《心理学报》审稿意见与作者回应

题目：重参数化的多分属性诊断分类模型及其判准率影响因素
作者：詹沛达　边玉芳　王立君

第一轮

审稿人 1 意见：

意见 1：文章进行多分属性的认知诊断研究，该类问题较新颖且具有较好的应用前景。文章重参数化了 3 个分别满足连接、分离和补偿缩合规则的 Pa-DCM 表达式，简化读者对 Pa-DCM 理解，并且对多级评分法进行了归纳和梳理，丰富了认知诊断模型理论；此外，文章首次探讨了多分属性情境下模型判准率的影响因素，有一定的创新性和理论及实践意义。但文章中仍有一些地方需要进行修改和说明：自检报告中提到 pG-DINA(Chen & de la Torre, 2013) 是成熟模型，但似乎有关 pG-DINA 的参考文献尤其实际应用极少，请作者对其成熟性进行说明并给出参考文献；

回应：很好的问题！自检报告中是“相对成熟”，用词相比于正文确实不够严谨。我们欲表达的更为准确的意思应该是正文中这段：“目前关于 Pa-DCMs 的研究还处于初期探索阶段，仅有少许 DCMs 可以处理多分属性(e.g., Karelitz, 2004; Templin, 2004; von Davier, 2005; Chen & da la Torre, 2013)，而其中较为完整的、针对 EDPA 的纯心理测量模型仅有 pG-DINA(Chen & de la Torre, 2013)。”即，仅仅是相对完整，但仍处于初期探索阶段，因此其在现实中的实用性还需要进行后续验证！另外，pG-DINA 相对难理解和难解释，这也必然会导致其在实际应用中存在困难。而这也正是本文的目的之一：简化模型复杂性，便于后续在实际中使用，且为现实使用提供理论基础(设定多少多分属性、多少水平是合适的？)。

意见 2：Sun, J., Xin, T., Zhang, S., & de la Torre. (2013)也使用了多分属性，其同时还是多级评分；

回应：是的。Sun et al.(2013)一文所使用的多级评分方法已经在原文讨论部分(6.1.1)提及了，如原文所述其多级评分方法是对祝玉芳和丁树良(2009)、田伟和辛涛(2012)等“1 属性 1 分”的拓广，即“1 水平 1 分”。而该方法是一种约束性或假设性很强的方法，主要有：“属性外显假设”、“属性与分数相对应假设”、“属性间满足连接缩合规则假设”等，详细见原文。这些假设将大大限制该多级评分方法的实际使用。

另外，您可以试想下，如果让评分者去根据考生的作答来评分，不仅要评他掌握哪些属性，还要评出来掌握了这些属性的哪个水平，我想这是一件极具挑战的事情，且即使完成也肯定有很大的误差。

当然，关于在 DCA 中如何更好地实现多级评分，尤其是能被普遍接受的评分方法，仍然是一件需要进一步探索的事情，非常值得今后关注！另外，也请参考我们对您第 12 个问题的相关回答。

意见 3：2.1 节最后一句，且当 L 和 K 均较大时(L+1)K>>2K。（只要 L>1,K>0 就可以）；

回应：非常感谢审稿人的提醒！作者在这里欲表示的意思是“远大于”而非“恒大于”，但经过查证发现数学中对“远大于”有较严格的定义，因此我们在修改稿中将“(L+1)K>>2K”替换为了文字描述 “(L+1)K 远比 2K 大”。

意见 4：DINA 获得广泛关注的原因还有其容易解释；
回应：是的，感谢您的提醒，我们已在修改中文中添加。

意见 5：文章中缺少连接缩合、分离缩合和补偿的严格定义；
回应：感谢您的建议。Maris(1995, 1999)在提出 conjunctive、disjunctive 和 compensatory 等概念时是直接使用公式的，我们同时参考了詹沛达等(2015)一文中对这三个概念的描述(其文在描述 conjunctive model 时使用了“非补偿”，通常我们可以认为 conjunctive 与 noncompensatory 是等价的)，在修改稿中分别给出了“连接”、“分离”和“补偿”3 个概念。

意见 6：实验中聚合模型和补偿模型（比如文章中的 RPa-DINO）结合时，被试得分该怎么算？
回应：我们想审稿人的意思应该是“当属性间为聚合层级结构时，使用满足分离缩合规则的 RPa-DINO 时如何判定被试得分？”。层级结构和模型本身的缩合规则是两个独立的概念，尽管层级结构限制了被试必须掌握属性 1 和 2 后才能掌握属性 3，即属性层级结构描述了属性之间的先后关系，但缩合规则假设了只要被试掌握属性 1 和 2 中任意一个后，其理想作答就应该是 1，即缩合规则描述了题目是如何考查这些属性的。
举一个不严谨的例子，假设被试必须先掌握“加”、“减”、“除”和“乘”之后才能掌握“四则运算”，这是聚合层级结构，而某题目是“有六个人，每人手里有 2 个苹果，则共有几个苹果？”，这时候该题只考查了“加”“减”和“四则运算”这 3 个属性，则只会加法的学生也能答对(2+2+2+2+2+2)，只会乘法的学生也能答对(6*2)，当然会四则运算的也能答对(2*2+2*2+2*2)……等等，这点在 Maris(1995, 1999)文中已经提到，分离缩合规则相当于题目允许被试采用多策略（利用不同的属性或属性组合）回答。

