《心理学报》审稿意见与作者回应

题目：认知诊断测评中缺失数据的处理：随机森林阈值插补法
作者：游晓锋，杨建芹，秦春影，刘红云

第一轮

审稿人 1 意见：

认知诊断评价中经常会遇到缺失值的问题，因此探讨缺失值的处理问题的研究有一定的理论及现实价值。本文在 DINA 框架下对已有随机森林插补法（RFI）进行改进，提出采用个人拟合指标（RCI）确定插补阈值的随机森林阈值插补方法（RFTI）。

综合而言，本文的创新型在于提出了新的认知诊断评估中缺失数据的处理方法 RFTI，并且发现与 RFI 和 EM 方法相比，RFTI 在被试属性模式判准率和边际判准率上表现出明显优势，创新性、理论意义及应用价值高。建议大修后再审。

意见 1：主要问题。（1）作者在文中以 DINA 模型为基础提出使用将个人拟合指数 RCI 应用于动态阈值的确定，但是审稿人认为在 CDM 实践应用中，即便是同一个测验中的各个测验项目的合适的拟合模型也可能是不同的，例如，在同一个测验中，可能有些项目适合 DINA，有些项目适合 A-CDM，有些项目可能还适合饱和的 GDINA 模型等等（如，Liu et al., 2019）。那么，在其他模型中，作者提出的这个方法的表现是怎么样的？（2）实践中，Q 矩阵中的元素有可能存在错误设定（例如，刘彦楼 等, 2019），在这种情况下，作者提出的方法的表现如何，与其他方法相比可能的优势是什么？（3）在其他非理想条件下（如，Liu et al., 2021），新方法的可能表现是什么？

Liu, Y., Xin, T., & Jiang, Y. (2021). Structural Parameter Standard Error Estimation Method in Diagnostic Classification Models: Estimation and Application. Multivariate behavioral research.

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, 1251-1259.

Liu, Y., Andersson, B., Xin, T., Zhang, H., & Wang, L. (2019). Improved Wald Statistics for Item-Level Model Comparison in Diagnostic Classification Models. Applied Psychological Measurement, 43, 402-414.

回应：非常同意审稿专家提到的这几点。对于认知诊断模型缺失数据的处理也和其他认知诊断分析的结果类似，的确会受到所选择的模型、Q 矩阵定义是否正确等因素的影响。在修改稿中根据您的建议，我们对内容作了相应的修改。具体修改如下：

（1）由于个人拟合指数 RCI 适用于所有明确定义项目反应函数的认知诊断模型，因此我们提出的将个人拟合指数 RCI 用于动态阈值确定的思想可以推广到其他认知诊断模型或同一套测验中不同题目拟合模型不一致的情况。我们在文章中 2.3 部分对这一点进行了进一步的阐述。具体见 P9 页 3 段和 P11 页 4 段。

（2）我们同意审稿专家指出的认知诊断测验中 Q 矩阵设定是否正确的重要性，以及探讨 Q 矩阵设定错误情况下不同缺失数据处理方法表现的价值和意义。但是由于本研究的主要目的是提出随机森林阈值插补方法以解决随机缺失机制复杂和随机缺失比例高的问题，因此，作为初步的研究我们基于 Q 矩阵正确设定的前提假设，而没有考虑 Q 矩阵设定错误的情况。我们在讨论部分就这一局限性和未来研究方向进行了解释和说明，同时也就认知诊断测验中可

能出现的一些特殊情况进行了说明。详见讨论部分 5.2 的研究局限性与展望。

意见 2：文中，$\eta_{ij}^{Q_t}=\prod_{k=1}^{K}\alpha_{ik}^{q_{jk}}$ 可以直接表达为$\eta_{ij}=\prod_{k=1}^{K}\alpha_{ik}^{q_{jk}}$吗？其中，$Q_t$为测验蓝图或测验的 Q 矩阵，t 是什么？

回应：感谢审稿专家的仔细审阅，原稿中$\eta_{ij}^{Q_t}$表述有误。已经在修改稿中做了修改，并补充了对应的参考文献，详见 P7 方程（2）以及对应的解释。所补充参考文献如下：
丁树良, 汪文义, 罗芬. (2012). 认知诊断中Q矩阵和Q矩阵理论. 江西师范大学学报(自然科学版), 36(5), 441–445.
Ding,S.L., Wang W.Y.,Luo,F.,Q Matrix and Q Matrix Theory in Cognitive Diagnostic. Journal of Jiangxi Normal University Natural Sciences, 36(5), 441–445.

