

## $p$ -最优化题库设计法在国外计算机自适应考试中的应用及其启示

杨丽红

山东建筑大学外国语学院 山东济南

**【摘要】** $p$ -最优化题库设计法是指根据考试目标人群的特点确定符合考试要求的最佳题库蓝图，目的在于确保测试的准确度和题目的均衡利用。 $p$ -最优化题库设计法可在题目数量最优化、测量精度最优化和题库设计过程最优化三方面展现其优势。研究发现，此方法可以解决题库题目冗余，试测成本大，测量精度不高等问题，并可简化测量过程。作为低成本、易操作的启发式算法， $p$ -最优化题库设计法的推介旨在为国内计算机自适应考试题库设计和建设提供参考，为命题专家提供试题参数的统计学依据，降低出题成本，提高考试的信度和效度。

**【关键词】**项目反应理论； $p$ -最优化法；题库设计；计算机自适应考试；启发式算法

**【基金项目】**山东省教育科学“十四五”规划课题招生考试一般专项“基于P-最优化设计法的大学英语四、六级计算机口语测试题库研究”（项目编号：2021KYB002）

**【收稿日期】**2024 年 9 月 12 日

**【出刊日期】**2024 年 12 月 26 日

**【DOI】**10.12208/j.ije.20240082

### A systematic review of the application of the $p$ -optimality method in computerized adaptive testing item pool design and its inspirations

Lihong Yang

School of Foreign Languages, Shandong Jianzhu University, Jinan, Shandong

**【Abstract】**The  $p$ -optimality method refers to determining the best blueprint of the test item pool that meets the examination requirements based on the characteristics of the target population of the examination. The purpose is to ensure the accuracy of the test and the balanced use of items. This article discusses the advantages of this method from three perspectives: the optimization of the number of items in the pool, the optimization of measurement accuracy, and the optimization of item pool design process. It is discovered that the  $p$ -optimality method can address issues such as the redundancy of pool items, reduction of high item trialing costs, and improvement of low measurement precision, simplifying the whole measurement process. As a low-cost and easy-to-operate heuristic algorithm, the promotion of the  $p$ -optimality method aims to provide a reference for the design and construction of computerized adaptive testing item pools in China, and provide support for test development experts, reducing the cost of item writing, and improving the reliability and validity of the test.

**【Keywords】**Item response theory;  $p$ -optimality method; item pool design; computerized adaptive testing; heuristic algorithm

#### 引言

随着各类考试考生规模的不断扩大和部分考试需要满足一年多考的要求，同时为提高测试的信度和效度，急需设计高质量且具规模的题库。20 世纪 60 年代中期，英国首先使用“题库”（item pool）一词。van der Linden et al.（2006）<sup>[1]</sup>给题库的定义是最大数量的题目集合，这些题目需要同时满足以下两个条件：a）满

足测试的目的、所有内容规范和要求；b）在考生一系列能力水平上能反应最大信息量，并能反映考生群体的能力估计分布情况。题库设计不同于题库组卷，后者主要研究如何把题库中的题目有效地组合到一起，形成信度和效度均比较理想的试卷供考生使用。而题库设计主要针对考试目标人群的特点设计出符合考试细目要求的统计学参数，目的在于指导出题者进行高效

出题，使题库最利于测试效果，确保测试的准确度和题目的均衡利用（刘庆思，2006；Yang & Reckase, 2020；Liu et al., 2024）<sup>[2-4]</sup>。然而比起题库组卷来说，国内的题库设计研究相对比较匮乏，建设题库也面临巨大的挑战（杨涛等，2012；修美丽等，2024）<sup>[5-6]</sup>。在当前计算机自适应考试正在逐步推广的前提下，本研究拟回顾国外先进的题库设计理念，*p*-最优题库设计法，分析其题库设计优化路径，在测量准确度和效率方面的优势以及可以克服的问题，以期为国内的题库设计和题库建设提供参考，为命题专家提供试题参数的统计学依据，降低出题成本，提高考试的信度和效度。

