

《心理学报》审稿意见与作者回应

题目：引入眼动注视点的联合-交叉负载多模态认知诊断建模

作者：詹沛达

第一轮

审稿人 1 意见：

稿件 xb21-253 基于联合-交叉负载建模法开发了一种可处理多模态数据的学习诊断模型，在实证研究中用作答数据、反应时和注视点数作为数据来源进行了新旧模型的比较，并进一步在模拟研究中比较了几种模型的功效。总体而言，该稿件结构完整，逻辑比较清晰，但创新性略显不足，实验内容不够丰富。有一些建议和想法供作者参考，希望与作者讨论。

意见 1：

Zhan 等人(2021)用联合-层级建模法提出 H-MLDM，首次在学习诊断模型中同时分析 RA 数据、RT 数据和 FC 数据，适用于处理多模态数据。本文基于联合-交叉负载建模法提出的新模型仅仅是更换了一种建模思路，本质上没有太大变化，更像是小修小补式的创新，其创新意义需要作者进一步阐述。此外，根据研究结果表明，仍然是已有模型表现更好，新模型的表现不如已有模型，新模型存在价值就不高了。

回应：

感谢您的建议。

本研究理论创新：首次将联合-交叉负载建模法引入认知诊断领域，提出三种不同假设的多模态认知诊断模型；

本研究理论贡献：填补了在认知诊断领域缺少联合-交叉负载模型的空白；

本研究应用价值：从全面反馈视角出发，以认知风格和认知流畅性为例，首次尝试在认知诊断中提供认知结构以外其他认知特征的反馈。丰富了认知诊断反馈的范围，增加了认知诊断的实践价值。

实证研究结果表明新模型对数据的拟合优于 H-MCDM。另外，原文讨论部分已经指出三个新模型与 H-MCDM 是基于不同理论假设构建的，在本研究中，尽管三个新模型对实证数据的拟合程度优于 H-MCDM，这并不代表它们三者在任何测验情境下都优于 H-MCDM；比如，H-MCDM 的相对优势是理论结构简单、待估计参数数量较少。我们建议后续使用者

针对特定的实证数据，同时使用多个 MCDM 对数据进行联合分析，并基于数据-模型拟合指标来选择相对最合适的模型，并结合模型的构建理论对分析结果做进一步解读。

意见 2:

学习风格与学习或作答策略本质上很接近，不同学习风格的学生会选择不同的策略，因此作者提出的模型和多策略诊断模型之间的优势与区别是什么，需要加以论述。

回应:

感谢您的建议。在认知诊断领域，多策略是指同一群体中的不同个体会使用不同的潜在属性(模式)来作答同一道题目，即同一题考查了两或多套不同的属性模式。已有的多策略 CDM 也仅是基于传统 RA 数据构建的，既不涉及多模态数据也不涉属性以外其他构建。

修改稿中，我们在“2.3.5 与已有模型的差异”中增加了相关内容。另外，考虑到后续研究中的确可以从多模态 CDM 视角探讨多策略的分析，修改稿中，我们在讨论部分也提及了相关内容。

意见 3:

本文提出的 C-MLDM- θ 、C-MLDM-D 和 C-MLDM-C 是基于了三种不同的假设，这三种假设有相关的参考文献或依据吗？

回应:

三个假设本身都是可以解释的，符合逻辑；模型提出时并没有参考文献，是作者独立思考的结果。但修改本文时，发现 Wang 和 Chen (2020)使用了与 C-MCDM-D 和 C-MCDM-C 类似但不一样的假设。修改稿“2.3.5 与已有模型的差异”中已经指出 Wang 和 Chen (2020)模型中的假设与 C-MCDM-D 和 C-MCDM-C 的类似，同时也阐述了它们之间的差异。

意见 4:

帕斯卡注视点模型中的 ϵ_n 为被试 n 的潜在视觉参与度或参与度作者是如何将本文关注的学习风格与之关联起来的？学习风格的操作性定义是什么其心理学意义是什么，是如何通过参数反映出来的。这点非常重要，否则就变成了建模游戏。

回应:

感谢您如此宝贵的建议。

1. 修改稿引言中增加了有关认知风格和认知流畅性的描述；
2. 修改稿增加“2.3.4 认知结构诊断及认知特征推断”一节内容，阐述如何基于新模型的参数估计结果实现对个体认知结构的诊断及其他认知特征(如，认知风格和认知流畅性)的推断；

意见 5:

