换为中文图。

意见 2：目前的论文篇幅过长，纯文字已经 2 万多字了，考虑公式图表的篇幅更大，需要压缩篇幅。

回应：我们将文章篇幅进行了压缩，纯文字压缩在 1.5 万字左右，并删除两个图。

第四轮

主编意见：本文作者在前人的“属性多级化认知诊断”思路下，拓展出适合属性多级化的认知诊断模型：PA-RRUM模型(R-RUM for polytomous attributes)和PA-DINA模型（DINA model for polytomous attributes）；并测验Q矩阵的设计、新模型的参数估计实现及其性能，等。作者在指出前人研究的不足之处后，如“但Chen等人（2013）研究中忽略了多级化Q矩阵设计（即认知诊断测验蓝图设计）的重要性，及其对认知诊断测验编制的指导作用（而这正是认知诊断测验编制与传统测验编制的不同之处），以及属性多级化认知诊断模型的性能到底如何等”，却没有在本文的结论中给予回应，即并没能真正地补足前人研究的不足，没能为认知诊断在实践的应用提供实际的方法学支持。比如作者在结尾讨论部分提及编制测验时说：“...这意味着属性多级化认知诊断测验包含的题数一般会更多，项目测量属性的模式也更为复杂（例如：如何控制不同项目测量同一属性的不同水平），这其实对于属性多级化认知诊断测验的编制又提出了更高的要求”；在提及标定时则说：“...表明测验Q矩阵的标定是一项十分复杂的任务，而对于多级属性的标定显然更为复杂。因此，属性多级化认知诊断在实践中应用还需同步解决其属性标定的技术等问题。”这两句话等于又回到了原点。因此，1、建议作者：再次压缩文章的内容，不要将重点放在Monte Carlo验证的过程，而是如何解决作者提出前人研究的不足，并给出自己研究后得出的实际解决之道。2、建议将主编的意见以及作者依次建议修后的文章一并交给编委，再审并予以判断。

回应：非常感谢主编专家提出的宝贵修改意见及建议。

前人研究更多是提出了属性多级化的思想（Karelitz, 2004; de la Torre, Lam, Rhoads & Tjoe, 2010），但具体如何实现新条件下的测验Q矩阵设计等具体方法与技术均未探讨；本文正是基于此开展的，本文主要贡献在于探讨了两种属性多级化CDMs的性能以及提出了属性多级化认知诊断测验Q矩阵设计原则及其算法。当然，属性多级化的认知诊断是近年来才引起学者们的关注，因此当前国内外对此研究还处于起步阶段，研究的还较薄弱，限于时间及篇幅，本文也仅重点探讨了该领域的两个根本性问题：属性多级化的统计分析模型及其测验Q矩阵设计。对于诸如原稿件讨论部分及主编专家提出的涉及的具体测验编制及多级化的属性标定等领域仍需未来研究进一步探讨。

当然，原稿件中我们在文章最后结论及讨论部分并没有仔细回应引言部分中提出的前人研究的不足及本研究的解决之道，这的确是我们原复件行文中的缺陷，现根据主编专家的建议，我们重新修改了引言、结论及讨论部分，认真回答了引言部分中提出的前人研究中存在

的问题以及本文是如何来解决这些问题的，补充了本文的贡献之处；同时，为了避免误解、减少文章篇幅以及突出本研究的贡献，我们对文章结论及讨论部分进行了合并及删减；详见文章引言部分（第8页）及结论讨论部分（第17-19页），修改了的地方我们用粉红色进行了标注。

最后，根据主编专家的意见，在不影响原文整体结构的前提下，我们再次合并、删减了文章篇幅：（1）将原文实验 1 与实验 2 进行了合并描述（变为现在的实验 1），将原文表 4、5、6 相关表格结果进行了合并；（2）将原文实验 3 与实验 4 进行了合并描述（变为现在的实验 2），将原文表 8、9、10、10 相关表格结果进行了合并；经合并删减后，全文字数（除去历次修改说明及论文自检报告）缩减至约 1 万字（不含公式）。

编委复审意见：本论文已经经过多轮次的修改，而且在前期论文评审人的要求下论文的篇数在膨胀，后来又要压缩，从论文修改的角度说已经很难再进一步修改了。应该说这篇论文还是有一定的新意，主要表现在论文探讨了两种新的可以处理属性多级化数据的认知诊断模型（PA-RRUM 和 PA-DINA 模型），并提出属性多级化认知诊断测验 Q 矩阵设计原则及其算法。当然一个研究也不可能回答现在认知诊断所存在的很多问题。因此我建议发表。请编辑部及主编老师斟酌。

主编终审意见：可发。