意见 7：4.1.2 节“实际中使用多分属性时数量不易高于 5 个”，改为不宜；
回应：感谢您的审稿，已修改！

意见 8：新模型存在参数估计困难，比如 RPa-DINA：K=5,L=2，比 DINA 参数个数多 242 个！作者打算如何解决这个问题？
回应：参数数量多是多分属性模型的特性，欲用此模型，则必须要面对这个问题。如果想减少待估计参数数量，我们觉得可以利用属性层级结构来减少可能的多分属性模式数量，即在参数估计时直接删除不满足属性层级结构的模式。

但本文为保证纯心理测量模型包含所有可能属性模式的假设（i.e., 不改变多分属性模型的本质假设），我们将所有可能存在的属性模式全部纳入参数估计当中，并未按照可能的属性层级结构去删减部分属性模式。当然，这样得出来的结果能起到“基线”作用，更具参考性。

意见 9：文中有部分错误单词“polytoumous”，请改正，并认真检查其它；
回应：已修改！

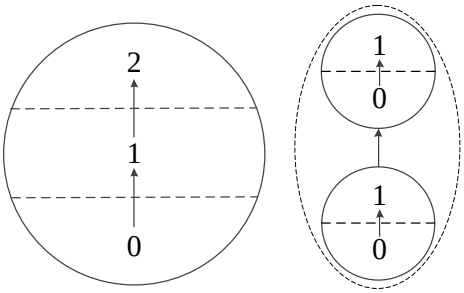
意见 10：文中对判准率的影响结论中提到“反比、正比”等，这有严格的数学定义，只能够说有某种趋势，如 x 上升 y 也上升，等等；

回应：感谢您的提醒，已修改！

意见 11：6.1.1 节式 9，等式左边的 j 是冗余，甚至是错误,因为右边没有出现指标 j ,左边只要 score[n,i]即可；
回应：感谢您的提醒，已修改！

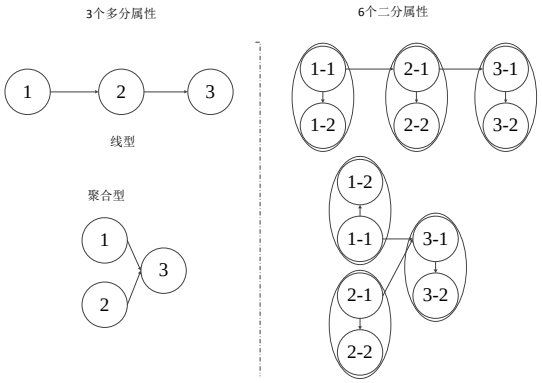
意见 12：6.1.2 中，关于是否有可能存在(1,2,3)或(1,1,3)的属性掌握模式，基于多分属性层级结构建构 Rp 矩阵及简化 Qp 矩阵的方法也是否需要修改的讨论，
如果去掉，问题是那些图还存在吗？
回应：首先，我们已经在修改稿中将该问题进行了更为准确的描述，将可达矩阵替换成了理想掌握模式，原因是后续研究发现该逻辑问题其实是有点复杂的：因为题目对属性的考查是可以不满足属性层级结构，也不满足水平约束的。
另外，关于您问题中“如果去掉，问题是那些图还存在吗？”这一问题，我们的理解是您在询问如果去掉部分不满足属性约束假设的模式之后，则多分属性之间的层级结构是否还存在？
答案是存在的！这点我们已经另攥文阐述，为避免不必要学术问题，我们仅在这里简单阐述下，烦请审稿人自己思考：

(1)多分属性与二分属性之间的对应关系：K 个多分属性相当于 $\sum_{k=1}^{K} L_k$ 个二分属性。其中 L 是每个属性的最高水平数，见下图。可发现，如果把多分属性拆分则相当于多个存在线型层级结构的二分属性。



多分属性与二分属性对应关系示例

为了让您更好地理解多分属性与二分属性之间的关系，我们把多分属性(Lk=2)和二分属性的之间的对应关系以图表示出来，见下图，其中二分属性中的“1-1”和“1-2”是指由多分属性“1”拆分而来的(见上图)，其余属性同理。您亦可尝试推导下图右侧中 6 个二分属性的 Rd 矩阵，之后可倒推出图 3 左侧 3 个多分属性的 Rp 矩阵，具体方法参见丁树良、罗芬等(2015)一文中的压缩算法。



在不同层级结构下多分属性与二分属性的对应关系

观察上图(以线型为例)，可发现当假设 3 个多分属性之间存在线型关系时，其对应的二分属性之间的线型关系仅存在于属性“1-1”、“2-1”和“3-1”之间，而“1-2”、“2-2”和“3-2”之间相互独立。聚合型和发散型与之类似。这表明仅从图式角度讲是可以存在被试仅掌握多分属性 1 的低水平(i.e., 二分属性中的“1-1”)，就掌握多分属性 2 的高水平(i.e., “2-2”)等类似的父属性低而子属性高的情况，这也就是现有的计算 Rp 矩阵、简化 Qp 和 IMPp 方法中存在“只要被试掌握了父属性(哪怕掌握地水平再低)，就有可能高水平地掌握子属性”的逻辑问题的根源！