ϕ_i 和 λ_i 参数的本质需要进一步解释清楚 它们的指标含义是反映了什么问题,服从什么分布,取值范围及对应的心理学含义等信息。

回应:

感谢您的建议,修改稿讨论部分已经添加了相应内容。

意见 6:

尽管诊断更加强调的是诊断准确率,但从建模角度来看,为了实现对知识结构的精确估计,仍需要考虑其他情况,例如 τ_{ao} 对 RA 的影响,建议作者可以在讨论部分加以论述。

回应:

感谢您的建议,修改稿讨论部分已经添加了相应内容。

意见 7:

图 2 的 c 和 d 图式为何有 θ ? 根据建模的逻辑,这里 θ 起到什么作用? 如果要保留 θ 那 θ 、 τ_{ao} 与 e 之间为何不考虑多元联合分布情况?

回应:

感谢您的建议。首先,包含高阶潜在结构目的有两个:(1)便于新模型与 H-MCDM 对比;(2)为推断诸如认知风格和认知流畅性等其他认知特征提供额外参考信息。其次,不考虑三元正态分布的原因涉及到模型可识别性问题。

修改稿中已添加相关说明:

“为保证与 S-MCDM 和 H-MCDM 具有可比性,本研究设定在 C-MCDM-D 和 C-MCDM-C 中也存在高阶潜在结构。再有,为了保证模型的可识别性(即 θ_n 和 τ_n 之间的以及 θ_n 和 ε_n 之间的相关性已经被交叉负载解释),在三个 C-MCDM 中并没有使用三元正态分布来联接潜在能力、潜在加工速度和潜在视觉参与度这三个潜在变量(Bolsinova & Tijmstra, 2018; Molenaar et al., 2015)。”

意见 8:

在实证数据分析中, C-MLDM-D 和 C-MLDM-C 中的参数未能收敛,作者仅说是由于数据本身的原因,请作者阐述一下具体的原因,供读者和研究者参考。而所进行的模拟实验条件又较为局限,且 Q 矩阵是满足 DINA 模型参数可识别要求设置的,但现实中很可能无法满足,该如何处理? 因此,本人无判断 C-MLDM-D 和 C-MLDM-C 的真实表现。作者应该加入对使用这两个模型的讨论,例如什么条件下才能够使用,以及什么原因导致模型不收敛等。

回应:

感谢您的意见。修改稿中,通过模拟研究已经表明新模型在不同模拟测验条件下的参数估计

均可以有效收敛。而实证研究中可能导致参数估计不收敛的原因有很多，难以寻求真实原因且寻求真实原因也并非本研究的重点。实践应用中，在开发认知诊断测评时就应该考虑到模型可识别性问题，具体而言：专家预设的有关认知过程的假设决定了分析理应使用的模型(e.g., 专家预设只有被试掌握了所有考查属性才能正确作答，则数据分析时理应使用满足连接缩合规则的 CDM，如 DINA 模型)，此时，构建 Q 矩阵时就应该考虑到 Q 矩阵需要具备完备性且满足 DINA 模型的可识别性要求。因此，实际测验中 Q 矩阵是否满足可识别性要求是实践应用问题，不涉及到本研究的模型开发。

意见 9:

表 3 虽然呈现了相同知识结构的学生具有不同的 τ_{ao} 和 e 的取值，但作为诊断结果是相同的学生话，那么不同的 τ_{ao} 和 e 在实际中能起到什么作用呢？例如，都是 1111 的学生，应该如何使用 τ_{ao} 和 e 来帮助学生提升学习成绩？本人不太理解。

回应:

感谢您的问题。修改稿中我们已经增加了相应内容，详见表 5 及相关描述。

意见 10:

作者写到：“另外，应注意到，对不同题目而言， ϕ_i 和 λ_i 的估计值并不总是保持相同正负号，表明潜在能力对 RT 和 FC 的影响在不同题目上具有不同的方向性(即促进或抑制)。”该如何理解这两个参数的反向作用，请作者加以阐述。

回应:

感谢您的建议。修改稿中我们已经增加了相应内容，详见表 1 及相关描述。

意见 11:

模拟研究是一个什么类型的实验设计没有交代，另外，实验条件比较简单，没有充分讨论在不同情景下，如不同被试人数，不同题目质量，不同题目数量等的结果，导致无法全面评估模型的表现。部分参数估计结果没有报告，如 ϕ_i 和 λ_i ， m 和 d ，以及公式 11 中的 $n\Theta$ 等。MCMC 有几条链等。

