

《心理科学进展》审稿意见与作者回应

题目：潜变量建模视角下 AHS 的分析:方法、问题与改进

作者：毛秀珍;黄林超;徐丙慧;郑巧玉

第一轮

审稿人 1 的意见：

该综述选题具有很高的学术价值和实践意义，对属性层级结构（AHS）的分析框架进行了系统梳理。作者尝试从潜变量建模、诊断分类模型以及贝叶斯网络三个视角展开论述，涵盖面广，能够帮助研究者建立对该领域方法论的初步认知。

为了论文质量的进一步提高，我有以下具体修改建议：

意见 1：目前的章节结构（按模型类型分类）把文章切得太碎了，读起来更像是各种方法的罗列，而不是一篇深度的综述。读者看完后，往往还是不知道自己面对具体数据时到底该选哪种方法。建议你考虑重新梳理一下逻辑，不再以模型类型为主线，而是转为以“分析目的”（比如验证既有结构 vs 估计潜在结构）来组织全文。这样读者就能很直观地看到，当他们有某个具体目标时，应该从工具箱里挑出哪种方法。如果你能在文中增加一个决策路径矩阵或对比表，效果会更好，这能直接提升文章的实操价值。

回应：感谢专家的意见。我们最初基于“估计、验证和修正”的分析目的来组织文稿结构。但是当我们再多次阅读文献时，发现“估计、验证和修正”只能区分方法的目的，而难以回答“为什么某类方法适用于特定目的”。究其原因，正是其采用的模型往往决定了已知条件，从而方法的目的随之而定。撰写初稿的时候，我们尝试按照建模框架来组织内容，希望从方法本质上梳理其差异，并为未来研究提供思路。

但是，正如专家提到的“LCM/DCMs 与 BNMs 已高度融合，而不是相互割裂的关系”，同时，某一类方法的思想也并非仅适用于某一类模型框架。专家建议的文稿组织方式更有助于读者的阅读理解不同方法的思路和方法的选用逻辑。因此，修改稿根据专家的意见，重新按照分析目的调整文章的结构，并在适当地方保留基于模型框架的讨论。相应地，我们还将论文题目由“潜变量建模视角下 AHS 的分析：方法、问题与改进”修改为“潜变量建模视角下属性层级结构的检验与估计”。

具体而言，修改稿将原稿第三、四、五部分重新整合，按照“属性层级结构的检验”和“属性层级结构的估计”分别组织为第三、第四部分。第三部分聚焦 AHS 检验方法，包括“基于模型拟合比较的 AHS 检验方法”、“基于结构参数检验的 AHS 检验方法”以及“不同 AHS 检验方法评价”三个部分；第四部分聚焦 AHS 估计方法，包括“基于结构参数检验的 AHS 估计方法”、“基于贝叶斯网络结构学习的 AHS 估计方法”、“基于参数估计的 AHS 估计方法”以及“AHS 估计方法的评价”四个部分。其中，基于参数估计的 AHS 估计方法又进一步分为基于惩罚 MMLE 和基于贝叶斯参数估计的两类方法。

针对专家提出增加决策路径矩阵或对比表以提升文章实操价值的建议，我们在修改稿第五部分“讨论与展望”中补充了方法选择路径图，见图 3。

意见 2：另外，关于第 4 节和第 5 节的逻辑，我还有个强烈的建议。目前作者把 LCM 和 BNMs 当成两个完全平行的独立章节来写，但这种划分方式其实是人为地割裂了该领域发展的内在脉络。LCM/DCM 是底层的建模基础，而 BNMs 是描述属性层级结构的数学语言，

这两者在现代认知诊断研究中早就高度融合了，根本不是‘井水不犯河水’的两个领域。例如，作者在第 4 节提到‘基于正则化的参数估计’，这本质上就是目前贝叶斯网络结构学习中前沿的‘连续优化方法’。作者非要把这东西拆在两章里讲，既造成了大量的理论重复，也遮蔽了这些方法在应对‘高维属性估计’这一核心挑战时的内在统一性。建议作者要么对这两个章节进行彻底的重组，将重点转向‘方法论的融合’，要么至少需要在文中明确阐述这两者是如何在现代研究中相互渗透的，而不是让它们各立山头，这样才能给读者提供一个连贯的学术图景。