最后回到审稿人的问题，多分属性之间的层级结构仅架构在属性的最低水平之间，因此去掉不符合逻辑约束的属性模式，并不影响层级结果关系。

意见 13： 属性与分数不独立法为什么不适合多选和建构题？多选也有答对一个给一分的情况；

回应：其实在建构反应题下，如何在 DCA 或题目内多维 IRT 框架下实现多级评分是一个相对较复杂的问题，主要的难点并不在于“如何描述多级分数与潜质之间的关系(i.e., 建模)”上，而真正的难点在于实际操作中评分者所使用的评分方法(i.e., 评分细则)是如何去规定分数与维度之间的对应关系的！当然，由于本文仍属于研究范畴，所以我们更关注“如何描述多级分数与潜质之间的关系”这个问题。下面回答审稿专家的问题：

就作者目前了解到的评分细则(e.g., 在全国基础教育质量检测中使用的)，往往评分是与采分点对应的，而与维度并不是严格对应关系的，因此属性与分数相独立的评分方法可能更符合实际操作。当然，采分点在某些时候也确实是与维度相关的，所以属性与分数相对应法也有其适用情境。但综合来看，仍是属性与分数相独立法的适用范围更广。

另外，有两点需要向审稿人说明，(1)multiple-choice (MC) item 是“多项选择题”，即有多个选项但正确答案只有 1 个的题目。如果换回中文常用的名称，对应的是“单选题”，但我们仍认为在学术文章中用直译会相对准确。因此，审稿人所说的“多选题”并非 MC item；(2)原文存在表达问题，我们认为“属性与分数不独立法”更适用于建构题，而不适用于多项选择题，修改稿已经进行了添加修改。

再有，我们之所以认为“属性与分数相对应法”在现实中不适用于多项选择题的原因是：如果将该方法用于多项选择题，限于多项选择题的选项数量(通常为 4 个)，当属性(或属性水平)数量之和大于选项数量时，则分数与属性之间的关系无法被合适地描述。

最后，关于审稿人所说的“多选题”，其实是同样存在题目选项数量无法与属性数量相对应的关系的。

意见 14：属性与分数相独立法，比如文中的 RPa-DINA，其等级之间差别的实质是什么？如

加法的“掌握很差”、“掌握一般”和“掌握很好”可否认为是不会 10 以内的加法、会 10 以内的加法、会 10 以上的加法，如果是这样那等级差别的实质可否认为还是属性（进位）；那么 RPa-DINA 的评分就变成和属性与分数不独立法的评分过程上再进行分数的二元化，而进行二元化之后，为了实现其诊断等级，其 Rp 矩阵、需要估计的参数大大增加，这样做是否值得？

回应：首先，我们认为该问题和您下一个问题是有联系的，综合来看这是一个很好的研究点，恕于能力和文章主题的限制，暂时无法从更为准确和客观地角度回答该问题，今后会专门针对该问题进行研究与探讨。

其次，依据我们的现有水平尝试回答您的问题。其实，在回答您第 11 个问题时，已经涉及到这个问题。其实多分属性每个水平之间的差别可转化为每个二分属性之间的差别，这就可能跟转化后的二分属性的颗粒大小有关。下图是 Chen 和 de la Torre(2013)一文中对属性水平差异的描述，可发现多分属性水平之间的差异其实是比通常意义下的两个属性之间的差异要小的，此时如果将其转为两个二分属性，或许这两个二分属性之间的区别就有点小了。所以，多分属性水平之间很可能是有量的差别但又未达到质的区别的情况，这也同时涉及到了专家定义问题。专家或许会针对实际问题认为划分为两个属性更合适，也或许会认为划分为 1 个多分属性的两个水平更合适，相当于 Pa-DCM 的存在为实际应用提供了另一个可选项。

Table 1. Polytomous Attributes in the Eighth-Grade Proportional Reasoning Assessment.

Attributes	Level 1	Level 2
Comparing and ordering of fractions	Students should be able to compare two fractions and determine whether one of these fractions is equal to, less than, or greater than the other.	Students should be able to order three or more fractions.
Constructing ratios and proportions from a situation	Given a problem situation involving ratios, students should be able to construct a single ratio to describe the situation.	Given a proportional situation, students should be able to construct an appropriate proportion.

Note: These attributes were adapted from de la Torre, Lam, Rhoads, and Tjoe (2010). Level 0 is defined as lack of attribute mastery.

另外，关于审稿人举得加法的例子，我想您说的是有道理的。如果单建构一个多分属性，则“掌握很差”、“掌握一般”和“掌握很好”可以分别对应不会 10 以内的加法(i.e., 不会基本加法)、会 10 以内的加法(i.e., 会基本加法但不会进位)、会 10 以上的加法(i.e., 会加法也会进位)，其对应地就可转化为 2 个存在线型关系的二分属性：基本加法→进位，此时被试只有(00)(10)(11)这三类属性模式，且分别对应了多分属性中的水平 0，1 和 2！

另外，根据对您第 11 题的回复可知，多分属性模型与其等价的二分属性模型的待估计参数数量是一样的，再举一个例子，假设有 K=4, L=3 的离散关系的多分属性，其应该对应的是 12 个二分属性（其中二分属性的层级结构为