意见 3：建议规范数学符号。（1）α_i表示知识掌握状态向量或属性掌握模式向量，可以表达为$\boldsymbol{\alpha}_i$。（2）$i_{mis}^{(s)}\in\{1,2,\dots n\}$中 i 是被试，还是被试作答数据集？（3）m 在研究中具体指的是什么，按照我的理解应该是项目数吧？

回应：感谢审稿人仔细审阅，修改稿中仔细检查了符号表示，并加了必要说明。
（1）将向量α_i加粗表示；

（2）$i_{mis}^{(s)}\in\{1,2,\dots n\}$中 i 指的是被试，$i_{mis}^{(s)}$是被试 1, 2, ⋯, n 组成的集合，描述了在变量X_s上存在缺失数据的被试集合。原稿中存在笔误，在修改稿中已经作了修改，详见 2.2 小节中的第 2 段。
（3）m 指测验包含的项目个数，在修改稿中作了补充说明。详见 2.2 小节中的第 2 段。

意见 4：作者在文中陈述“采用传统的缺失数据插补方法，如均数插补法计算X中所有缺失值的初值，然后按照缺失值的数量升序将所有含缺失的变量X_s进行排序，得到的结果的矩阵记为X_{old}^{imp}。”，然后又说“直到最新的一次插补结果X_{new}^{imp}与上一次的插补结果X_{old}^{imp}相比不再变化或变化很小（达到收敛指标）时停止”，这让审稿人比较迷惑，是将采用传统的缺失数据插补方法获得的X_{old}^{imp}作为初始值，然后使用随机森林算法依次迭代吗？
回应：是的，在随机森林阈值插补法实现的过程中，迭代的初始值采用传统的缺失数据插补方法获得的X_{old}^{imp}，然后使用随机森林算法依次迭代。原文表述的确存在容易混淆的地方，在修改稿中我们进一步明确了迭代初始值以及估计值更新的过程，对这一段的表述作了修改。详见 2.2 中关于插补步骤的修改。

意见 5：公式$\Delta_F=\frac{\sum_{j\in F}\sum_{i=1}^{n}I_{X_{new}^{imp}\neq X_{old}^{imp}}}{\#NA}$中的 F 指的是什么？
回应：F表示两次迭代中被插补数据的集合，Δ_F描述了前后两次插补值变化的个数占总缺失数据个数的比例，值越小表示两次迭代得到插补数据集的差异越小。

意见 6：公式（4）中的$I_j(\alpha_i)$是否就是η_{ij}，或者是表达为$\hat{\eta}_{ij}$?
回应：由于公式（4）不仅适用于 DINA 模型，广义而言，$I_j(\alpha_i)$为属性掌握模式为$\boldsymbol{\alpha}_i$的被试在试题 j 上的理想作答反应。但是对于本文关注的 DINA 模型，$I_j(\alpha_i)$即为η_{ij}。在修改稿中

我们进一步明确了这一点，具体参见 P9 中第 2 段。

意见 7：公式（5）中的 na 是否可以统一表达为 NA？
回应：已修改

意见 8：文中“可以得到 K 个不同的阈值（例如，当 δ=0.01 时，$K = 50$）。根据不同阈值$\tau^{(k)}$（$k = 1,2,\ldots,K$）插补，得到 K 个插补后的作答矩阵$X^{(k)}$（$k = 1,2,\ldots,K$）。”，K 具体是指什么，k 指的是属性 k 吗？
回应：这里的 K 不是属性个数，指的是数据插补过程中指定变化步长 δ 对应的阈值个数。为了避免可能引起的误解，在修改稿中将 K 换成 T。具体修改见 2.3.2 最后一段。