回应:

感谢您的建议。修改稿中已调整模拟研究内容，探讨新模型在不同模拟测验情境下的表现。但考虑到《心理学报》对方法类文章偏应用的要求，我们刻意降低了模拟研究的比重(包括复杂性和篇幅等)。

意见 12:

作者写到：“换句话说，具有相同知识结构的学生并不一定具有相同的学习风格。若后续实

施补救教学，除针对知识结构的缺失外，也应考虑到不同学习风格对补救教学的影响。”该结论其实是符合常识性的结论，本人更关心的是除了从建模角度反映了该问题，其反应风格是如何影响到学生的知识结构的，从模型中能否反映某个参数位于什么范围对应的是一个什么结果，不同参数的不同范围之间的组合是如何对知识结构产生影响的。而不仅仅是报告 PCCR 等这些结果。

回应：

感谢您的建议。首先，量化研究本身就是要把“符合常识性的结论”以量的形式表现出来，新模型提供了方法学支持。另外，您所欲了解的认知风格是如何影响学生的认知结构的问题，这是本研究以及绝大多数联合建模研究都难以回答的问题。主要原因是联合模型假设各潜在构建之间是相关关系。修改稿中我们增加了如何根据分析结果实现对个体认知结构的诊断及认知特征的推断的内容，详见表 2、表 5 及相关描述。

意见 13：

学习诊断的关键在于对知识状态的识别精度，但在实证和模拟研究中均未充分体现新模型对知识状态识别的精度优于旧模型，普遍均要差一些，尤其是尤其是 θ 的估计结果的估计结果（RMSE 很大，相关性很低）很大，相关性很低）。那新模型的意义和。那新模型的意义和价值如何体现？价值如何体现？

回应：

感谢您的建议。

- (1) 实证研究中根据模型-数据拟合已经表明新模型对数据的拟合由于 H-MCDM;
- (2) 修改稿中模拟研究部分已经调整;
- (3) HO-DINA 模型中高阶潜在能力参数的返真性理论上就不高（由所考查的属性数量导致的），与以往相关研究比本研究结果符合预期(如, de la Torre & Douglas, 2004; Zhan, 2020; Zhan et al., 2018);
- (4) “5.2 局限与展望”部分最后一条已说明为何难以过模拟研究对比新模型与 H-MCDM 的表现。

意见 14：

图 9 建议需要呈现具体数值，具体估计结建议需要呈现具体数值，具体估计结果果无法从图中无法从图中反映。反映。

回应：

感谢您的建议。修改稿图 6 已呈现具体数值。

意见 15：

研究结果对于如何推进学习诊断，或识别学习风格等方面的探索较少。

回应：

感谢您的建议。修改稿中我们增加了如何根据分析结果实现对个体认知结构的诊断及其他认知特征的推断的内容，详见表 2、表 5 及相关描述。

.....

审稿人 2 意见：

文章基于多模态数据，提出了三个基于联合-交叉负载建模法的学习诊断模型，并通过实证研究和模拟研究验证了模型参数估计结果的返真性，以及使用该模型分析数据的优势。文章结果丰富，逻辑清晰，具有一定的创新性与现实意义。但是，作为一个方法类研究，文章有一些地方显得不够准确、清晰。具体意见如下：

意见 1：

引言部分

(1)第一次出现“模态”这一概念时作者并未加以解释，仅在后文提到“多种类型数据(即多模态数据)”，也没有正式、清晰地解释这一概念。模态从含义来看应当就是信息来源的意思，一般中文论文中很少用到这个概念，因此第一次出现时最好附上英文原文及其解释。

回应：

感谢您的建议。修改稿中已经添加有关多模态数据的概念描述：

“多模态数据是指对于同一个描述对象，通过多种仪器、测量设备或采集仪器获得的互补的多样性数据(Lahat et al., 2015)。”

(2)作者说“个体的学习行为是多变量的产物，这些变量之间彼此相关；因此，为了增进我们对个体学习行为的了解，应该联合分析这些变量。”那么，是否有研究已经应用了多模态数据的分析技术？并与传统方法比较体现出其优势？希望作者在文献回顾部分提供相关证据。

