

《心理学报》审稿意见与作者回应

题目：考虑题目选项信息的非参数认知诊断计算机自适应测验

作者：孙小坚, 郭磊

第一轮

审稿人 1 意见：

稿件《考虑题目选项信息的非参数认知诊断计算机自适应测验》在认知诊断计算机自适应测验（CD-CAT）情境下提出了两种考虑题目选项信息的非参数选题策略和两种变长终止规则，在 CD-CAT 充分考虑题目选项信息方面具有理论价值及实践价值。不过研究还存在一些问题值得作者进一步思考和改进：

意见 1：作者在提出非参数选题策略时，对参数选题策略的不足花费了较多的笔墨，而忽视了参数选题策略的优势。审稿人觉得在阐述某方法或策略时，应辩证地进行阐述，而不应只阐述对自身有利的方面。

回应：感谢专家给出的建设性意见，我们已在正文中分别对参数选题策略的优势和不足进行了相应的论述，以尽可能实现对参数选题策略的辩证阐述。

意见 2：作者对于其所提出的两种变长终止规则的理由较为薄弱，张淑君（2019）已经在非参数 CD-CAT 下提出了两种终止规则，如此，作者提出新的终止规则的理由又是什么？

回应：感谢专家给出的中肯意见。我们提出新终止规则的出发点主要有两点：（1）虽然张淑君（2019）提出了 D1 和 D3 两种非参数的变长终止规则，但相对于参数情境中的变长终止规则，非参数的变长终止规则依然比较少，本研究提出的额外的变长终止规则可以进一步丰富非参数情境下 CD-CAT 的终止规则研究。（2）张淑君（2019）提出的 D3 终止规则思路在于每次估计被试的属性掌握模式（AMP）时，要求具有最小非参数距离（如汉明距离）所对应的 AMP 唯一，但三次估计的 AMPs 之间可能各不相同。举例而言，以 3 个属性为例，被试 i 连续三次的 AMPs 均唯一，此时三次的 AMP 分别为{010}、{011}和{110}，这种情况下，基于 D3 终止规则，被试 i 最终的 AMP 为{110}。由于三次的 AMPs 各不相同，如此 D3 可能产生估计误差。

我们已在正文（第 2 页）对提出新终止规则的理由进行了阐述。

意见 3：引言第 3 段和最后一段，作者陈述到“小样本情况下...”，此处的小样本是指正式测试样本，还是预测试阶段的样本量，作者并没有阐述清楚。

回应：感谢专家的细致审阅。此处的小样本是指预测试样本，因为参数选题策略需要进行预测试以获得题目参数估计值，当题目参数值确定后，正式测试样本量的大小对参数估计结果的影响非常小。我们已在正文中给予说明。

意见 4: 部分语句存在重复, 如 2.2 节第一段和第二段中重复出现了“具有最小距离的 AMP 将作为个体最终的 AMP”。

回应: 感谢专家的细心审稿, 我们已将 2.2 节第二段中的重复语句给予删除。

意见 5: 3.2 节中, 建议增加关于什么是缩减 q 向量的描述。

回应: 感谢专家的意见。我们已在正文 3.2 节增加了关于缩减 q 向量的描述, 我们用数学符号将题目 j 的缩减 q 向量描述成: 即 $\underbrace{\{1, \dots, 1\}}_{K_j^*}$, K_j^* 为正确选项所考察的属性个数。

意见 6: 3.4 节, 对 D3 规则的陈述并不是很清晰, 建议重新描述。

回应: 谢谢专家给出的意见。我们已对 D3 规则进行了重新描述: D3 规则下, 每次估计被试 AMP 时, 要求具有最小 HDDmc 只对应唯一的 AMP, 且这种一一对应关系需连续出现 3 次才能结束测验。

意见 7: 4.2 节, 对 Q 矩阵结构的阐述上不清晰, 是题目考察各属性的概率为 20%, 还是题目选项考察各属性的概率为 20%?

回应: 感谢专家的细致审阅。此处应该表述成题目的正确选项考察各属性的概率为 20%, 而错误选项的 q 向量则为正确选项的子集, 且选项之间具有包含关系 (de la Torre, 2009)。如正确选项为 {110}, 则错误选项为 {100} 和 {000}, 或者 {010} 和 {000}。我们已在正文中给予更正。

de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163–183.