意见 9：“假设属性间关系是相互独立的”指的是“假设属性之间不存在层级”吧？
回应：是的，原文中指的是属性之间不存在假设的层级结构，为了避免混淆，修改稿中将其修改为：假设属性间不存在层级关系。

意见 10：“试题属性分配方式是随机的。”“测验 Q 矩阵（随机产生）”是每次重复都随机产生还是在重复前随机产生，然后在每次重复中固定使用先前随机产生的？另外，Q 矩阵随机产生的条件是什么，如何保证模型的可识别性？
回应：谢谢审稿专家指出这一不严谨的表述。已修改为“假设属性间不存在层级关系”。详细参见：章节 3.1 的最后一段。另外，原文说的随机是指每次都重新生成 Q 矩阵，但是对每一次 Q 矩阵生成均需保证模型可识别。具体生成方法如下：在设定 Q 矩阵时，我们根据 Q 矩阵完备性的假设以及 Xu & Zhang (2016)的研究，首先随机生成只测量 K 个属性中某个单一属性的 K 道项目，即在 Q 矩阵中存在一个 K×K 的单位矩阵；然后测量每个属性的项目个数至少为 3 个，以保证模型的可识别性”。在修改稿中我们补充了这一信息并补充了相应的参考文献。详细参见 3.2.2。

意见 11：被试知识状态是通过多元正态阈值模型生成，因此“Φ^{-1}为标准正态分布的累积分布函数的逆函数”这个表达不恰当，应该是多元正态分布，需要修改。另外，$(\frac{k}{K+1})$指的是什么，请作者解释一下？
回应：已修改，见修改稿中公式 7 的下面的说明。具体修改见 3.2.1 第一段。

意见 12：通过“被试知识状态”生成部分可以发现属性之间是相关的，而不是“属性间关系是相互独立的”，因为协方差矩阵中的部分元素等于 0.5。
回应：谢谢审稿专家指出这一表述不严谨之处，的确属性之间是相关的，但是不存在指定的属性层级关系。在修改稿中已经修改了表述。详见 3.1 小节。

意见 13：DINA 模型中的猜测参数 g 和失误参数 s 均从均匀分布中抽取，是每次重复都随机产生，还是在重复前随机产生然后在每次重复中固定使用先前随机产生的值？
回应：每次重复都随机产生，在修改稿中进一步明确了这一点。

意见 14：公式 8、9、10、11 中，“$\sum_{r=1}^{R}$ ”中的 k=1 到 R 是什么意思？
回应：感谢 仔细审阅，原文存在公式中符号说明不清楚和前后不一致的问题。修改稿中公式 8、9、10、11 中的 R 表示独立重复模拟的次数，已经在修改稿中做了进一步说明。

意见 15：文中，属性是用 k 表示还是 s 表示，以及“X_s为任意一个可能存在缺失值的变量”建议符号统一起来，还有就是，前面用 n 表示被试的个数，后面部分用“N 表示被试的人数”，其他部分也是类似的。
回应：类似的情况都已经修改了

意见 16：项目参数的大小是不同的，可以求平均的 Bias 吗？
回应：感谢审稿专家指出这一点。本研究中主要是对不同方法之间的估计精度进行比较，所有项目参数估计偏差的平均可以在一定程度上描述整体上这种方法估计的精度。根据您的问题，我们也仔细检查了单个题目估计偏差在不同方法之间的差异，其结果与平均后结果一致。因此，由于不同方法对应的测验项目相同，平均后的结果对于不同方法的比较是公平的。为了表述方便和节约篇幅，我们在文中呈现了所有项目平均估计的偏差。同时，这种将测验中所有项目参数估计结果平均的方法，在测量和统计方法的文章中也非常常见（例如，Han, Liu, Ji, 2021）。修改稿中我们对这一点进行了说明，详见 4.3 小节的修改。

