

《心理学报》审稿意见与作者回应

题目：CD-MST 初始阶段模块组建方法比较

作者：高椿雷，罗照盛，喻晓锋，彭亚风，郑蝉金

第一轮

感谢两位评审专家对论文提出的宝贵审稿意见。经过论文作者共同对审稿意见的仔细阅读，我们针对评审专家们的意见，查阅文献认真修改，逐条做出了解释以及修改说明（正文中红色字体内容为本次修改或增补内容）。具体如下，如有不妥之处，敬请批评指正。

审稿人 1 意见：

CD-MST 作为本研究的背景，具有一定的前沿性。为保证自适应测验的精度和效率，初始阶段的选题是相对较为重要的，该问题值得探究。本文中，作者提出了 7 种方法并进行了结果比较。全文整体质量较高，仍有如下建议或问题：

意见 1：文章篇幅尚可，建议把 7 种方法描述的更为详尽。包括但不限于(1)3.1.6 和 3.1.7 中将几种方法结合的原因是什么？7 种方法提出的原因，以及更为清晰的描述；

回应：感谢审稿人的意见，已经进行了修改。

意见 2：Chang & Ying (1999)一文建议在初始阶段选择质量相对较差的题目，而把好的题目用在刀刃上。但作者在文中提出的几种方法均把质量好的题目放在了初始阶段，请作者解释下原因；

回应：Chang & Ying(1999)年提出了 a 分层 b 分块的方式用于基于 IRT 的 CAT 的曝光控制方法，这种方法简单有效，得到业内的认同。本研究是基于 CD 的 MST 的初始阶段模块组建方法，将质量较好的题目放在初始阶段原因有二：

一方面，是基于 IRT 和 CDA 的 CAT 的区别。认知诊断中的知识状态是离散的偏序集合，CDM 是离散偏序的分类模型。不管是基于 IRT 的 CAT 还是基于 CD 的 CAT 的选题策略都是要寻找到一批合适的项目实现最快的从估计值到真值的收敛。离散偏序分类模型的最优收敛速度是指数收敛，而 IRT 中 $\hat{\theta}$ 的最优收敛是 \sqrt{n} （n 是施测项目数），从 $\hat{\alpha}$ 到 α 的速度要快于从 $\hat{\theta}$ 到 θ （Wang, Zheng & Chang, 2014）。因此在基于 IRT 的 CAT 初期使用质量较差的项目对其收敛的影响较小；在基于 CD 的 CAT 初期使用质量较好的项目可以加速其收敛。

另一方面，本研究同 Chang & Ying(1999)的目的不同，他们的研究在于控制 CAT 的项目曝光，本研究目的在于提高初始阶段的测验判准率。

在以后的研究中可以考虑初始阶段的曝光控制等思想，如在 CTTIDR*等方法的基础上采用 RP 等方法控制曝光。

意见 3：全文格式排版质量较差；

回应：感谢审稿人的意见，已经按照心理学报的论文模板进行了修改。

意见 4：英文参考文献中添加对 7 种方法的描述。

回应：感谢审稿人的意见，已经添加，并使用红色标注。

意见 5：4.2 中，3000 人不能被 32 整除，如何重复生成？

回应：感谢审稿人的提问，在 4.2 中对该问题进行了阐述。

意见 6：英文摘要中写 6000 被试，与正文不同。

回应：感谢审稿人的细心审查，已经进行了修改。

审稿人 2 意见：

意见 1：文章概念性错误。全文主要围绕在线多阶段自适应测验 on the fly multistage adaptive testing (OMST) 即依照第一阶段的项目的作答反应为被试自适应组建第二阶段的模块。因此文中应增加对 OMST 的介绍。

回应：非常感谢审稿人的意见。本研究没有采用在线组卷的方式，而是使用 PWKL 信息量进行预先组卷。预先组卷的介绍在 2.2 的第四段和 3.2 进行了介绍，以及讨论部分的第四段对这种组卷方式进行了讨论。

OMST 是传统 MST 的变式，根据被试临时能力采用在线的方式组成下一阶段的模块，这种方式估计精度更高，同时它也缺失了传统 MST 的部分优势，如专家预先检查试卷、容易实现内容约束、组织者的能动性。