回应：首先，我们非常赞同专家的观点：“LCM/DCM 是底层的建模基础，而 BNMs 是描述属性层级结构的数学语言，这两者在现代认知诊断研究中早就高度融合了”。根据这一意见，我们在修改稿第二部分“认知诊断评估中潜变量建模”的倒数第二段，对 LCM、DCMs 和 BNMs 的关系进行补充说明，并结合新增的相关文献进一步阐述三者的联系。

其次，我们也非常认同专家指出的“造成了大量的理论重复”问题。初稿撰写过程中，我们已经感受到相关内容存在重复，但对其原因思考不够深入。专家进一步指出“遮蔽了这些方法在应对‘高维属性’估计这一核心挑战时的内在统一性”，这让我们重新理解方法本质和实践困难受到了启发。修改稿在第二部分“认知诊断评估中潜变量建模”最后一段、“4.3.1 基于惩罚 MMLE 的 AHS 估计方法”部分的最后一段和“4.4 AHS 估计方法的评价”部分的最后一段都做了分析。

这部分可能讨论不到位，还请专家指点，我们继续改进。

意见 3：在方法论深度上，我觉得还可以再挖深一点。目前对两阶段法和一阶段法的讨论还是点到为止。建议你直接点出两阶段法可能存在的“插补偏差”（即忽略了第一阶段 AMP 估计的标准误），以及一阶段法在计算上的“拦路虎”（收敛难、参数多），告诉读者在什么测量精度或样本量条件下该如何做取舍，而不是泛泛地罗列。这种针对性的探讨，对研究者来说才是“干货”。

回应：感谢专家的指点。我们再次分析了各类方法在恢复 AHS 的具体思想，并结合专家关于一阶段法与两阶段法优劣的建议，在修改稿“4.4 AHS 估计方法的评价”第二段中进行了补充分析和讨论。具体内容如下：

首先，AHS 的分析思路不同。结构学习算法采用“先估计被试 AMP 再学习结构”思路，但 AMP 与 AHS 相对独立，被试 AMP 的估计误差会影响后续 AHS 的学习，产生类似“插补偏差”的问题。假设检验和参数估计将结构参数和项目参数视作整体，要么依据结构参数估计值建构可能 AMP 空间再恢复 AHS；要么根据 AHS、AMP、项目特征与作答反应的联合后验分布来抽样估计 AHS。这两类方法同时处理 AMP、AHS 和项目参数的不确定性，但也面临参数多、模型空间大、收敛困难和计算成本高等问题。实践中应结合样本量、题目质量、计算资源和结果需求进行方法选择。

这部分可能讨论不够准确，还请专家指点，我们继续改进。

意见 4：文末那个关于“目标、内容、模型基础”的演进脉络总结得非常好，这其实是这篇综述的灵魂。建议你不要把它仅放在结尾，而是把它当作叙事的主线贯穿在全文里，引导读者感受这个领域是如何从早期的显著性检验，一路演进到现在的数据驱动联合估计的，这样文章的逻辑张力会强很多。

回应：非常感谢专家的建议。修改稿尝试将 AHS 分析方法的演进脉络当作叙事的主线贯穿全文。相关论述只要体现在引言部分倒数第二段、第二部分“认知诊断评估中潜变量建模”的最后一段、第四部分“属性层级结构的估计”的第一段、“4.4 AHS 估计方法的评价”部分最后一段，以及第五部分“结论和展望”的第一段。

意见 5: 全文的符号定义（尤其是被试和题目的索引）请务必再核对一下，保持一致性是学术表达最基本的要求

回应: 谢谢专家的意见。我们仔细阅读并检查了全文中的字母符号，并将有误之处和索引不一致的地方进行了修改。

第二部分第二段中“作答项目 j 的概率向量 $\Theta_j=(\theta_{j1},\theta_{j2},...,\theta_{jm})$ ”改为了“作答项目 j 的概率向量 $\Theta_j=(\theta_{j1},\theta_{j2},...,\theta_{jM})$ ”；公式(1)“ $l(\Theta,\pi|X)=\sum_{i=1}^I \log[\sum_{m=1}^M \pi_m \prod_{j=1}^J \theta_{jm}^{x_{ij}}(1-\theta_{jm})^{1-x_{ij}}]$ ”中第一个求和符号的上限改为了 I ；2.3 部分第一段末尾关于属性节点的表示均修改为了 N_k 。

.....