### 1 *p*-最优题库设计法

Reckase (2003)<sup>[7]</sup> 提出的 *p*-最优题库设计法的理论基础是 Rasch 模型。这个模型属于二元评分题目，其具体模型公式如下：

$$P_{ij}(U_i = 1 | \theta_j; b_i) = \frac{e^{1.7(\theta_j - b_i)}}{1 + e^{1.7(\theta_j - b_i)}}$$

在此模型中， $U_i$ 表示考生  $i$  在题目  $j$  的得分， $\theta_j$ 表示考生  $j$  的潜在特征， $b_i$ 是题目  $i$  的难度。模型中的 1.7 是常数，目的是将对数几率模型的参数放置在概率单位模型的尺度上（de Ayala, 2009）<sup>[8]</sup>。该方程表示任

何考生  $j$  正确回答任何题目  $i$  的概率。

针对计算机自适应考试题库设计，*p*-最优题库设计法（Reckase, 2003；2010）<sup>[7][9]</sup>的基本思路是：首先根据目标人群的特征进行其分布情况模拟，比如正态分布，或正、负偏态分布等，然后从考试的目标人群中随机选择考生，并向他们施测一系列题目属性比较理想的试题。然后把被选择的试题放入“bins”内（相当于难度区间，以 Rasch 模型为例，被称为 *b*-bin），这些难度区间跟考生的能力在同一量规上，因为 Rasch 模型没有区分度和猜测系数，bin 和试题的难度相同，每个难度区间内的试题被视为具有同等效度。区间宽度根据测试目的和需要确定，所依据的理论是项目反应理论各模型。如图 1 所示，在应用 Rasch 模型时，如果可选择试题的信息量在最大可能性的 95% 范围以内（记为  $p, 0 < p < 1$ ），考试信度为 0.95，所选择的试题和考生能力值基本和两者完全匹配范围差值约 0.35 logit，整个 bin 的宽度为 0.7 logit。如设计好的题库中始终具有可供选择试题信息量最大可能值的 95% 或更高标准，该题库则被称为 0.95 *p*-最优题库。如下图所示，如果题库中试题的难度与当前能力估计相差 0.35 logit，那么考生选做此题与选做和其能力估计值完全匹配的题相比，所损失的信息量最大值为 5%。