在 IRT 中，项目的难度和被试的能力在统一量尺上，因此在 MST 的组卷中，采用难度进行模块组卷。在 CDA 中知识状态是离散的偏序集合，它的项目难度难以界定，采用同 IRT 中相同方式组卷是不可能的。本研究提出采用信息量进行预先组卷，如 PWKL, SHE 等信息量指标，这种组卷方式判准率相对较高。它同样存在着缺点，当属性个数过多时，穷尽所有可能的知识状态，需要组建的模块数量大大增加，因此我们也在寻找更简单方便的组卷方式。

因为没有采用 OMST 的思想组卷，文中为了不引起歧义，尽量减少了 OMST 的介绍。

意见 2：文章全篇有一些表述不准确，内容存在重复，段落间逻辑性有待修改。

回应：感谢审稿人的意见，已经进行了修改。

意见 3：【摘要】a) 首次的字母应有解释。如 CD-MST；b) 部分对文章研究设计的介绍过于简单，包括具体模拟研究过程的介绍；c) 介绍的结论不够充分，不丰富。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 4：【关键词】不应使用缩写。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 5：【1 引言】a) 第一段对于“认知诊断评价”的两个过程描述不恰当，应当包括①工具编制及其质量的评价；②对被试的诊断；③进行诊断评价。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 6：【1 引言】b) 中第一段逻辑顺序需要进一步调整。（如首先介绍 CDA 的概念、意义和优势）；

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 7：【1 引言】中第二段将纸笔测验、CAT、MST 进行比较介绍。其中第二句“具有认知诊断功能 CAT...Mao&Xin,2013)”可以考虑删除。与前一句的内容逻辑性不强。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 8：【2 相关概念和问题的提出】2.1 第一段建议将第二句提前，介绍 MST 的基本概念。之后逐一介绍其成分。2.1 第一段最后一句“几个模块组成阶段”应明确具体模块的特征，是怎样的模块。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 9：【2 相关概念和问题的提出】2.1 第二段第三句表达不准确，中应改为“在第一阶段...，选择第二阶段难度水平同被试能力相匹配的某个模块，并获得被试临时能力估计；同理...”；

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 10：【2 相关概念和问题的提出】2.1 第三段应表述为“CAT 是 MST 的一个特例”。因为特例是有一般推向特殊，CAT 只是针对 MST 题组中具体的一个题目。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 11：【2 相关概念和问题的提出】a)2.1 第四段的第一句“研究了不同结构的...”应明确指出具体是指什么结构？（如模块结构、）；b)2.1 第四段内容与【引言】重复。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 12：【2 相关概念和问题的提出】2.1 第五段应提出 MST 中主要的研究问题，并提出本文的研究重点。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 13：【2 相关概念和问题的提出】2.2 CD-MST 中第二段应简短介绍 CD-MST 的概念，操作特点等，重点介绍预先组卷。

回应：感谢审稿人的意见，已经进行了修改，预先组卷的介绍放在了 2.2 的最后一段介绍。

意见 14：【2 相关概念和问题的提出】2.2 CD-MST 中第三段逻辑不清晰。“在 MST 中，预先组卷需要...(Zheng, Wang &Chang,2014)”这是 MST 的组卷策略中自下而上（bottom-up）策略。另外还有在 MST 中还有自上而下（top-down）策略，应进一步介绍。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 15：【2 相关概念和问题的提出】2.3 中“问题的提出”第一段落中第三句可修改为“CD-MST 不仅需要项目参数，还关注每个项目考察的不同的属性。”

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 16：【2 相关概念和问题的提出】2.3 中“问题的提出”第二段最后一句可修改为“这些方法可以作为 CD-MST 初始项目选择方法”。

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 17：【3 研究方法和采用的认知诊断模型】3.1 中第一段第一句初始项目选取方法是 CD-CAT 的一个重要组成内容。在现有研究中，CD-CAT 初始阶段的项目选取...