意见 17：按照我的理解，项目参数估计越准确，模式判准率越高。但是我不明白作者给出“如果研究者关注的重点是被试知识状态的估计，这也往往是认知诊断测验本身要解决的问题，是实际应用关注的焦点，我们推荐使用新提出的 RFTI 方法；但是如果研究者的目的是对项目参数进行准确估计，如建立题库等，这一方法的使用则要相当慎重，我们则推荐采用 EM 方法。”这个结论的原因是什么？
回应：这里主要讨论的是存在缺失数据的认知诊断评估，基于缺失数据的认知诊断评估比基于完整数据认知诊断评估往往更加复杂，在估计项目参数和被试知识状态前，需要对缺失数据进行插补。但是基于本研究的结果，基于不同插补方法的插补数据集，再作项目参数估计和被试知识状态估计，结论并不一致。所以我们针对感兴趣问题的不同，做出了这一推断。在修改稿中我们对这一原因参考已有的研究进行了简要的解释分析。详见讨论 5.1 小节中的最后一段。

审稿人 2 意见：

本研究是对 RFI 方法的应用拓展， 即使用“穷举”法依次尝试比 0.5 更适合的切点将缺失值转换为 1。 本研究对认知诊断缺失值数据处理领域具有一定的方法学贡献， 但研究设计有一些不足需要修改且研究结果的一些“不合理”。 审阅后我有一些建议， 仅供作者参考。

意见 1：引言部分，1. CDA 翻译为“认知诊断测评”更符合其本质含义；英文单词首字母不需要大写；
回应：谢谢审稿专家建议，已经按照您的建议进行了修改。

意见 2：研究 1：3.1 中“属性间关系是相互独立的”是说统计独立还是结构独立？ 请准确表达
回应：已修改为“假设属性间不存在层级关系”。详细参见：章节 3.1 的最后一段。

意见 3：不能采用随机方式生成 Q 矩阵， 对 DINA 模型而言， Q 矩阵的设定需要保证模型的可识别性， 具体参考 Xu & Zhang (2016)等相关研究；
回应：谢谢审稿专家指出这一不严谨的表述，原文说的随机是指每次都重新生成 Q 矩阵，

但是对每一次 Q 矩阵生成均需保证模型可识别。具体生成方法如下：在设定 Q 矩阵时，我们根据 Q 矩阵完备性的假设以及 Xu & Zhang (2016)的研究，首先随机生成只测量 K 个属性中某个单一属性的 K 道项目，即在 Q 矩阵中存在一个 K×K 的单位矩阵；然后测量每个属性的项目个数至少为 3 个，以保证模型的可识别性”。在修改稿中我们补充了这一信息并补充了相应的参考文献。详细参见 3.2.2。

意见 4：为保证研究结果与实际相符， 需使用二元正态分布生成 DINA 模型中的 s 与 g 参数， 具体参考 Zhan, Jiao, Liao, & Bian (2019)等相关研究；
回应：感谢审稿专家的建议。本研究考虑的条件是假设 DINA 模型中 s 与 g 参数相互独立，所以在生成数据时采用分别生成 s 参数和 g 参数的方法。这一生成数据的方法也是认知诊断模型中缺失数据处理常用的方法（e.g., Dai，2017）。 当然，我们同意考虑认知诊断模型中项目参数相依的意义。在未来研究中可以考虑项目参数特征不独立的情况（Zhan, Jiao, Liao, & Bian，2019)的相关研究。我们在讨论部分对这一点进行了说明并补充了相应的参考文献。

意见 5：3.2.3 节中， 如果“实际中完整的训练数据集并非必须的”， 那为何作者要使用完整的训练数据集？ 作者需要明确解释每一个步骤的设定和操作理由。
回应：如果没有完整的训练集数据，在已知项目参数的情况下可以通过项目参数生成模拟的完整数据集；在未知项目参数的情况下，可以直接基于现有的作答模式，模拟生成符合作答模式的完整数据训练集。由于本研究关注的主要问题是认知诊断模型中缺失数据的处理，为了控制其他因素的干扰，在模拟研究中假设 20%的数据是完整的，这一假设在实际测试中也不难满足。这一领域的研究还可以参考随机森林方法的相关文献（e.g., Stekhoven, 2012）。