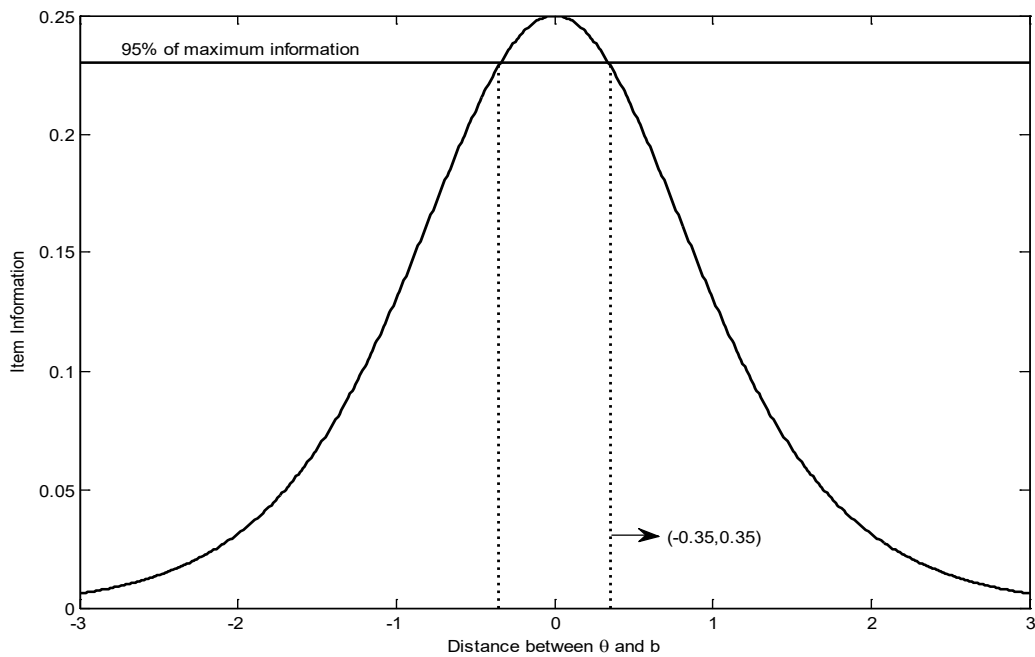


图 1 Rasch 模型下的试题信息函数（Yang & Reckase, 2020, p. 957）<sup>[3]</sup>

### 2 *p*-最优题库设计法的优化路径

*p*-最优题库设计法在美国多个大型计算机自适

应考试题库设计（比如护士执照考试，经企管理研究生考试，军队职业倾向测验）中成功运用多年，且效果良

好。相关实证研究和模拟研究的结果<sup>[3][10-12]</sup> (Yang & Rackase, 2020; He & Reckase, 2014; Mao, 2014; Zhou, 2014) 表明, 此方法主要从以下三方面优化了题库设计, 解决了诸如题库题目冗余, 试测成本大, 测量精度不高和测量过程复杂不易操作等问题。

### 2.1 题库题目数量最优化

最优化题库是指在选择题目时, 题库中包含与考生当前能力估计值完全匹配参数的题目。比如对于 Rasch 模型, 题库大小通常是  $2^{n-1}$ , 其中  $n$  是给考生施测题目的数量。对包含 20 道题目的考试, 最优化题库需包含 1048575 道题目, 这对任何题库设计来说皆不切实际。Reckase (2010)<sup>[9]</sup> 对  $p$ -最优化题库给出的定义是每当计算机自适应考试利用选择算法搜索将要测试的题目时, 题库中恰好有所需题目。Reckase (2003; 2010)<sup>[7][9]</sup> 使用 Rasch 模型展示了在计算机自适应考试题库设计题目数量最优化的具体路径。比如一名考生参加完考试, 根据最优化标准把他/她的最佳题目分配到难度区间 (bin) 内。然后选择另一名考生, 并重复相同程序, 直到考生共同受测题目的数量达到最大值 (Reckase, 2010)<sup>[9]</sup>, 所产生的题库即最优化题库。通过此方法, 能力值相近的考生可在同一难度区间内共享相似的题目, 因此减少了题库所需题目的数量。

题目数量最优化的效果明显, 在 Reckase (2010)<sup>[9]</sup> 的一项研究中, 利用 Rasch 模型将两个具有可变难度区间的  $p$ -最优化题库与实际应用题库中的题目特征进行了比较, 结果表明, 根据  $p$ -最优化题库设计法得到的题库题目数量是 1603, 远远小于实际题库 2000 个题目的数量, 相比之下, 实际题库居于难度中间值的题目超出了需求, 且比  $p$ -最优化题库的难度系数分布范围更小, 如在能力的两个极端值附近题目数量较少, 测量靠近极端值附近考生能力的准确度大大降低。题目数量最优化的效果还体现在可变长度计算机自适应考试中, He 和 Reckase (2014)<sup>[10]</sup> 依据 Rasch 模型把  $p$ -最优化题库设计法应用于此类资格考试中。研究结果表明, 题目分布呈负偏态, 峰值集中在接近分数线的题目难度值附近, 难度值离分数线越远, 题目数量就越少。题库中的题目涵盖了能力量规上所有题目的难度等级, 并提供了足够数量的题目来支持曝光控制程序。此外, 难度区间越窄, 题目数量就越多, 意即题库中题目的特性受难度区间宽度的影响比较大。这一结果也在多维 Rasch 模型研究中得到了验证 (Mao, 2014)<sup>[11]</sup>, 研究者利用多维补偿模型分析了  $p$ -最优化题库设计法在多维计算机自适应考试中 (multidimensional computerized