1→2→3　　：共 4 种模式

4→5→6　　：共 4 种模式

7→8→9　　：共 4 种模式

10→11→12）：共 4 种模式

则，多分属性共有 4^4=256 种属性模式(e.g., (1,3,2,0))，不考虑属性层级结果的二分属性有 2^12=4096 种，而考虑完属性层级结构后仍为 4^4=256 种属性模式组合(e.g., 与(1,3,2,0)对应的(1,0,0,1,1,1,1,1,0,0,0,0))，所以两者是等价的，进而也就没有值不值得的问题了(相关模拟研究已另攥文阐述，研究结果也表明两者基本等价)。

意见 15：作者可否对属性与分数不独立法于属性与分数相独立法中的代表模型进行认知过程原理、性能和代价上的详细比较，以便大家从理论或实际应用中更清楚的认识这两类的区别和联系。

回应：审稿人这个建议非常好!但限于本文主题原因，我们就不在修改稿中涉及过多讨论，我们会把该内容作为我们的后续研究问题之一。

审稿人 2 意见：

意见 1：与 Chen 等人(2013)研究相比，该本并没有太多创新之处，从模型设置基本原理、模型数理基础等角度来看只是简单重复了 Chen 等人(2013)的模型思想；

回应：感谢您的审稿意见。但本文的重点本就不是进行新的建模，而是重参数化已有模型来简化模型的理解和解释难度，以期促进 Pa-DCMs 在实际中的应用，并且探讨 Pa-DCMs 的特性为实际应用提供适当的理论参考和支持。

根据 DINA 在当前研究中和应用中使用的比例来看，简单的易解释的模型显然更受欢迎。类似，在 IRT framework 下 Rasch type 模型的研究和应用比例依然高于其他模型(可参考 ETS 和 PISA 的做法)。因此，出于简化理解、解释和实际使用困难的角度出发，对 pG-DINA 进行重参数简化是有其价值的。这也回答了第一位审稿人的第一个问题，pG-DINA 的复杂性必然会导致其在实际应用中存在困难。

意见 2：作者研究了几个自变量对模型判准率的影响，结论如“属性越多判准率越差”、“多分属性水平数越高判准率也越差”，“模型判准率受多分属性层级结构的影响较大”，这些结果结论基本上都是可以预见的，这些结果也基本上是认知诊断中的一般常识性的问题；

回应：得到符合逻辑预期的结论说明了本研究的设计是正确的。另外，“‘可以预见的’‘常识性问题’”并不等同于“研究结论是‘常识性问题’”，这是基本的科研逻辑。

我们认为不是每个人的每篇论文都是 big idea，相反往往为了做一个 big idea，前期是要铺垫一些东西的，这就是本文的定位。

当然，我们也向往和追求审稿专家的“研究结果一定要超出预期才有意义和创新性”这一如此 high level 的要求。

意见 3：文章没有报告研究的变量又是如何影响项目参数估计精度；

回应：从文章标题就知，这并非本演关注的重点。况且在 DCA 中，人们对属性模式的关注度通常是高于对题目参数的关注度，同样以往类似主题研究(e.g., 涂冬波, 蔡艳, 戴海崎(2013); 蔡艳, 涂冬波, 丁树良(2013))也没报告题目参数的估计精度。

意见 4：作者参数估计的方法也没有交待；

回应：这不是本文重点，再有我们不认为在本文中放上一堆参数估计推导公式能对提升本文的水平有多大帮助，因为它不但不符合本文主题。因此，我们认为无需浪费篇幅提及相关内容。参数估计代码可向作者索取。

意见 5：文章无故创造出一些新名词（如聚合性属性层级结构），这直接影响到文章的可读性。

回应：感谢您的建议，首先“聚合”这个词不用看示意图也应知道是什么意思，我们认为这不会对理解文章有什么影响；当然，不同人的知识背景是不同的，考虑到也可能有读者与审

稿人一样会觉得这些词会影响文章可读性，我们希望得到审稿人的建设性意见，应该将这些您觉得难以理解的英文名词翻译成什么中文名词更合适？谢谢！

意见 6：研究的自变量中，属性间的相关程度与层级结构变量，这两个变量本来就有一定的内在关系，可惜作者并未深入挖掘。
回应：我们认为仅从模拟研究角度讲，属性间的层级结构和属性间统计相关性是两个可以独立操作的概念。从独立操作的结果来看，属性间统计相关性增加对属性判准率有提升作用，因此我们在统计相关性为 0 的条件下做后续探讨是一种“基线”研究，没有问题。另外，属性层级结构是可以人为操控的，而属性间相关性很大程度上依赖于受测样本，实际上是一种不可控因素，因此我们认为这两个概念是可以相互独立的。

另外，我们暂未想到如何在属性存在层级结构(除线型层级结构)的前提下仍能有效控制属性间相关性的方法，如果有好的方法，希望审稿人能够提供，以便我们改进研究，这样的意见才有建设性。

审稿人 3 意见：
意见 1：应该更加具体论述如何利用三种规则从原始的模型推导出三个新模型。这个是本文的创新点之一，但是原文中没有提供 Chen 等人的原始公式，也没有论述如何利用缩合规则如何进行推导。详细的展现推导过程有利于读者更好的阅读，也有利于充分展现作者的原创性工作。
回应：感谢您的建议。我们想如果在正文中添加相关内容或许会影响本文论述的流畅性或喧宾夺主。经考虑，我们将 pG-DINA 约束后的 3 个模型与本文重参数化的 3 个模型进行了对比，并放在了附录部分，以供读者阅读并判断重参数化的模型是否更易于理解。