意见 6：结果解释部分作者需要谨慎， 4 种缺失类型的生成方式不同， 尽管它们可以有相同的缺失比例， 但同一比例下每一种缺失类型的生成方法都可以变化(作者只是用了一种)。因此， 作者在结果解释时应谨慎对待不同缺失机制下的对比；
回应：我们同意您所说的缺失数据生成的方法非常多，生成方式的不同有可能会影响结果的判断。但是在模拟研究中，往往会基于某一假设模式生成不同机制的缺失数据，不仅在认知诊断模型中如此，而且在项目反应理论模型中也是这样的。这样做的好处是容易解释清楚数据缺失机制特点的影响。缺失数据生成模式我们参考了这一领域的相关研究，具体见 Dai（2017）。

意见 7：研究 1 中每个条件下生成几批数据？
回应：每一种条件下生成 100 批模拟数据，修改稿中进一步明确了这一点。

意见 8：研究 2，Bias 和 RMSE 的公式错误， 参数缺少迭代(循环)次数角标； 另外， 请注意， RMSE 和 bias 的计算是针对单独某一个参数的（https://www.geeksforgeeks.org/root-mean-square-error-in-r-programming）， 而非某一个参数在所有题目上的平均。 正确的公式应为 bias(其中 R 是中迭代次数。 而当作者想报告所有题目的 bias 和 RMSE 时， 需要再对每道题目的 bias 和 RMSE 再求均值， 即均 bias 和均 RMSE。
回应：结合审稿专家一的建议，仔细检查了这部分的结果，已做相应修改。也核对了分析的结果，是原文表述记号有误，具体计算是每个项目计算 Bias 和 RMSE 后的平均。

意见 9：参数估计是自编程序还是使用了软件包？ 与已有研究相比，研究 2 中 PMR 偏低，

比如，当缺失值仅有 10%时，PMR 最高连 0.6 都不到；这是一个不符合常理的结果，不知作者是否思考过原因？

回应：采用 R 软件中 CDM 包进行的参数估计。我们仔细检查了程序也核对了结果，并查找了以往类似研究，我们的结果与 Dai(2017)关于 MNAR 机制下采用 EM 算法插补的结果基本一致（详见 Dai(2017)研究的 P106 页）。

意见 10：另外， 为何 MAR 和 MCAR 的 PMR 相对低于另外两种的？是数据生成原因导致的， 还是插补方法导致的，缺失数据机制导致的？作者在引言及讨论部分都提及已有方法(EM)可能会受到缺失机制的影响，但研究结果(图 1)似乎并没有发现该问题，即所有 3 种插补方法在 MIXED 和 MNAR 上的 PMR 均优于 MAR 和 MCAR 的。

回应：我们在修改稿中参考了更多关于分类数据缺失值处理的文献，并对这一看似奇怪的结果进行了解释和说明。详见讨论部分第三段的内容。

意见 11：图 1/2 和表 4 的信息重复， 呈现一种即可；

回应：修改稿中删掉了图 1 和图 2。

意见 12：其他问题：1．公式 5 中缺失值通常用大写 NA 表示；

回应：已修改。

……

审稿人 3 意见：

本研究尝试将机器学习中随机森林缺失数据的插补（RFI）方法应用于 DINA 模型，提出了一种基于 DINA 模型中的个人拟合指标 RCI 来动态确定阈值的新方法，即随机森林阈值插补方法（RFTI）。该方法实现了缺失数据插补过程中，机器学习方法与认知诊断模型的结合应用，正确率和插补率的结果证实了这是一种有效的动态选择阈值的方法。研究具有一定的理论意义和实践价值，还有以下问题和作者探讨。

回应：感谢审稿专家对本研究意义和价值的肯定，根据您的宝贵意见和建议，我们注意做了详细的思考和修改。具体回复如下。

意见 1：随机森林插补法处理的问题什么？是对具有缺失数值的情况下，替换缺失值后，对项目参数和被试状态进行估计吗？考察这种缺失值方法的表现就是判断缺失值替换的精度和模型参数估计的精度，是吗？如果是，摘要中的结果为什么会出现缺失值替换精度高，反而项目参数估计精度不高的情况？摘要与自检报告中研究亮点 1 的回答有矛盾之处。