adaptive testing, MCAT) 的应用, 研究结果表明, 在 MCAT 中, 就能力的测量准确度而言, 模拟题库与基线题库表现相似, 但前者包含的题目更少 (比如要少 100 多个题目), 并具有更好的题库使用率。以上研究表明, 通过  $p$ -最优化题库设计法设计的题库, 当试题的参数和考生能力估计值不能精确匹配时, 所选题目也会对考生当前能力估计值提供最大信息量, 从而减少出题或试测成本。

### 2.2 题库测量精度最优化

$p$ -最优化题库设计法的一个基本原则是根据考生的能力分布状况进行设计, 题库中题目的统计学参数是根据考生的能力量身定制, 所以根据这些参数得到的考试细目蓝图可以最大限度地覆盖考生的能力范围, 尤其是能力极端值附近的考生, 减少测量误差, 在不增加题库题目数量的情况下, 达到测量精度最优化。比如在研究一个可变长度计算机自适应资格考试题库设计中, Reckase (2010)<sup>[9]</sup> 把  $p$ -最优化题库设计的结果和实际题库进行了对比。研究发现, 对于分数线附近的能力值, 不同题库具有相似的估计精度。然而, 在能力值为极端值的情况下, 实际题库的条件标准误差会增加, 而模拟题库的误差则保持不变。同样, He 和 Reckase (2014)<sup>[10]</sup> 依据 Rasch 模型把  $p$ -最优化题库设计法应用于一个可变长度的计算机自适应资格考试中。此研究设计了七个候选题库, 研究因素包括可变难度区间、曝光控制和内容平衡。相比之下, 有曝光控制的题库比没曝光控制的题库考生分类准确度更高, 其题库中题目的数量增加了约三倍。为评估模拟题库的可行性, He 和 Reckase (2014)<sup>[10]</sup> 把两个选定的模拟题库性能与相同试题的退役题库进行比较。结果表明, 退役题库的平均测试时间略长, 分类精度也略低, 尤其是在极端能力值附近的偏差较高。以上结果也在三参数项目反应理论中得到了验证。Gu 和 Reckase (2007)<sup>[13]</sup> 依托 3PL 模型, 把  $p$ -最优化题库设计法应用于美国武装部队职业能力测试。研究结果表明, 无论是否实施曝光控制, 模拟的最优化题库都比实际题库表现更好。此外, 模拟题库中曝光不足题目的百分比也明显低于实际题库。

以上结果也适用于比较复杂的项目反应理论模型, 比如广义分步计分模型和多维 Rasch 模型。Zhou (2014)<sup>[12]</sup> 利用广义分步计分模型 (Generalized Partial Credit Model, GPCM) 探讨  $p$ -最优化题库设计法在计算机自适应考试中的应用。其研究结果表明, 用  $p$ -最优化方法设计的题库在测量准确度和题库使用率方面都较理想。具体来说, 模拟题库能提供足够多的题目, 使整个

考生能力量规都具有充足的信息量,并且在 a-分层曝光控制法约束条件下,区分度参数的分布都保持均衡。利用多维 Rasch 模型, Mao (2014)<sup>[11]</sup>对 p-最优化方法在多维计算机自适应考试中的应用进行了尝试。研究表明, MCAT 的模拟最优化题库可以支撑大量考生的考试,且题库题目也涵盖了整个量规上的考生能力范围,对考生能力估计展现了较高的精度。