审稿人 2 意见:

意见 1: 关键词中有错别字。

回应: 感谢专家的意见。修改稿已将初稿关键词中的“性层级结构”更正为“属性层级结构”，并对关键词的顺序进行调整。

意见 2: 论文仍处于“修订”状态，有批注存在。

回应: 谢谢评阅专家的意见。第一次投稿前，我 Word 中执行了“接受所有修订”的操作，但文档左侧始终有一条竖线。当时我在修订模式下多次尝试未能取消，后来在学生帮助下取消了该竖线，便误以为问题已经解决就上传了稿件。

当我重新下载上传的 Word 版稿件，用手提电脑打开时没有显示任何批注，但是我在台式电脑上打开的时候却显示了批注。经检查，我发现笔记本电脑的审阅选项为“修订模式+显示最终状态”，台式电脑则为“修订模式+显示标记的最终状态”。即使切换为“非修订模式+显示标记的最终状态”，文档仍会显示批注。由此我才意识到，此前并未彻底清除修订标记。正确操作应是在修订模式下接受所有修订，取消修订选项中的左右侧竖线等标记，再切换至非修订模式并保存文档。

第一次上传稿件中仍残留批注，主要是我对 Word 修订模式操作不当所致。感谢专家指出这一问题，今后我会更加耐心、细致地检查稿件修订状态，避免类似问题再次发生。

意见 3: 文章很多公式都是嵌套在段落中，需要单独分行，不然阅读起来很困难。

回应: 谢谢专家的意见。我们在修改稿中将核心公式做了单独分行处理。详见修改稿中公式(1)到公式(7)的内容。

意见 4: 摘要中写“参数估计、假设检验和贝叶斯结构学习三个视角”，但是正文的第 2-3 节并不是这个结构。

回应: 感谢专家的意见。第一位专家也指出了文章结构方面的问题。经认真思考后，我们对论文结构进行了调整：以 AHS 分析方法的目的一级分析框架，以“参数估计、假设检验和贝叶斯结构学习”三个视角作为二级分析框架，重新组织了第三部分和第四部分内容。

具体而言，修改稿将原稿第三、四、五部分重新整合，按照“属性层级结构的检验”和“属性层级结构的估计”分别组织为第三、第四部分。第三部分聚焦 AHS 检验方法，包括“基于模型拟合比较的 AHS 检验方法”、“基于结构参数检验的 AHS 检验方法”以及“不同 AHS 检验方法评价”三个部分；第四部分聚焦 AHS 估计方法，包括“基于结构参数检验的 AHS 估计方法”、“基于贝叶斯网络结构学习的 AHS 估计方法”、“基于参数估计的 AHS 估

计方法”以及“AHS 估计方法的评价”四个部分。其中，基于参数估计的 AHS 估计方法又进一步分为基于惩罚 MMLE 和基于贝叶斯参数估计的两类方法。

意见 5：目前稿件侧重于“如何获得”AHS，但对如何评价所获 AHS 的优劣讨论较少。

回应：感谢专家的建议。初稿确实主要围绕如何获得 AHS 以及各研究所报告的 AHS 判准率展开，对估计所得 AHS 的优劣评价讨论不足。修改稿已在“4.4 AHS 估计方法的评价”第三段中对此进行了分析和讨论。对应修改内容如下：