回应：随机森林插补法在处理缺失值时包含两个阶段，首先是按照随机森林方法对缺失数据的取值进行估计，进而根据一定的阈值将其进行插补；然后再基于插补后的数据集，结合认知诊断模型进行项目参数和被试状态进行估计。对于认知诊断测评的应用和实践，评估其效果主要指标是最后的被试知识状态估计的精度；有时也会关注项目参数估计精度，例如 CD-CAT 题库的开发。

作为一种新提出的方法，我们用两个研究同时关注了两个阶段的结果。即：在摘要中有一句话“RFTI 在插补正确率上明显高于 RFI 方法”，以及研究一探讨的问题是关注 RFTI 的插补正确率。其主要原因有：（1）本研究的重点是在 RFI 的基础上提出了动态阈值插补的思想，其出发点是为了解决 RFI 插补正确率过低的问题。因此，想要验证与 RFI 方法相比，动态阈值的处理是否可以提高插补的正确率就是首先要回答的问题。因此，我们单独作了研究一以检验在 RFI 基础上提出的 RFTI 方法是否符合我们的预期。（2）由于新的动态阈值的

缺失数据插补方法是一种非完整的非参数插补方法，即 RFTI 插补后还会存在没有被插补的缺失值，如果这一比例很大，基于以往研究这一方法在实际中的表现也会存在问题。为了检验这一点，研究一中也同时检查了 RFTI 插补后的缺失率（表 3）。这些研究结果有利于帮助读者更好理解新方法的特点，以及解释为什么 RFTI 方法的表现会由于 RFI 方法。

另外，我们对自检报告也进行了进一步修改。

意见 2：引言第三段提到实际中缺失比例高并不少见，建议补补充相应的例子或证据来支撑说明。

回应：按照您的建议，补充了相关的信息和文献。详见引言第三段蓝色字体, 补充了以下主要文献。

Graham, J.W., Taylor, B.J., Olchowski, A.E., Cumsille, P. E. (2006) Planned missing data designs in psychological research. *Psychological Methods.* 11:323–343.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research*, 29, 409–454.

Johnson, E. G. (1992). The design of the National Assessment of Educational Progress. *Journal of Educational Measurement*, 29, 95–110.

意见 3：引言中介绍了缺失值处理研究中基于 DINA 模型下 EM 表现更好，GDINA 模型的相关研究中 EM 和 FIML 表现较好，本研究二和 EM 和 RFI 进行比较，为什么没有和 FIML 方法比较，建议在修改稿中增加 FIML 的实验结果。

回应：感谢审稿专家提出这一关键问题。我们重新梳理了本部分的研究，修改了引言部分的表述。本文中我们没有对 FIML 进行探讨的最主要原因是已有研究发现，较好的缺失数据处理方法 EM、MI、FIML 和 ZR 等方法中，EM 的表现是最好的（宋枝璘, 郭磊, 郑天鹏，2022）或； Dai，& Valdivia, 2022 的研究发现，DINA 模型中 EM 与 FIML 结果差异不大。因此，在本研究中，我们选择了 EM 方法与我们提出的新方法进行比较。详细修改见引言的第二段。本段论述中增加的参考文献如下：

Dai, S.; Svetina Valdivia, D. (2022). Dealing with missing responses in cognitive diagnostic modeling. *Psych* , 4, 318–342. https://doi.org/10.3390/psych4020028.

Song, Z. L., Guo, L., & Zheng, T. P. (2022). Comparison of missing data handling methods in cognitive diagnosis: Zero replacement, multiple imputation and maximum likelihood estimation. *Acta Psychologica Sinica*, *54*(4), 426-440.

[宋枝璘, 郭磊, 郑天鹏. (2022). 认知诊断缺失数据处理方法的比较：零替换、多重插补与极大似然估计法. *心理学报*, 54(4), 426-440.