### 2.3 题库设计过程最优化

p-最优化题库设计法的设计理念是基于难度区间(bin 范围)的提出,因此其设置对于题库设计过程最优化显得尤为重要。难度区间宽度是根据所依托的项目反应理论类型(如 Rasch 模型、1PL、3PL 等)及所要达到的理想考试信度来决定。如 2.1 所述,如果理想考试信度为 0.95,那么相应的难度区间宽度则为 0.7 logit (Yang & Reckase, 2020)<sup>[3]</sup>。另外难度区间宽度还可以根据考试目的进行调整,如在分数线附近缩小范围以提高测量准确度。这种题库设计法比较灵活易掌握,且不需要任何复杂的软件工具辅助就能达到题库最优化设计的目的。难度区间设置好后,可以结合计算机自适应考试的选题方法和终止条件进行模拟测试,把所选择的试题放入难度区间内,重复执行一定量的模拟次数以减少测量误差,这样就能够得到最终的题库设计蓝图,供出题者参考使用。以下是如何根据项目反应理论的各种模型设计难度区间及题库设计蓝图的最优化路径。

在单维 Rasch 模型中,难度区间和考生能力在同一量规上,因此可以根据理想信度的大小直接确定难度区间,是属于最容易设计的一种类型(Reckase, 2003, 2010; He & Reckase, 2014)<sup>[7][9][10]</sup>。在 3PL 模型中,根据预测模型策略(prediction model strategy),区分度的对数和难度之间存在一定的线性关系(Gu & Reckase, 2007)<sup>[13]</sup>。于是利用实际题库中的回归模型可以从难度系数预测区分度的对数及区分度。猜测系数为常数项,从 beta 分布中选取,原来的难度区间则变成难度与区分度组合区间(ab-bin)。如图 2 所示(Gu & Reckase, 2007)<sup>[13]</sup>,对于能力估计值在(-1.12:-0.84)的考生来说,在 ab-bin 中的 A(-0.84:-0.56, 1.26:1.55)比位于(-1.12:-0.84, 0.00:0.89)中的 B 题目所提供的信息量大,尽管看起来 B 离考生的能力值估计较近且有可能提供更大的信息量(Gu & Reckase, 2007)<sup>[13]</sup>。

多维 Rasch 模型和广义分步计分模型的难度区间设置比较复杂。在 Mao (2014)<sup>[11]</sup>用多维 Rasch 模型研究的 MCAT 中,多维概念分为题目间多维和题目内