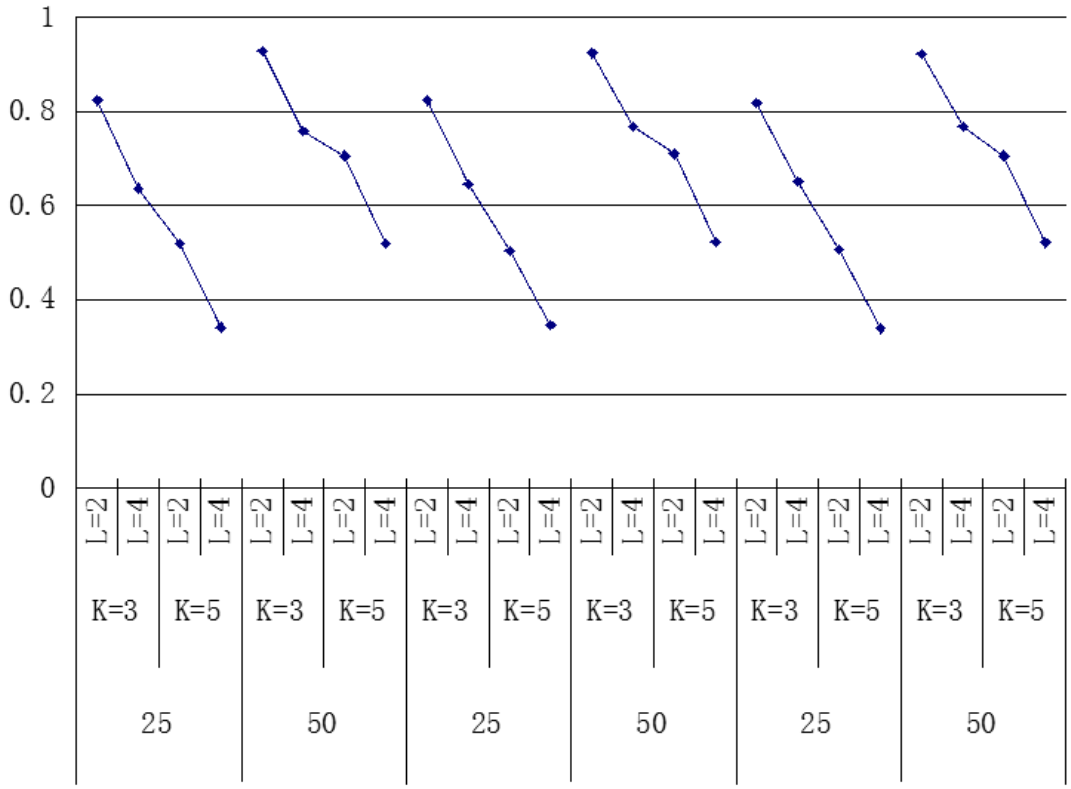
意见 2：研究二标题中专业术语使用不当。 “不同自变量交互时”中“交互”在心理设计中指的是交互作用，但是不同自变量交互时指的是什么，请作者明确表达含义。作者应该指的是“完全/部分交叉（ fully/partially crossed）的多因素设计”。
回应：您的理解是正确的，我们指的是多因素设计。已在修改稿中修改！

意见 3：研究二的研究结果与结论部分同样使用了“交互”一词（“且当两者交互出现时”），导致这句话非常费解。同时，本人通过画折线图验证了属性数量与属性最高水平并不存在交互作用。因此，请作者修改有关的语句，清晰地表达重点信息。
回应：感谢您的建议，已修改！

意见 4：通读研究二的研究结果与结论部分，非常遗憾没有发现任何有关交互作用效应的报告，因此没有从结果部分很好体现多因素设计的优越性。作者已经完成了这个复杂设计的所有模拟研究，把结果整理成一张大表，但是本人更希望作者能够在此基础上找出最有意义（或者是最有意思）的结果部分以图的形式呈现给读者。对于一个复杂的多因素设计，本人非常期待看到有趣的交互作用。
答：您的建议非常好，但由于这是模拟研究，所以多因素在模拟生成时本就是只考虑了主效应，而未涉及交互效应，因此结果反映出没有交互作用也是符合预期。
我们以 DINA 为例，下面将正文中表 7 转为图的形式呈现(有 3 组 25/50，分别对应 500 人，1000 人，2000 人)，从图中可以得到和数据表相同的结论：
1，被试数量对判准率的影响较小

2，题目数量的增加对判准率有影响但不大(0.1 左右)
3，水平数量的增加对判准率有较大影响(0.2 左右)
4，属性数量的增加对判准率有较大影响(0.3 左右)
这个结论可以通过待估计参数的数量来解释，RPa-DINA 中待估计参数数为(L+1)^K，显然，K 增加对数量的影响大于 L 增加对数量的影响。