意见 4：对于缺失值的处理方法为什么要在不同缺失机制和缺失比例下比较 RFI 和 RFTI 缺失值的插补精度，而又在第二个而研究中比较 RFTI 和 EM、RFI 在知识状态和项目参数估计精度的表现？换句话说，缺失值的替换和参数估计是不是应该在相同的实验条件下进行呢？还有，具有缺失值的处理问题中是不是分为了两类，一类是缺失值的替换方法，一类是不替换缺失值直接进行参数估计。这样，本研究才在两个模拟实验下进行比较的原因？如果是，需要在文献综述中说清楚，如果不是，需要解释开展两个独立实验的原因。

回应：感谢审稿专家的提问，类似您所提出的第 1 个问题，本文涉及的 RFTI 和 EM、RFI 三种缺失数据插补方法都是属于缺失值替换方法一类。分两个研究的原因见问题 1 的回答。

意见 5：文中提到的 DINA 模型的理论和应用研究都是 2015 年前的文献，而且目前 CDM 越

来越多，越来越一般化，DINA 是 GDINA 的特殊情况，近年来关于 CDM 的研究都倾向于在更一般的模型下开展，那么本文的方法是否只适用于 DINA 模型，如果不是如何推广到其他模型，有必要在文稿中进行说明。

回应：非常感谢您提出的关于方法可拓展性的问题。在修改稿中我们补充和丰富了本领域的相关文献，研究所提出的方法是可以拓展到任意具有明确项目反应函数定义的认知诊断模型的。在修改稿中我们明确了新方法的适用范围和拓展性。具体修改见 2.3.1，2.3.2 以及 5.2 的相关内容。

意见 6：被试的作答数据庥，X_s为任意一个可能存在缺失值的变量，$i_{mis}^{(s)} \in \{1,2,\ldots n\}$为$X_s$上包含缺失值的被试作答数据集。可以将其分成 4 部分：

图片中，黄色背景部分达标不准确，前者是一个数，后者是一个集合或者向量。

回应：感谢审稿专家的仔细审阅，原文表示有误。$i_{mis}^{(s)} \in \{1,2,\ldots n\}$中 i 指的是被试，$i_{mis}^{(s)}$是被试 1,2，…,n 组成的集合，描述了在变量X_s上存在缺失数据的被试集合。原稿中存在笔误，在修改稿中已经作了修改，详见 2.2 小节中的第 2 段。

意见 7：对下面黄色背景后面加上一句“依次对缺失值升序排序后的缺失变量进行插补替换”

其次，对于每一个变量$\mathrm{X_s}$，使用随机森林算法对缺失数据进行插补，分为如下两步：第一步，用因变量$\mathrm{y_{obs}^{(s)}}$和自变量$\mathrm{x_{obs}^{(s)}}$训练出一个 y~x 的随机森林模型；第二步，将$\mathrm{x_{mis}^{(s)}}$作为特征变量输入，用训练出的随机森林模型预测缺失值$\mathrm{y_{mis}^{(s)}}$。对所有$\mathrm{X_s}$预测插补完成后，对所有$\mathrm{X_s}$预测插补完成后，所得到的矩阵记为$\mathrm{X_{new}^{imp}}$。

回应：其次，对于每一个变量$\mathrm{X_h}$，使用随机森林算法对缺失数据进行插补，分为如下两步：第一步，用因变量$\mathrm{y_{obs}^{(h)}}$和自变量$\mathrm{x_{obs}^{(h)}}$训练出一个 y~x 的随机森林模型；第二步，将$\mathrm{x_{mis}^{(h)}}$作为特征变量输入，用训练出的随机森林模型预测缺失值$\mathrm{y_{mis}^{(h)}}$。对所有$\mathrm{X_h}$预测插补完成后，所得到的矩阵记为$\mathrm{X_{new}^{imp}}$。详见 2.2 中关于插补步骤的修改。