回应：

感谢您的建议。实际上，只要超过 1 种数据模态就可以被称为多模态数据；如正文中述，目前在心理与教育测量中，已经有大量的联合分析 RA 和 RT 数据的研究，研究结果表明考虑 RA 数据以外的其他模态数据不仅可以丰富测量结果所包含的信息，还可以提高模型参数的估计精度(即减少测量误差增加测量信度)。另外，近期也有一些研究使用了 RA 和 RT 数据以外的其他类型数据(如, Jeon et al., 2021; Man & Harring, 2020; Bezirhan et al., 2021)，在修改稿中也有所提及。

意见 2:

多模态数据的联合分析部分

(1)图 1 (b) 中的联合-交叉负载学习诊断建模示意图, 为什么没有潜在能力, 后面的图 2 又包括了潜在能力指向属性部分。

回应:

该问题与审稿人 1 的问题 7 类似。首先, 新模型包含高阶潜在结构目的有两个: (1)便于新模型与 H-MCDM 对比; (2)为推断诸如认知风格和认知流畅性等其他认知特征提供额外参考信息。

(2)LRT 模型第一次出现时是什么的缩写? 没有说明。

回应:

感谢您的建议, 已修改。

(3)与另外两种方法相比, 独立建模法的缺陷, 应当不是文中提到的“忽略了参数估计误差带来的影响”, 而是忽略了交叉负载的关系?

回应:

您说的对, 但考虑到分离建模章节更多地是为了引出三个测量模型, 且 S-MCDM 的表现相对较差, 修改稿中我们将这句话删除了。

(4)联合-交叉负载建模法是否有相关的理论依据? 即一些研究证明这些路径之间是存在的。

回应:

感谢您的建议。本研究的主要目的是将 IRT 领域中的联合-交叉负载建模法引入到认知诊断领域, 而至于这些路径是否存在则需要根据数据分析结果(即 ϕ_i 和 λ_i)来判断。

(5)公式 (10) 下面的解释不够准确, 应为“ ϕ_i 和 λ_i 分别表示理想作答为 1 的学生和理想作答为 0 的学生之间 RT 和 FC 的均值的差异。”

回应:

感谢您的建议。已作出相应修改。

(6)在实证研究中, 各参数的先验分布是如何确定的? 有无进行先验分布的敏感性分析?

回应:

感谢您的建议。修改中已增加参数估计对包含不同信息量的先验分布的鲁棒性分析, 详见附录 S1 节。其中, 研究所用的中信息量先验分布是根据已有经验和已有研究结果(如, Man &

Harring, 2019; Zhan et al., 2021)设定的。

意见 3:

实证数据分析部分

(1)“在三个模型中，我们均分别评估不同模态数据与其测量模型之间的拟合关系”。

这里使用的什么指标没有明确指出？从后文看是 $-2LL$ 和 ppp 。

回应:

这句话跟上面两句是连着的。经过思考，应该不需要额外补充说明。

(2)这部分应给出参数估计收敛情况。那么，对拟合情况的评价是否应当仅使用参数估计结果都收敛的 *replication* 计算？

回应:

感谢您的建议。当固定随机种子后，实证数据分析不需要重复，1 次估计即可。此时，C-MCDM- θ 所有参数都达到收敛标准，而 C-MCDM-C 和 C-MCDM-D 并非所有参数都达到收敛标准，表明两者参数估计结果不准确，结果仅供参考。

(3)由于真值未知，在实证研究中如何评价参数估计精度？估计标准误小能说明估计精度高吗？估计精度高能说明模型和数据拟合更好吗？这里的逻辑希望作者思考。

回应:

感谢您的提问。或许我们对“精度”一词的理解有差异，原文中所述的“精度”是波动性小的意思(类似于“信度”概念)，而非估计值与真值之间的匹配程度(类似于“效度”概念)；修改稿中，我们明确了这点，见脚注 3。如果我们有理解不到位的地方，也烦请您明言。

(4)作者通过表 3 意图说明使用 C-MLDM- θ 模型的优势。其实，使用 S-MLDM 和 H-MLDM 也会得到这样的结果，只要潜在属性上诊断结果相同的被试，潜在加工速度和潜在视觉参与度的值不同即可。这个例子只能说明利用多模态数据的优势而不是联合分析的优势。为体现该模型优势，需要呈现潜在加工速度或潜在视觉参与度值相同的被试，由于属性诊断结果不同，在同样题目上得到的 RT 或 FC 不同。