其次，对 AHS 的评价方式不同。假设检验方法重在报告每一类 AMP 的一类错误控制率和统计检验力(Zhang et al., 2025b)；贝叶斯结构学习算法则多报告 AHS 估计中平均多弧、少弧和逆向弧数量(喻晓峰 等, 2021; 马奕帆, 2021)；惩罚 MMLE 方法中关注 AMP 中正确偏序的判准率(Ma et al., 2022)；贝叶斯参数估计中除报告了 AHS 中各边的后验纳入概率、估计 AHS 属于真实结构子图的比例以及估计 AHS 包含错误边或错误方向的比例(Chen & Wang, 2023) 外, Wang 等人(2026)还报告了邻接矩阵和可达矩阵中真实直接边的检出率 and 不存在直接边的排除率。这些指标除了反映方法特征外，还从不同侧面评价了 AHS 估计结果的错误类型和结果稳定性。实践中应选择多维评价指标，以全面分析结果优劣。

修改稿若有对专家意见理解不够准确之处，还请专家批评指正，我们将继续修改完善。

意见 6：文中多次提到 RLCA、LVS、PLA 以及 K2、HC 等算法对无结构型或复杂结构的判准率较低。未进一步讨论其原因。

回应：初稿中的这句话“RLCM 和 LVS 对线性和无结构型的判准率更低”，我们重新查阅文献后发现，RLCM 方法表现虽远不及 LVS 方法，但是 LVS 方法在不同 AHS 类型下的结果相似。根据专家的意见，修改稿在“4.4 AHS 估计方法的评价”中第三段对“对不同 AHS 类型的返真性结果不同”进行了分析和讨论。对应修改内容如下：

再次，对不同 AHS 类型的返真性结果不同。迭代 Z 检验中对线性结构的一类错误率更高；惩罚 MMLE 方法中，PLA 对无结构型的恢复略差；结构学习算法中，K2、HC 和 MMHC 在收敛型或无结构型的表现更差；贝叶斯参数估计中，Wang 等人(2026)提出的联合贝叶斯估计方法在线性结构上恢复能力相对较弱。其原因可能是：Z 检验方法易受 AMP 的数量和结构参数标准误的影响；正则化参数估计方法易受 AMP 数量的影响；结构学习算法易受结构复杂度和搜索策略的影响；而贝叶斯联合估计方法则易受项目考察的属性模式和被试 AMP 分布均衡度的影响。当然，研究展现出的差异和原因还有待进一步比较分析。

修改稿若有对专家意见理解不够准确之处，恳请专家批评指正，我们将继续修改完善。

意见 7：作者在展望中分了 3 个小节来描述未来的方向，这太冗余了，建议凝练一些，写一些主要的方向。

回应：谢谢专家的意见，初稿的展望部分确实存在过于详细，又显得不够深入。我们对该部分进行了重新梳理和凝练，删除了原稿中过于详细的表述，并增加了一些新的思考。

意见 8：文献综述不够全面，有一些最近的研究成果没有收录，需要更全面和综合地对关键研究进行综述。

回应：谢谢专家的意见。我们进一步查阅文献，补充了 6 篇文献，包括 1 篇 AHS 的最新研究成果，即 Wang 等人(2026)；2 篇从数据挖掘和机器学习角度探索 AHS 的文献，即 Chen 等人(2015)和 Chen 等人(2016)；以及 3 篇围绕 AHS 条件概率参数开展认知诊断建模的文献，即 Zhan 等人(2020)、Zhan 和 He(2021)和 Li 等人(2022)。修改内容请参见修改稿第二部分“认知诊断评估中潜变量建模”的倒数第二段、“4.3.2 基于贝叶斯参数估计的 AHS 估计方法”

部分，以及“5. 结论与展望”的最后一段。

第二轮

审稿人 1 意见：作者已经回答了我的所有问题。我没有更多问题了。

回应：非常感谢专家对修改内容的肯定。我们对第二版修改稿再次进行了阅读校阅，使逻辑更清楚，表达更准确、流畅和简洁。

审稿人 2 意见：

作者在修改后，文章的质量有了较大的提高，基本解决了审稿人前面提出的多数问题，但是对于文献的把握仍存在不足，比如下面的两篇高度相关的文章，作者都没有纳入综述。颜玉枝. (2022). 排序理论在认知诊断中辅助标定属性层级关系 (硕士学位论文), 江西师范大学.