回应：感谢审稿人的意见，已经根据审稿人意见进行了修改。

意见 18:【3 研究方法和采用的认知诊断模型】“3.1.2 选题策略法”中初始阶段的知识状态是随机的还是哪种。“3.1.3R 矩阵”中在“在 CD-MST 初始阶段模块组建时...考虑项目和属性的关系属于 R 矩阵所描述的关系”。

回应: 感谢审稿人的意见, 已经根据审稿人意见进行了修改。

意见 19:【3 研究方法和采用的认知诊断模型】3.1.4 中第一段 a)可改为“项目区分度法”。因为本文描述方法本质上是根据 CTT 中的区分度定义的思想推广到 CDM 的区分度中。b)Pah 与 Pal 定义描述不当, 应去掉“某一项目”。

回应: 感谢审稿人的意见。3.1.4 中, 如审稿人所言, 本质上是根据 CTT 中的区分度定义的思想推广到了 CDM 的区分度中, 是可以称为项目区分度法的。但是 3.1.5 也是一种区分度方法, 是基于全域信息量的区分度方法, 为了将两种方法进行区分, 将 3.1.4 的方法称为基于 CTT 的项目区分度法, 将 3.1.5 称为全域项目区分度指标法。

第二个问题已经根据审稿人的意见进行了修改。

意见 20:【3 研究方法和采用的认知诊断模型】3.1.5 中第一段第二句去掉“参数的个数”。因为上一方法中, p 值更多的是由所采用的模型来决定的。

回应: 感谢审稿人的意见, 已经根据审稿人意见进行了修改。

意见 21: 本文结论过于简单, 可以进一步扩展, 例如, 题库使用情况的介绍。

回应: 感谢审稿人的意见, 已经根据审稿人意见进行了修改。

第二轮

审稿人 1 意见:

意见 1: 修改稿较好地回答了审稿人的问题, 整体质量比初稿有所提升, 可发表;

意见 2: 存在部分病句及上下文书写逻辑问题;

回应: 非常感谢审稿人的仔细审阅, 对全文进行仔细审阅, 根据审稿人的意见进行了认真修改。

意见 3: 因为摘要中对 7 个方法的描述前添加序号;

回应: 非常感谢审稿人的仔细审阅, 已经根据审稿人的意见进行了修改。

意见 4: 在最后讨论部分增加部分内容: 把优质题目提前对题目曝光率可能带来的影响等;

回应: 非常感谢审稿人的仔细审阅, 已经在讨论部分增加了优质题目提前对题目曝光率可能带来的影响。

意见 5: 中文摘要中不需要呈现所有的英文全称;

回应: 非常感谢审稿人的仔细审阅, 已经根据审稿人的意见进行了修改。

意见 6: 作者数量大于 5 时请作者确定是否可直接采用“等”;

回应: 非常感谢审稿人的仔细审阅, 已经根据审稿人的意见进行了修改。

意见 7: “在小的纸笔测验基础上自适应化”中的“小的”是什么意思？建议更换为“简短的”

回应: 非常感谢审稿人的仔细审阅，已经根据审稿人的意见进行了修改。

意见 8: “2.2 CD-MST”前一段，建议修改为“现阶段，MST 主要建构在 IRT 的基础上……”

回应: 非常感谢审稿人的仔细审阅，已经根据审稿人的意见进行了修改。

意见 9:“2.3 问题的提出”前一段，“von Davier 和 Chang(2014)……” 可以考虑更换书写逻辑：“本研究尝试将 CD-CAT 的选题策略应用于 CD-MST 中，这与 von Davier 等(2014)的观点一致。”等类似的写法，就是强调方法是你“提出”的，而不是别人说怎么合适就怎样，因为你没有进行论证；

回应: 非常感谢审稿人的仔细审阅，已经根据审稿人的意见进行了修改。

审稿人 2 意见:

修改稿较初稿在内容上表述更清楚，逻辑性，合理性和可读性更强，但仍有几点意见仅供参考：

意见 1: R 方法的优缺点阐述不够，是不足测验考查 k 个属性的。

非常感谢审稿人的仔细审阅，已经根据审稿人的意见进行了修改。

意见 2: CTTID 方法中所有的人的初始阶段的模块项目都是一样的么？还有多个平行模块。

非常感谢审稿人的仔细审阅，在文章的讨论部分对这个问题进行了讨论。

意见 3: 文中仍有格式和错别等小错误。需进一步修正。

非常感谢审稿人的仔细审阅，已经根据审稿人的意见进行了修改。