意见 8: 6.1 节第二段, 部分表述存在错误, 如“...在简单 Q 矩阵和混合题目质量下的分类准确性要高于所有 JSD 方法”。图 1 可知, 在简单 Q 矩阵、混合题目质量和 2K 题目长度下, JSD 要优于非参数策略。

回应: 感谢专家的细心审阅, 诚如专家所言, 我们的表述确实存在错误。我们已对全文进行了仔细检查, 并将有错误的表述给予更正, 如将专家提到的内容理性为“...在简单 Q 矩阵和混合题目质量下的分类准确性整体要优于 JSD 方法, ...”。但由于是人工检查, 难以做到百分百的准确, 若专家还发现相关错误, 烦请专家批评指正。

意见 9: 部分语句相对拗口, 如 Yigit 等 (2019) 基于选择题 DINA 模型提出了同时考虑正确选项和干扰选项所提供诊断信息的 JSD 策略。

回应: 谢谢专家的意见。我们重新阅读了全文, 并对文中觉得拗口的语句进行了更正。

意见 10: 部分参考文献的格式不符合 APA 第七版要求; 另外, 正文中, 当只有两个作者时, 需要把所有作者呈现 (如 3.2 节中郭磊等 (2021) 应为郭磊和周文杰 (2021)); 再如, 4.2 节第一段的参考文献 (e.g., 郭磊, 周文杰, 2021; Chiu & Chang, 2021; Chang et al., 2019) 中, Chang et al., 2019 应该在 Chiu & Chang, 2021 前面。

回应: 感谢专家的细致审稿。我们已根据 APA 第七版要求, 对文章中的参考文献进行了排

版。

审稿人 2 意见：文章提出了考虑选项信息的非参数 CD-CAT 中的选题策略和终止规则，方法使用正确，文章逻辑清晰，结构完整规范，结果可靠。有一些建议供作者参考：

意见 1：作者提出两种选题策略和两种终止规则，是在前人的基础上修改的，则请作者阐述清楚自己修改的是哪个部分，为什么这么修改，要在突出的位置，让读者一目了然的掌握。

回应：感谢专家给出的建议。

(1)选题策略方面。正文中的 3.2 和 3.3 两个章节专门介绍了基于 MC 汉明距离和基于 Jaccard 距离的 mcCD-CAT，其中 3.2 节阐述了本文所采用的 MC 汉明距离相对于郭磊和周文杰（2021）中的不同之处及其原因：郭磊和周文杰（2021）中对于理想选项的计算值要求个体 i 对题目 j 所考察属性的掌握情况与选项 h 的缩减 q 向量（即 $\underbrace{\{1, \dots, 1\}}_{K_j^*}$ ， K_j^* 为正确选项所考察的属性个数）完全匹配，其计算过于严苛。本文提出的理想选项值则放松了该理想值的计算，以使其更好地适用测验情境。本文的 3.3 节则重点介绍了 Jaccard 距离如何适用于 mcCD-CAT。本文中 Jaccard 距离的计算较何明霜（2021）文章中 Jaccard 距离的主要不同在于理想反应模式的计算，其中本文的理想反应模式的计算见公式(3)。我们已在正文第 7 页阐述了此二者的不同。

(2) 终止规则方面。MR 变长终止规则借鉴的是限制性 MHRM (cMHRM) 算法的算法终止标准的思想，cMHRM 算法需计算前后两次迭代的所有参数估计值间的差值 δ ，并取最大差值 $\max(\delta)$ ，将每次迭代的 $\max(\delta)$ 组成向量 $\Delta = \{\max(d^1), L, \max(d^L)\}$ ，当 Δ 中连续四个 $\max(\delta)$ 均小于预设标准时，算法结束。MR 规则下，当连续四次所估计的 AMPs 均相同时，测验终止，并将该 AMP 作为个体最终的 AMP。而 DR 规则则是借鉴单因子分析中确定单个因子的思想，即当第一个因子特征根远大于第二个特征根时，确定量表测量一个维度。DR 规则就是计算 d_{1st} 和 d_{2nd} 之间的比值，其中 d_{1st} 和 d_{2nd} 表示在计算属性掌握模式 (AMP) 所对应的非参数距离时，距离最小的两个 AMPs 所对应的距离。

意见 2：在研究设计部分，内容太多太长，又没有标题，作者能不能分点阐述研究设计的各个部分，否则读者认知加工负担较重。

回应：感谢专家的建设性意见，我们已在研究设计下面增加三级标题以增加研究设计的可读性。由正文可知，研究设计部分主要阐述了自变量和控制变量的内容，故我们分别将自变量和控制变量作为两个三级标题。

意见 3：写研究结果时，能不能不要用所有人都可以用的标题，如判准率，题库使用等，而应该用突出本研究结果的标题。

回应：感谢专家给予的建设性意见。根据专家的意见，我们在模拟研究一的结果中将 4.4.1

（分类准确性结果）和 4.4.2（题库使用情况）的标题分别更改为“HDDmc 和 JDDmc 的分类准确性整体优于 JSD”和“HDDmc 和 JDDmc 的题库使用情况较 JSD 更加均衡”。