Yan, Y. Z., Dong, S. H., & Yu, X. F. (2025). Using ordering theory to learn attribute hierarchies from examinees' attribute profiles. *Journal of Educational and Behavioral Statistics*, 50(6), 773-796.

建议作者对领域的最新研究进行认真检索，确保文章能体现最新的研究进展和水平。

回应：非常感谢专家对修改稿质量提升给予肯定，并指出仍存在的不足之处。感谢专家的意见，在今后的研究中一定更加注意权威文献检索的全面性。根据专家指出的重要文献，我们进行了认真阅读和梳理，并在第二版修改稿中作了多处补充和完善。具体情况如下：

第一部分，倒数第二段段首补充了 AHS 分析方法的分类方法，具体内容为：

按照是否依赖模型，可以分为参数化和非参数化方法；按照分析的数据依据可分为基于 Q 矩阵(Tatsuoka,1995)、作答数据和被试 AMP 的方法。

第四部分中“4.2 基于贝叶斯网络结构学习的 AHS 估计方法”的最后一段中补充了颜玉枝(2022)和 Yan 等人(2025)的研究，具体内容为：

这些方法都可以采用参数化或非参数方法估计 AMP，并将其作为观测数据来分析 AHS。类似地，颜玉枝(2022)基于排序理论提出，首先根据被试 AMP 计算属性相关强度矩阵 (Attribute Correlation Intensity Matrix, ACIM)，用于表征“未掌握前提属性却掌握后续属性的比例”；然后，基于 ACIM 计算层级指标矩阵用于表征属性所在的层级顺序；最后依据临界值判断不同层之间是单独成层还是合并层级，在此基础上恢复属性层级结构，简称为 ACIM 方法。ACIM 方法可不依赖 DCMs，属于非参数化方法。该方法不受样本量影响、计算简单，在小样本(50 人)中同样表现较好(颜玉枝, 2022; Yan et al., 2025)。另外，Yan 等人(2025)Y 研究表明，ACIM 方法在不同实验条件下对 AHS 的判准率和计算效率都显著优于 MMHC 方法。

第四部分中“4.4 AHS 估计方法的评价”的第三段增加了颜玉枝(2022)和 Yan 等人(2025)采用的评价指标；最后一段增加了颜玉枝(2022)提出的 ACIM 方法作为非参数方法的代表对今后研究的启示。

第三轮

编委 1 意见：该文经过两轮审稿与修改，已经达到发表水平。有几处小问题请考虑修改：

意见 1: 评估通常是与测验并列的概念。如果这里说的是对测验的评估，原文没有问题；如果是对教学的评估，摘要中“在测验设计与评估、教学设计”改成“在测验设计、教学设计与评估”。

回应: 非常感谢编委专家对修改稿的肯定。摘要部分的原文是“属性层级结构(Attribute Hierarchical Structure, AHS)刻画了属性之间的先决关系，在测验设计与评估、教学设计、学习路径规划和个性化学习推荐方面具有重要意义。”经过专家的提示和思考，原文是想表达“AHS 对测验设计、教学设计和教学评估方面具有重要意义”。于是，根据专家的意见，我们对摘要部分的表达进行了修改。

意见 2: 公式 (1)， $\log$ 是函数符号，用正体。其他函数符号类似。

回应: 非常感谢编委专家指出修改稿中函数符号表示的不规范之处。我们统一了修改稿中相关函数的表示，并进行了检查和修改。

意见 3: 公式 (2)，乘积符号与求和符号一样用正体。

回应: 非常感谢编委专家指出的不足。我们已将公式(2)中的乘积符号修改为正体，并对修改稿中所有公式进行检查和修改。

意见 4: 表示矩阵或向量的英文字母用粗斜体，希腊字母用粗正体。

回应: 非常感谢编委专家指出修改稿中矩阵和向量表示的不规范之处。我们已对修改稿中相关内容进行检查和修改。

编委 2 意见: 同意两位审稿专家的意见，作者对论文做了较大程度的修改，修改后达到发表水平，建议录用。

回应: 非常感谢编委专家的审阅和对修改稿的肯定。

主编意见: 同意发表。