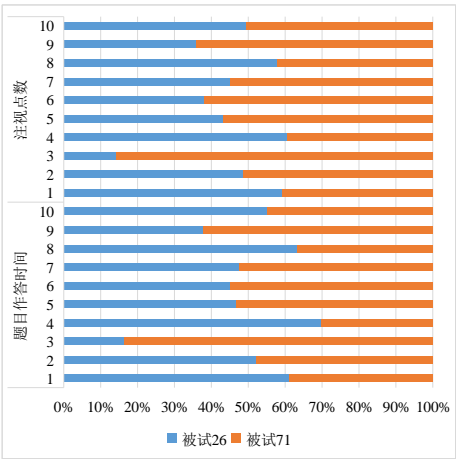
回应:

感谢您的建议。您说的没错，修改稿中表 5 (原表 3)并非为了说明 C-MLDM- θ 模型的优势，而是从一个新的视角解读 C-MLDM- θ 模型的估计结果。修改中也提及“此外，实证研究也向读者展示了如何根据数据分析结果来实现对个体认知结构的诊断及其他认知特征(如，认知风格)的推断”。

对实证数据而言，我们主要还是以模型-数据拟合指标来判断各模型的相对优劣(且该优

劣也仅是针对该数据集而言)。另外, ϕ_i 和 λ_i 的参数估计值一定程度上已经表明了潜在能力会影响 RT 和 FC。进一步, 我们对比被试 26 和 71, 可发现两者的潜在加工速度和潜在视觉参与度相近但属性掌握模式和潜在能力不同, 百分比堆积图显示两者在一些题目上的 RT 和 FC 均不相同。但考虑到撰写逻辑和篇幅问题, 我们并没有将下表和下图放入正文中。

被试	α_1	α_2	α_3	α_4	θ	τ	ε	认知特征推断
26	1	0	1	0	0.299	0.270	-0.321	认知流畅+非聚焦者
71	1	1	1	0	0.389	0.279	-0.279	认知流畅+非聚焦者



潜在加工速度和潜在视觉参与度相近但属性掌握模式和潜在能力不同的两被试的题目作答时间和注视点数(百分比堆积图).

意见 4:

模拟研究部分。

(1)模拟研究的目的包括“新提出的三个 *C-MLDM* 在不公平环境下也具有与 *HMLDM*(数据生成模型)相近的表现。”什么叫不公平环境？需要表述准确。

回应:

修改稿中模拟研究部分已经重写，不涉及该问题。

(2)三个潜在变量之间的相关系数矩阵，其实包含了三个水平，是模拟研究中的一个模拟因素。这里表述不够清晰，作者应当在开始交代模拟研究的模拟因素（自变量）及其水平

回应:

修改稿中模拟研究部分已经重写，不涉及该问题。

(3)结果部分 *PSRF* 采用了严苛程度不同的三个标准，是否都有文献支持？

回应:

修改稿中仅涉及 1.2 和 1.1 两个标准，且都标准了参考文献。

(4)从表 4 中看出，潜在能力的 *RMSE* 普遍较大（最小为 0.523），是否与同类研究结果一致？应当如何解释？有没有可解决的方式设想？

回应：

HO-DINA 模型中高阶潜在能力参数的返真性理论上就不高（由所考查的属性数量导致的），与以往相关研究比本研究结果符合预期(如, de la Torre & Douglas, 2004; Zhan, 2020; Zhan et al., 2018)。

(5)图 8 信息量较小，建议用表同时呈现 *ppp* 和 *DIC* 结果（*ppp* 结果没有报告）。

回应：

修改稿中模拟研究部分已经重写，不涉及该问题。

意见 5：

总结与展望部分

(1)文章对新模型优势总结为“新模型可实现对知识结构和学习风格的双重诊断”，其实旧模型也可以，只要能够得到潜在能力、潜在加工速度和潜在视觉参与度的估计值即可诊断。

新模型的优势主要体现在融合和交叉载荷上。作者应当进一步思考和明确新模型独特的优势。

回应：

修改稿已经调整引言及问题提出，本研究旨在实现对认知结果的精准诊断及与其他认知特征的反馈，由于数据模态类型的限制，仅以认知风格和认知流畅性为例来说明多模态模型为实现对个体的全面反馈提供了可能性。

本研究理论创新：首次将联合-交叉负载建模法引入认知诊断领域，提出三种不同假设的多模态认知诊断模型；

本研究理论贡献：填补了在认知诊断领域缺少联合-交叉负载模型的空白；

本研究应用价值：从全面反馈视角出发，以认知风格和认知流畅性为例，首次尝试在认知诊断中提供认知结构以外其他认知特征的反馈。丰富了认知诊断反馈的范围，增加了认知诊断的实践价值。