而对于模拟研究二的结果，两个表格阐述的是相同的内容，故无法增加子标题，因而在 5.4（研究结果）的第一段用一句概括性的话来对 5.4 节的内容进行说明：表 2 和表 3 分别呈现了多元正态阈值分布下，HDDmc 和 JDDmc 在四个和六个属性的表现情况，整体而言，HDDmc 和 JDDmc 在 MR 和 DR 两种终止规则下的分类准确性较 D1 和 D3 高，但测验长度更长；同时二者在曝光不足率上的表现优于 D1 和 D3。下面分别对两个表格进行阐述。

意见 4：讨论部分能不能写出各部分的讨论主旨作为标题，否则要读者在文中寻找你到底讨论的是什么。

回应：感谢专家给出的建设性建议。在 6.1 研究讨论章节，我们增加了 4 个三级标题以增加讨论部分的可读性。三级标题分别为：6.1.1 控制简单 Q 矩阵和混合题目质量条件下，属性个数对非参数选题策略有消极影响；6.1.2 MR 和 DR 规则在平衡准确性和题库使用间的表现较 D1 和 D3 规则稍差；6.1.3 Q 矩阵复杂程度对分类准确性有正向影响；6.1.4 不足与展望。

意见 5：请作者再次检查文字表述和参考文献的规范性，如图中用 HDD,JSD 显然不对等。

回应：感谢专家的细致审阅。我们重新对文章进行了仔细审读，对于文章中可能出现表述上的问题以及参考文献进行了修改。

此外，由于 EXCEL 的图例中无法使用字母作为下标，故本研究将使用小写字母 mc 表示选择题的缩写，我们将全文中的 HDD_{MC} 和 JDD_{MC} 全部写成 HDDmc 和 JDDmc，该写法没有改变 HDD 和 JDD 所表达的最初含义，依旧表示“考虑题目选项信息的非参数诊断方法”。

最后，再次感谢两位审稿专家对本论文提出的建设性意见和建议，两位专家提出的意见对于丰富和完善本论文具有非常重要的作用。

第二轮

审稿人 1 意见：经过一轮修改后，稿件《考虑题目选项信息的非参数认知诊断计算机自适应测验》的质量有了较大提升，作者较全面地回答了审稿人提出的问题。还有以下几个小问题供作者参考：

意见 1：个别语句还可以更加精简和准确，如正文第 2 页“要求具有最小非参数距离（如汉明距离）所对应的 AMP 唯一”比较拗口。

回应：感谢专家的建设性意见，我们对文章进行了多次审读，并将其中可能存在拗口的语句

进行了修改，如我们将“要求具有最小非参数距离（如汉明距离）所对应的 AMP 唯一”更改为“最小非参数距离（如汉明距离）只能对应一个 AMP”；又如，我们将“郭磊等（2015）提出了六种变长终止规则，其进一步将此六种规则分成基于绝对标准的终止规则、基于相对标准的终止规则以及结合绝对和相对标准的混合终止规则”精简为“郭磊等（2015）提出了六种变长终止规则，其进一步将此六种规则分成基于绝对标准、基于相对标准以及结合绝对和相对标准的混合终止规则”。再如，正文第 5 页，我们将“候选题目集中具有最大 JSD 值的题目将给个体进行施测”修改为“候选题目集中具有最大 JSD 值的题目将提供给个体作答”。

意见 2：建议统一相关用语，如文章中既使用“个体”，又使用“被试”，此二者可以进行统一。

回应：感谢专家的建议，我们已将文章的“被试”全部替换成了“个体”，以实现相关用语的统一性。

意见 3：部分英文缩写首次出现时，建议写全称，如 KL 信息、SHE 等。

回应：感谢专家的建议，我们已将文章中首次出现的相关英文缩写补全，具体涉及到的英文缩写包括 GDINA、KL 信息、SHE、GDI、AMP、SICM 等。

意见 4：论文还存在个别错字和别字，建议作者再次认真阅读论文，避免出现此类错误。

回应：感谢专家的细致审阅，我们对文章进行了多次阅读，并将其中的错误和别字给予纠正，也请其他读者帮忙对文章进行了核心，以尽量保证文章不出现此类错误。

审稿人 2 意见：

建议作者讨论部分再充实丰富一点儿。

回应：感谢专家给出的建议，我们在讨论部分额外增加了关于 HDDmc 和 JDDmc 不依赖于预测试样本量的讨论，增加该部分内容的原因在于不依赖于预测试样本量是 HDDmc 和 JDDmc 相对于 JSD 策略的特色，也是本研究的核心内容之一。

再次感谢两位审稿专家对本论文提出的建设性意见和建议。