当然，很遗憾，实验结果并没有显著的交互效应。



意见 5：公式（3）中的资格参数（qualification parameter），译为“限制参数”更加妥当。
回应：感谢您的建议，但经过思考和全文统筹，我们认为“资格参数”更合适本文欲表达的意思，也就是说当 ank≥qik，则表明被试 n 有资格掌握属性 k，因为他对该属性的掌握水平高于题目的考查水平。另外，该参数以及名词是我们自己设定的，应该是中译英的时候存在歧义，不知道审稿人能否推荐一个更为适合的英文翻译？

意见 6：3.5 模拟作答部分根据公式（8）无法得到被试作答，请补充完整。
答：这点可能是原文叙述问题，原文中的“得分”在二级评分情境下即为“得 1 分”，已修改。

第二轮

审稿人 1 意见：
意见 1：首先感谢作者的耐心回答，通过其反馈可以发现作者对该领域的问题有较深刻的思考，同时也使本人更清楚理解作者的文献。但是，仍有一些问题期待作者的解答。

回应：首先感谢审稿人对本文的认真审阅，你提出了一些非常有意义的问题，能够让我们与您一起思考，帮助我们完善和丰富了现有的想法和思路。但在回答您此次问题前还是要强调的是，本文的主题为简化已有的 Pa-DCMs 并探讨几个常见因素对其判准率的影响。只是在研究相关内容和攥写原文时因考虑到该逻辑性问题与模拟被试“真值”和参数估计时有一定的相关，故在讨论部分提及了，意在引起读者的兴趣而非对该问题进行详尽讨论和尝试解决，且原文中已经指出该问题还有待深入探讨。

经过您的讨论，我们认为目前我们对该问题尽管有所思考但还有欠缺，可能还无法较好地阐述出并解决之，且会占用较多的篇幅，恐会喧宾夺主和脱离主题。因此，我们在本次修改稿中已经删除了该部分内容，并根据之前与您的讨论内容添加了与主题相关性更高的一个小讨论（i.e., 多分属性与二分属性之间的关系）。当然，对该逻辑问题的思考和尝试性解决我们不会停止，期待在今后有一个更为丰满的思考时，再专文论述之。

意见 2：6.1.2 中，关于是否有可能存在(1,2,3)或(1,1,3)的属性掌握模式，基于多分属性层级结构建构 Rp 矩阵及简化 Qp 矩阵的方法也是否需要修改的讨论，

如果去掉，问题是那些图还存在吗？

回应：首先，我们已经在修改稿中将该问题进行了更为准确的描述，将可达矩阵替换成了理想掌握模式，原因是后续研究发现该逻辑问题其实是有点复杂的：因为题目对属性的考查是可以不满足属性层级结构，也不满足水平约束的。

……观察上图(以线型为例)，可发现当假设 3 个多分属性之间存在线型关系时，其对应的二分属性之间的线型关系仅存在于属性“1-1”、“2-1”和“3-1”之间，而“1-2”、“2-2”和“3-2”之间相互独立。聚合型和发散型与之类似。这表明仅从图式角度讲是可以存在被试仅掌握多分属性 1 的低水平(i.e., 二……

最后回到审稿人的问题，多分属性之间的层级结构仅架构在属性的最低水平之间，因此去掉不符合逻辑约束的属性模式，并不影响层级结果关系。

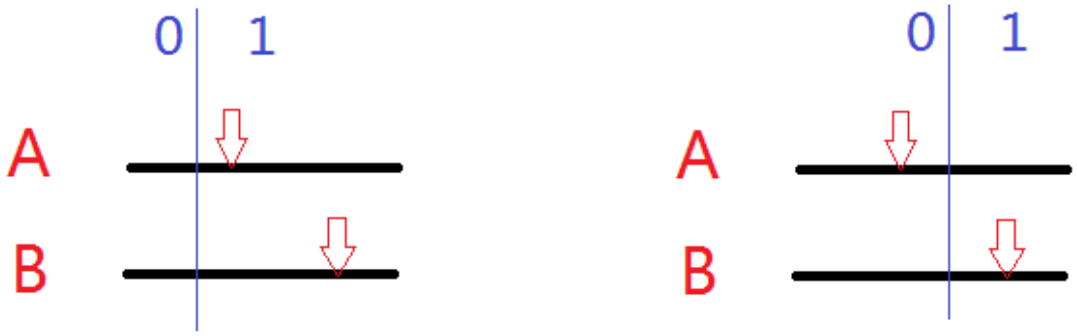
追加：个人认为构建 Rp 矩阵及简化 Qp 矩阵的方法是一般情况下的方法，而在多分属性下，以此方法构建理想掌握模式是有可能出现作者提到的子属性的掌握水平大于其父属性的掌握水平的情况，但是这并不意味着一定违反逻辑，因为父子两个属性划分的等级数和标准并未要求完全一致，1-1-2 完全有可能，相反 3-2 也有可能违反逻辑。作者从理想掌握模式入手避免了讨论 Qp 的一些麻烦，但是我们最终的诊断是依靠被试在试题上的反应而进行分析。因此在去除违反逻辑的模式后，有相当一部分被试在这种情况下就会认为存在猜测或失误从而误判；比如有两个属性父属性是结果百以内的加法（比如 25+47=72）、子属性是因子十以内的乘法（9*8=72），如果我就只做加法呢（9+9+9…），如果我会背乘法口诀表呢，如果我会减法呢（80-8），被试掌握子属性的高等级就一定会父属性的高等级吗？恐怕未必。如果按照作者的观点去除这部分期望反应模式，这部分被试恐怕会被误判，类似的情况恐怕有相当一部分，从而影响诊断效果。那么出现这种违反逻辑的情况，可否认为是强假设的的层级结构遇见约束性小的多分属性模型而导致的原罪呢，那么我们以后应用多分属性模型是否不要考虑层级结构，只要用离散型的比较好？