(2)在局限性部分作者提到“本研究只关注了一种学习风格的分类(即沉思型 vs.冲动型)”。其实，研究只得到模型中不同参数估计结果，并未实现学习风格分类。这种分类应当如何实现，是在潜变量上设定阈值吗？如何设定？可否与直接基于实际观测变量的 *LCA*

(*Latent Class Analysis, LCA*) 或 *LPA (Latent Profile Analysis, LPA)* 比较分类结果一致性?
可对在这部分表述更准确一些, 并做出一定的展望。

回应:

感谢您的建议。修改稿中我们已经增加“2.3.4 认知结构诊断及认知特征推断”一节内容, 阐述如何基于新模型的参数估计结果实现对个体认知结构的诊断及其他认知特征的推断。另外, 讨论部分也阐述了当前这种分类方法的局限性。

第二轮

审稿人 1 意见:

意见 1:

作者写到, Wang 和 Chen (2020) 提出的模型非常类似于联合交叉负载建模, 也就是说前人已经将这种建模思路引入了认知诊断。作者仅是在 Wang 等的基础上增加了额外的一些内容(如注视点, 理想作答反应等)。审稿人想要探讨一个问题: 若最终解决的问题是一样的, 只是基于不同的理论假设, 并且原有模型更加简洁的话, 那么为何不使用更加简洁的模型呢?

回应:

感谢您的建议。首先, “类似”≠“是”, 本研究更多地是对 Zhan et al. (2021) 的延伸, 而与 Wang 和 Chen (2020) 所关注的研究问题并不相同。其次, 修改稿中我们增加了一则模拟研究以表明考虑交叉负载的相对优势(即考虑交叉负载的 C-MCDM 比不考虑交叉负载的 H-MCDM 对不同测验的兼容性相对更好)。

意见 2:

修改后的题目其实也没有突出作者的研究主题, 可否考虑将具体方法“联合-交叉负载多模态”这种词汇替换成类似于能够提供 FC 等这些信息的诊断方式之类的。供参考

回应:

感谢您的建议, 修改稿中题目已修改为“引入眼动注视点的联合-交叉负载多模态认知诊断建模”。

意见 3:

Ranger (2013)、Bolsinova 和 Tijmstra (2018) 指出当潜在能力和潜在加工速度之间的相关系数不等于 0 时, 各模态属性间的信息可以被使用。为何作者此处阐述潜变量之间需要存在较高的相关?

回应:

已修改为“不等于 0”。

意见 4:

5 条链，每条链 6 万次迭代

回应:

感谢您的建议，根据已有研究经验，作者认为对于基于 Gibbs 抽样的贝叶斯 MCMC 算法，2 条随机起点链+30 次迭代足以，没有必要做冗余工作。

意见 5:

Ppp 值是否有临界值可以参考？该列数据中的 ppp 值离 0.5 有点远

回应:

与统计学常用 p 值类似，修改稿中已增加相应说明“通常， $ppp < 0.05$ 或 $ppp > 0.95$ 可被视为不拟合(Gelman et al., 2014)”。

意见 6:

参数估计值的标准误未呈现，因此审稿人无法判断各参数估计值的准确性如何

回应:

修改稿中已增加相应内容，见表 4.

意见 7:

模拟研究中，作者并没有对已有模型和新模型间的表现进行比较，故审稿人无法判断新模型是否真的比已有模型更好。也许作者认为，实证数据表明 C-MCMD- θ 的相对拟合度要好于 H-MCDM（且在参数未收敛的前提下），所以可以表明新模型表现要好于 H-MCDM，故模拟研究中不再使用 H-MCDM。但实证数据还表明 C-MCMD- θ 和 H-MCDM 均要好于另外两个新模型，按照该逻辑，模拟研究中的 C-MCMD- θ 也应要好于另外两种新模型，但模拟研究的结果并没有得到这种预期结果。因此，审稿人认为模拟研究中有必要加入 H-MCDM 进行比较，以明确新模型是否真的比 H-MCDM 更好。如果模拟研究的结果表明，新模型并没有表现出比旧模型更大的优势，那么如此复杂的新模型也没有存在的必要性了。

使用新模型生成数据的情况下，如果旧模型的诊断分类结果与新模型很接近，那么完全可以使用更加简单的旧模型；而使用旧模型生成数据情况下，如果新模型的诊断分类结果与旧模型很接近，且当真模型为新模型本身时，新模型的分类结果要好于旧模型，那么可以接受新模型。通过些情境的设定，一定程度是可以推断新模型性能的

回应:

感谢您的建议，在不明显增加篇幅的前提下，修改稿中我们增加了模拟研究 2 以表明考虑交叉负载的相对优势(即考虑交叉负载的 C-MCDM 比不考虑交叉负载的 H-MCDM 对不同测验的兼容性相对更好)。

.....