意见 8：表 2 为什么不同时呈现插补为 0 的正确率；表 3 为什么不呈现 FRI 对应的结果？

回应：动态阈值设置是大于 0.5 以上，动态阈值小于 0.5 的时候，RFI 和 RFTI 插补值都是 0，所以 RFI 和 RFTI 插补为 0 的数据个数相同，因此，只统计插补为 1 的正确率，以考察动态阈值的效果。

意见 9：模拟研究二中说“每个数据集分别采用 EM、RFI 和 RFTI 三种缺失数据处理方法进行分析”，为什么不在研究一种就用到 EM 方法？

回应：这是由 RFTI 提出的动机和研究一的目的决定的。研究一的目的为了检验相比于 RFI 方法，采用基于个人拟合指数确定动态阈值的 RFTI 方法的必要性，同时也想检验这种非完全缺失数据插补方法的插补率。

第二轮

审稿人 1 意见：作者已经较好地回答了审稿人提出的问题，但是文中仍有一些文字表达及公式符号使用的问题。这些问题对于读者理解这篇论文可能会造成较大困扰，建议作者仔细修改。

回应：感谢审稿专家的仔细审阅，结合另一位审稿专家的意见，我们认真通读了全文，尤其是仔细检查了公式和符号表示，对一些表述不清楚的内容作了进一步修改。具体修改见文中紫色字体标注的内容。

审稿人 2 意见：

改稿很好地回答了评审专家的问题，文稿质量在可读性、理解性、逻辑性和完整性方面都得到极大的提升。还有一些小的问题，请作者思考、回答和修改。

意见 1：1.2.2 部分第二段第二行最后的下标是 m 还是 n，根据后面的表述应该代表的第 h 题的做答数据，所以 X 代表的 m 个题目的作答列向量的集合。

回应：第二段第二行最后的下标是 m，原稿中存在笔误，在修改稿中已经作了修改，详见 2.2 小节中的第 2 段（紫色字体的内容）。

意见 2：2.2.2 部分第七段“简单来说，....”中应该是还是？

回应：重新修改了这一段的内容，具体见 2.2 中的第七段。

意见 3：3.2.2 部分第十一段“比较与的差异是否小于,...”,这里的表达是否准确？是不是应该比较和某个阈值的大小？因为我的理解是的值代表了与的差异。

回应：感谢审稿专家的仔细审阅，上次修改稿中这一段的确表述不准确，遗漏了迭代停止标准的信息。在修改稿中，我们补充修改了这一段的内容，具体修改见第十一段。

意见 4：4.2.3.2 第三段中阈值的确定方法，感觉是在规定的范围内找最优，带有随机的感觉，而且会比较耗时。有没有考虑过推演阈值变化与 mean_rci 的变化规律，从中给出可行的范围。

回应：感谢您的建议，理论上推演阈值变化与 mean_rci 的变化规律的确是一个很好的建议，实际中也可以为阈值的确定提供更加高效的方法。但从本研究模拟的结果可以推测，两者关系有可能受到缺失比例和缺失机制等诸多复杂因素（可能还包含本研究模拟设计中没有涉及到的其他因素）的交互影响，理论上推导二者的关系似乎存在困难。考虑到本研究重点是提出一种新的缺失数据处理方法的思路，本文没有就这一问题进行探讨。不过我们也认为这是一个值得进一步探讨的问题，在修改稿中我们对讨论部分的内容进行了拓展，拟在后续就中通过更多条件的模拟探讨二者的关系和变化规律，为理论上推导二者的关系提供思路。详见讨论部分 5.2 的修改。

意见 5：5.3 部分上面一段，注意标点符号的使用。

回应：修改稿中已经修改。

第三轮

审稿人 1 意见：建议接受

编委意见：该文经过修改，达到了发表的水平，同意两位审稿专家意见，建议发表。

主编意见：本论文针对认知诊断测评中缺失数据处理这一主题，在对已有方法进行改进的基础上，提出了采用个人拟合指标确定插补阈值的新方法，并对该方法的优势及特点进行了证明。本论文的研究问题具有一定新颖性，研究结论较为可信且具有较强理论和实践意义。