答：首先，无论原文中给出的“删除父(水平)＜子(水平)的模式”的逻辑约束方法是否恰当，我们应承认的是从事物发展的角度讲，“被试对父属性仅略知一二，而对子属性运用自如”这种倒三角的认知发展过程尽管有存在的可能但不应具有普适性，所以问题在于如何合理地对这个问题进行量化约束。

我们赞成审稿人指出的“父子两个属性划分的等级数和标准并未要求完全一致”，这点其实在我们后续的思考中已经想到了。

需要说明的是，该逻辑问题中的“掌握程度”和多分属性中的“掌握水平”是两个并不

完全相同的概念，如下图：两个属性 A 和 B，红色箭头为被试对两个属性的掌握程度，而蓝色竖线为掌握水平 0 和 1 的划分标准。左右两图一对比即可发现(即使在二分属性下)相同的掌握程度会随着划分标准的变化而被“标定”为不同的掌握水平：左图的掌握水平为(1,1)，而右图为(0,1)。



所以原文中提出的逻辑性问题是针对“掌握程度”而言，而非“掌握水平”，但为了能该逻辑问题进行量化约束，原文地给出了“删除父(水平)＜子(水平)的模式”这样一个初步逻辑约束，但仔细想来，由于“水平”和“程度”并不完全对应，所以该逻辑约束方法仅适用于“父(程度)＜子(程度)且恰好父(水平)＜子(水平)”这种情况(右图)，而当属性数量增加时，情况又会变复杂，所以原文的初步想法的确是有局限性的(实际上原文已经指出“但这样做是否合理还有待研究，且对该逻辑问题的探讨也有待深入”)

其次，无论是二分属性还是多分属性情境中，属性层级结构和属性水平约束(暂假设原文的逻辑约束可行)是两个需要分开的且具有递进关系的假设：属性层级结构假设→属性水平约束假设。只不过在二分属性情境中，两个假设被压缩到了一起：原文提到的水平约束假设，实质就是假设在“A→B”这个线型关系下约束“父(水平)≥子(水平)”，即只能有(1,0)而不能(0,1)。

1.在回答审稿人“那么我们以后应用多分属性模型是否不要考虑层级结构，只要用离散型的比较好？”时，我们首先要回答的问题是，是否承认属性层级结构这个假设？如果不承认或抱有怀疑态度(i.e.,施测样本或许有不少特例，或担心层阶结构界定错误)，则在题目编制时可以完全不或部分考虑下属性之间的先后关系，而在数据分析阶段选用不考虑属性层级结构的诊断模型(i.e., DINA)即可。

如果承认属性层级结构假设，则在就需要按照商定好的层级结构建立 Rp 矩阵和简化 Q 矩阵，以便指导题目编制。推算方法参见 Sun 等(2013)和丁树良、罗芬等(2015)，而多分属性之间层级结构的图式形式可见一审中对该问题的回答。【至此，当承认层级结构假设后，已有计算方法本身是没有问题的】但在分析数据之前就需要考虑第二个假设了，即是否承认属性水平约束假设？

2.如果不承认或抱有怀疑态度(i.e.,施测样本或许有不少特例)，则在分析数据时，仅需采用考虑层级结构的诊断方法即可(e.g., RSM, AHM, GDD)，或采用 DMCs 时从理想掌握模式矩阵中删除不满足层级结构的模式即可。但此时，也要考虑一种潜在危害：会把本不符合层级结构的被试也归为某个满足层级结构的模式之中（即任何假设都有不被满足的情况）。

如果承认水平约束假设，则当采用考虑层级结构的诊断方法或采用 DMCs 时就需要把理想掌握模式中不满足水平约束的模式删除。当然，也会存在一种危害(如审稿人所述的“误判”)：会把本不满足水平约束的被试也归为某个满足水平约束的模式之中。这和刚才承认属性层级结构假设一样，假设都是需要“冒风险”。

而至于，承认这些假设和不承认这些假设有什么区别，我觉得是承认了假设相当于“资源集中原则”，即把有限的信息量分配给那部分真的满足假设的被试，而忽略不满足假设的被试；而不承认假设相当于“资源均分原则”，即在有限的信息量情况下，考虑到所有的被试(属性模式)。孰优孰劣，或许就需要使用一些指标来判断了。