审稿人 2 意见：

作者回答了我上一轮的一些问题，并做了相应修改，文章质量有一定提高。但是有些地方仍然不够清楚，现陈述如下。

意见 1：

目前引言部分较长，且段落之间的逻辑关系不够清晰。例如，第一段和第二段的关系是什么？第一段在讲认知诊断，第二段是认知风格，之间应当有一些过渡和联系。而第三段又开始讲认知诊断模型，似乎是承接第一段的内容。

回应：

感谢您的建议，修改稿中已经调整这三段顺序，并已对引言进行精简。

意见 2：

表 2 中的认知特征推断仅涵盖了加工速度和潜在视觉参与度两个维度，而从表前三列的内容来看，应当涵盖三个维度，应加上潜在能力维度。

回应：

感谢您的细心审阅，实际上表 2 中的认知特征推断已经包含潜在能力维度。比如，“认知流畅性”是由高潜在能力+高加工速度共同推断的、“沉思型”是由高潜在能力+低加工速度共同推断的。

意见 3：

“2.3.5 与已有模型的差异”中，对新模型的特点总结不够准确。从前文的行文逻辑来看，先是从最初的联合建模的模型，拓展加入了多模态数据，然后在此基础上 Zhan 等(2021)提出的模型又在认知诊断框架下建模，最后拓展到本研究的模型，相对于以往模型的改进其实就是加入了交叉负载。因此这部分无需与每个最原始的模型比较以突出其创新性。而应当将重点放在为什么要加入交叉载荷，其必要性如何。我建议可以从两个方面论证，一是基于前人研究或者自己的实证数据，证明确实存在交叉负载，其含义和解释如何。二是在模拟研究中，加入用不同模型产生数据，再用所有模型拟合的情况。例如，用含交叉负载的模型产生数据，再用不含交叉负载的模型拟合，证明忽略交叉载荷的严重后果。同时，用不含交叉负载的模型产生数据，再用含交叉负载的模型拟合，证明该模型的稳健性。

回应:

感谢您的建议,在不明显增加篇幅的前提下,修改稿中我们增加了模拟研究 2 以表明考虑交叉负载的相对优势(即考虑交叉负载的 C-MCDM 比不考虑交叉负载的 H-MCDM 对不同测验的兼容性相对更好)。

意见 4:

结果部分对不收敛的解释为“或许是由于数据本身的原因”,这种解释似乎不应该出现在学术论文中。在实证数据中作者提出的两个模型收敛情况具体如何,不应当用“大部分”“小部分”这种不够具体的词语描述。其余模型的参数是否全部收敛,应明确表述。不收敛的情况可能的原因是什么?是否会影响模型的应用?

回应:

感谢您的建议,修改稿中已经添加相关说明:由于 C-MCDM-D 中约 39.71%的待估计参数(包括 g_i , s_i , ξ_i , ω_i , m_i , d_i , φ_i , λ_i , γ_{0k} , γ_{1k} , α_{nk} , θ_n , τ_n , ε_n , σ_A^2 , σ_T^2 , σ_E^2 , $\sigma_{\theta T}$, $\sigma_{\theta E}$, σ_{TE})和 C-MCDM-C 中约 29.35%的待估计参数(同上)并未达到收敛标准($PSRF < 1.2$),因此,这两个模型与数据的拟合结果仅供参考,在实证研究中不再参与后续探讨。

意见 5:

为避免歧义,实证研究中估计精度的表述最好直接使用估计的标准误。

回应:

感谢您的建议,已做出相应修改。

意见 6:

作者修改了原表 3 并回答了我的疑问,但是这部分实证研究的目的应当是证明新模型的必要性,表 5 所提供的结果原来的模型也能提供。作者提到 φ_i 和 λ_i 的参数估计值一定程度上已经表明了潜在能力会影响 RT 和 FC。那么作者应当给出这些参数估计值的置信区间,以帮助读者了解其是否显著不等于 0。

回应:

感谢您的建议,结合审稿人 1 的建议,修改稿中我们提供了参数估计标准误(后验标准差),同样能反映您所需内容。的确,根据表 4 中结果,发现个别交叉载荷较小的题目(e.g., 4、9)的 φ 和 λ 并非显著不等于 0。

意见 7:

附件“参数估计对先验分布的鲁棒性分析”,各参数的先验分布都没有设置超参数,是出于什么考虑?

回应:

感谢您的建议,考虑到鲁棒性分析是基于模拟研究进行的,数据分析者对数据已经有所了解。同时,在实证研究中所分析的数据也是有前期研究经验的,并非一无所知,所以我们没有使用超先验分布。在修改稿讨论部分,我们适当提及了超先验分布。

意见 8:

作者提出加入了交叉负载的模型和原有的模型一起,形成了同一体系下的多种不同复杂程度的模型。在实际应用中应当如何选择模型?例如,是否应当加入交叉负载?是否要通过模型选择的方式确定哪种模型与数据拟合更好?希望作者能给出实践的建议。

回应:

感谢您的建议,实际上原文已经提及“我们建议后续使用者针对特定的实证数据,同时使用多个 MCDM 对数据进行联合分析,并基于数据-模型拟合指标来选择相对最合适的模型,并结合模型的构建理论对分析结果做进一步解读。”

第三轮

审稿人 1 意见:

已无修改意见。

回应:

再次感谢您所有的审稿工作。

审稿人 2 意见

作者回答了我上一轮的问题,并且加入了模拟研究以探讨新模型的稳健性,使得整个研究的逻辑更加严密,文章质量有了非常大的提高。这一稿有一些小问题,现罗列如下,供作者参考。

意见 1:

目前引言部分经过修改后逻辑较为清晰。但是本研究主要是对已有模型的拓展,即加入了交叉负载的部分,然而引言部分没有体现这一部分内容。从现有的文献综述来看,也可以适用于提出已有的 H-MCDM 模型。希望作者在文献综述中加入对加入交叉负载的原因考虑。例如,理论上的分析,或者实证的证据。实际上在“2.3.3 联合-交叉负载建模法”中有一些描述,但是文献综述部分似乎应当更贴合本研究所提出的模型。

回应:

感谢您的建议，修改稿中我们进一步调整了引言内容，强调了 H-MCDM 的理论局限性及引入联合-交叉负载建模法的必要性。

意见 2:

文章提出的模型其实是在已有模型基础上加入交叉负载，也可以说是更 general 的一种模型。因此，在模型提出或优势总结方面，可以从这个角度切入。说明新模型包括了已有模型，即使不存在交叉负载，理论上新模型也能够处理。

回应:

感谢您的建议，修改稿中我们强调了这一点。

意见 3:

建议在新加入的研究 2 中不需要报告 DIC 相对优势和 PCCR 相对优势。一是很少看到报告这些指标，二是从公平比较的角度不用计算相对优势，其实从已有的结果就已经能看出新模型的稳健性。

回应:

感谢您的建议，修改稿中我们从表中删除了相对优势，而在正文中提及。

意见 4:

为和研究 1 对应，建议在研究 2 中报告参数估计准确性的结果。出于篇幅考虑，可放在附录中。另外，研究 2 中当产生模型为 H-MCDM 时，新模型估计的大部分交叉负载的置信区间应当包括 0。可以从结果上说明一下，进一步验证新模型是更加 general 的，体现其优势。

回应:

上一轮修改稿的研究 2 中已经报告了参数估计返真性结果（附录表 S5 和 S6）。

编委复审

尚有一些小问题，详见审改稿（请举一反三修改同类地方）。此外，文章经过两轮修改篇幅很长，建议作者将可有可无的内容删除。

回应:

感谢您的细心审阅，相关内容已经进行修改。

主编终审

经审阅，XB21-253 稿件，质量没有问题，可以刊发。

建议:

在文章结构上，请作者解释将实证研究放在模拟研究之前的原因。一般的论文结构都是先进

行模拟研究的论述，再加入实证研究。最后，文章结尾有一拼写错误：hype-prior 应为 hyper-prior。

回应：

感谢您的肯定和建议，修改稿中已经添加相关内容，并修订小错误。

“鉴于本文所提出模型中包含 φ_i 和 λ_i 两个新参数，暂缺乏对它们的取值范围的了解，难以进行恰当的模拟研究(即，不知根据何种分布来生成它们的真值)；因此需要先进行实证研究，以展现新模型的实践可应用性，并为模拟研究中参数真值生成提供参考依据。” 见第 10 页。