答毕。限于作者当前水平和能力有限，希望上述回答您能理解和满意。

附：针对审稿人“那么我们以后应用多分属性模型是否不要考虑层级结构，只要用离散型的比较好？”这一问题，我们再附加一些额外的思考内容：

当我们不停地对每个属性进行水平划分直至每个属性都接近连续变量(e.g., 30 个或 100 个水平)时，上图中的“掌握程度”和“掌握水平”之间的契合程度会增加，即“掌握水平”能够代表“掌握程度”的可能性将增加(假设水平数足够多时，两者甚至可以等价)，此时若仍假设多分属性层级结构成立，则原文中提出的逻辑约束“父(水平)≥子(水平)”方法很可能是行得通的。相关概念和内容其实在詹沛达和边玉芳(in press)的《概率性输入，噪音“与”门(PINA)模型》一文中已经有所涉及，只不过其文主题并不是在探讨我们当前所讨论的问题。其文在分析实证数据后发现，所有被试对属性 1~3 的掌握概率是递减的，而 Templin 和 Bradshaw(2014)在分析同一批数据时认为属性 1~3 为线型层阶结构。且由于“掌握概率”为描述“掌握程度”的一个量化连续变量，所以该概念与我们这里讨论的足够多的接近连续变量的“掌握水平”其实很接近的！那么既然当属性 1~3 存在线型层级结构时，被试对它们的“掌握概率”为递减的，那么是否能够从侧面说明本文提出的“父(水平)≥子(水平)”的逻辑约束也是有些道理呢？

再有，我们认为没有任何一个模型、方法能够 100%拟合数据。拟合度达到数据分析者能够合理解释，并且大多数人能够接受即可。并且，任何模型和方法都有一定的前提假设，现实中也一定存在不满足其假设的情境。我想审稿人应该能够认可我们上述观点。那么，用一个或多个相对指标(e.g., AIC BIC DIC)来判断考虑层级结构或考虑水平约束的方法更适合这批数据还是不考虑层级结构或不考虑水平约束的方法更适合是相对更合理的做法。

意见 3：属性与分数相独立法，比如文中的 RPa-DINA，其等级之间差别的实质是什么？如加法的“掌握很差”、“掌握一般”和“掌握很好”可否认为是不会 10 以内的加法、会 10 以内的加法、会 10 以上的加法，如果是这样那等级差别的实质可否认为还是属性（进位）；那么 RPa-DINA 的评分就变成和属性与分数不独立法的评分过程上再进行分数的二元化，而进行二元化之后，为了实现其诊断等级，其 Rp 矩阵、需要估计的参数大大增加，这样做是否值得？

……而考虑完属性层级结构后仍为 4^4=256 种属性模式组合(e.g., 与(1,3,2,0)对应的(1,0,0,1,1,1,1,1,0,0,0,0))，所以两者是等价的，进而也就没有值不值得的问题了(相关模拟研究已另攥文阐述，研究结果也表明两者基本等价)。

追加：从被试角度两者是等价的，这是肯定的，无需讨论。但是请注意我的提问是主要谈试题角度付出的代价，比如收敛型，多级评分的完备阵 Q 阵只要两列，多分属性的 Rp 阵就大得多，尤其是随着属性和等级数的增加差距更大，而考试的试题又往往是这部分题目的倍数。考虑到试题的编写、属性的标定，参数的确定等等，请问这样做是否值得？

回应：您提及的这个问题确实是存在的，多分属性下的 Rp 矩阵列数($\sum_{k=1}^{K} L_k$ 列)是要比二分属性下的 R 矩阵列数(K 列)多，但 $\sum_{k=1}^{K} L_k$ 随 K 和 L 的增加速度并不大。根据本文研究结果表明，为保证至少 0.6 以上的判准率，建议多分属性个数不易超过 5 个，最高水平数不易

超过 4。此前提下，Rp 矩阵应该不会超过 $\sum_{k=1}^{K} L_k$ =5×4=20 列，正常情况下是小于测验 Q 矩阵的列数，也小于题库可包含的题目数量的。且 Sun 等(2013)也指出并由丁树良，汪文义，罗芬，熊建华(2015)证明之，在多分属性情境下测验 Q 矩阵中包含 Rp 矩阵即可使理想反应模型与知识状态一一对应。所以，仅从需要编写的试题数量上看，增加的压力并不多。
而对于属性标定和参数确定的问题，我想肯定是要比二分属性情境下更有难度，但限于能力与知识储备有限暂无法准确回答。但我想，如果测验目标是为了给被试提供更为详细的反馈结果，我想前期的所有努力和付出应该都是值得的。

审稿人 2 意见：
*修改稿未再次发送给该审稿人。

审稿人 3 意见：
意见 1：第五部分“5　　研究二：多个自变量同时存在对 Pa-DCMs 判准率的影响”的标题不太恰当，建议修改为“Pa-DCMs 判准率影响因素的多因素设计模拟研究”。
回应：已修改，感谢您的建议。

意见 2：作者回答本人提出的第四个问题时，做了一幅图。这幅图比原文中的表格更加直观，因此，可以考虑把这个图插入正文。
回应：已添加，感谢您的建议。【但考虑到篇幅问题，最后可能会被修改掉】。

意见 3：英文摘要中的结论部分需要修改时态，具体包括结论部分的 2，3 点中的一般将来时需要修改为现在一般时，因为作者其他地方都是采用了现在一般时。
回应：已修改，感谢您的建议。

编委复审阶段：
Currently the article is 13000+ words, though there are 2 studies, but should cut down by 2000 words to 11000.
A: Actually, there are 5 studies in my paper, we try our best to cut down almost 1500 words. Currently the main body of my paper is approximate to 11300 words.

Recommend to change the title, and remove the word “研究”
A: Thank you for your recommend, we accept it.

I had polished the English abstract for the authors’ consideration
A: Thank you for your edit.