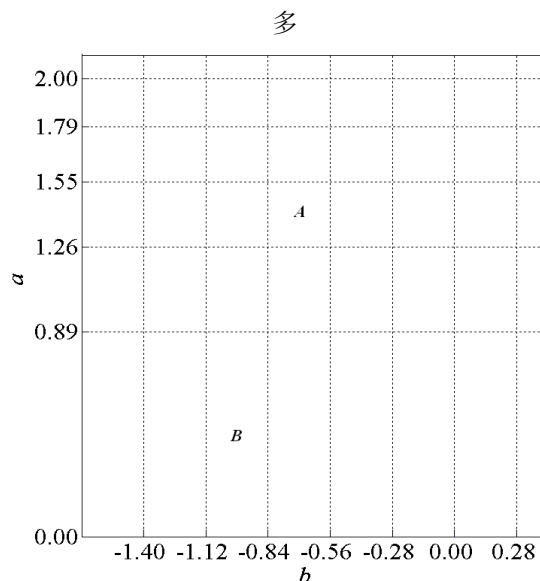


图 2 ab-bin 展示 (Gu & Reckase, 2007, p. 16)<sup>[13]</sup>

维,前者为简单结构,后者为非简单结构(Reckase, 2009)<sup>[14]</sup>。在多维 Rasch 模型中,区分度为项目反应面最陡处的斜率,难度系数是从原点到最陡处的距离,信息量则为一个矩阵。难度区间(MDIFF-bin)宽度取决于试题要求的信度和题目在多个维度的信息量。基于不同的测试蓝图, Mao (2014)<sup>[11]</sup>设计并开发了 24 个模拟 p-最优化题库,设计因素包括维度间不同的相关系数、不同难度区间和曝光控制条件。由于现实中并不存在多维题库,基线对比题库是根据过往文献中使用的题库设计。研究发现,题库的大小和以下几个因素有关:难度区间的大小、测试蓝图细目、维度间的相关系数和曝光控制条件(Mao, 2014)<sup>[11]</sup>。比如在曝光控制条件下,当难度区间缩小时,维度之间的相关度会变高,测试结构也会变成非简单结构,导致题库需要更多的题目来支撑。在 Zhou (2014)<sup>[12]</sup>利用 GPCM 模型来探讨多级计分计算机自适应考试中,考试过程实行了 a-分层曝光控制法(Pastor et al., 20020)<sup>[15]</sup>(最大速率为 0.20)和内容平衡。在此研究中,难度区间被定义为难度、能力值组合区间(aθ-bin)。在多级计分模型的所有参数中,包括题目的区分度、总难度参数和各阈值难度参数,任何单一参数都不能用来代表题目的信息量。但是如果给定题目参数,总会存在唯一的能力值 θ 对应这个题目的最大信息量。因此 b-bin 被扩展为 aθ-bin,其中 θ 表示题目信息最大化所在的位置(Zhou, 2014)<sup>[12]</sup>。比如此研究的目的是 0.98 p-最优化题库,在考生能力量规上以零为中心,除两端外,平均距离为 0.81 logit,确定难度区间。然后和 a-分层曝光控制法相结

合, 确定  $a\theta$ -bin。

在多阶段计算机自适应考试 (multistage computerized adaptive testing, MST) 中, 因为需要更复杂的组卷过程, 难度区间被分成了多个阶段。依托单维 Rasch 模型, Yang 和 Reckase (2020) [3] 把 *p*-最优化设计法应用于 MST 中, 在曝光控制和非曝光控制两种条件下, 选择了四种应用范围最广泛的 MST 面板设计类型: MST 1-2, MST 1-3, MST 1-2-2 和 MST 1-2-3。每种面板设计均包括三个模拟因素: 试题长度、路由测试长度比例和曝光控制, 生成 72 个最优化题库, 并通过总样本和条件样本的估计误差来评估题库的表现。以 MST 1-2-2 为例, 在创建 MST 第一阶段路由测试蓝图过程中, 首先设置一个从一阶段到二阶段的分数线 ( $M_i$ ), 将考生平均分配到二阶段模块。对于有不同模块的 MST 试题来说, 分数线的数量也各不相同。确定分数线后, 第一阶段路由测试的题目数量从含有  $M_i$  的难度区间内 (*b*-bin) 等机会随机选择。二阶段和三阶段过程类似, 考试直到满足两个条件之一将终止: 模块长度用尽或满足特定考生的预设信息。终止标准将用于决定第三阶段的难度区间和题目数量。综上所述, *p*-最优化题库设计法可结合多种类型的项目反应理论进行题库设计。因为整个难度区间内的考试题目是等效的, 这就解决了计算机自适应考试过程中选题唯一性难的问题, 不必拘泥于必须找到完全符合考生当前能力值的题目, 而是在预设的考试信度范围内做到灵活选择合适的题目, 提高组卷效率。

### 3 促进我国计算机自适应考试题库设计的建议

基于以上讨论, *p*-最优化题库设计法在计算机自适应考试领域已取得不错的成绩, 未来应用也具有广阔前景。近十年来, 关于题库设计的研究和应用在国外发展迅速, 但国内相关研究还比较匮乏。目前我国计算机自适应考试, 比如信息系统安全专业认证考试, 正在逐步大规模实施, 国外先进的题库设计理念和实践经验值得我们学习和借鉴。本文针对 *p*-最优化题库设计法在国内 CAT 或未来 MST 考试中的应用及研究方向提出以下建议:

(1) 因为题库建设的成本比较高, 对于国内已经建立起来的题库, *p*-最优化题库设计法可以指导其再次利用。研究者可以首先分析实际题库中题目参数的分布特征, 然后根据目标人群的能力分布状况进行调整。如果实际题库中的题目参数与测试目标人群的分布特征一致, 选定 IRT 模型和理想信度后, 可直接根据本文介绍的 *p*-最优化题库设计法步骤进行; 如果不

一致, 则在能力值不匹配的位置会出现较高的测量误差, 这时在设计题库时如以考生当前的能力分布状态为准, 会避免测量误差的出现。

(2) *p*-最优化设计法对国内题库的管理和运行提供了新的思路。众所周知, 题库不是静态的, 而是不断发展变化的动态实体。有些题目向公众发布后很快就会过时, 或由于重复使用而导致过度曝光。新题目不断涌现且需要被添加至题库当中。van der Linden et al. (2000) [16] 讨论了使用成本函数法进行题库管理的困难。因为需要旧题库来定义成本函数, 并使用新添加的题目来设计新题库。新题目应该与旧题库中的题目属性相同, 成本函数才能拟合。然而, 如 van der Linden et al. (2000) [16] 所述, 旧题库中的题目属性可能与新题目的属性不符, 如果引入新题目, 此方法“可能会面临无法解决的数据缺失问题” (van der Linden et al., 2000 年, p.148) [16]。相比之下, 用 *p*-最优化题库设计法设计题库无须考虑新旧题目间属性是否相符的问题。研究者可根据所依托的 IRT 模型和理想信度直接按照 *p*-最优化题库设计法步骤进行模拟题库设计, 得到题库细目蓝图, 使题库开发和管理更加科学与高效。

(3) *p*-最优化题库设计法虽然在实际或模拟测试中效果良好, 但目前此方法只应用于客观题的题库设计, 对于主观题, 比如口语考试方面还是空白, 未来研究可结合自动评分系统进一步探索此方法的可行性。另外在实际应用中, 题库设计还会受到众多其他因素的影响。比如虽然大多数情况下考生能力呈正态分布, 但不排除有负偏态或正偏态分布的情况; 另外某些试题类型, 比如阅读题或图表题具有题组效应。其他因素比如试题长度是否固定, 尤其针对多维 CAT 和 MST 测试, 是如何影响 bin 的宽度设置及题库测量的精度等问题都有待研究者进一步探索。

### 4 结语

论文从题库题目数量最优化、测量精度最优化和题库设计过程最优化三方面分析了 *p*-最优化题库设计法在国外计算机自适应考试题库设计中的优化路径, 介绍了题库设计的基本思路 and 具体操作步骤, 并就其未来应用和研究方向提出了建议。探讨 *p*-最优化设计法在国外计算机自适应考试题库设计中的应用能帮助教育从业者更深入地了解此方法在不同考试模式和不同测试模型框架下的表现, 给命题专家提供题库题量参考和试题参数的统计学指标, 避免或克服现阶段命题者对试题的预估难度和实测难度之间的差异, 以期降低出题和试测成本, 提高组卷效率和考试的信度及效度。

## 参考文献

- [1] van der Linden, W. J., Ariel, A., & Veldkamp, B. P. Assembling a CAT item pool as a set of linear tests. *Journal of Educational and Behavioral Statistics*, 2006, 31, 81-99.
- [2] 刘庆思. 英语等级考试题库介绍[J]. *中国考试*, 2006 (12): 3-8.
- [3] Yang, L., Reckase, M.D. The optimal item pool design in multistage computerized adaptive tests with the p-optimality method. *Educational and Psychological Measurement*, 2020, 80(5), 955-974.
- [4] Liu, Y., Li, Z., Liu, H., & You, X. The Impact of Non-Effortful Responding on Item and Person Parameters in Item-Pool Scaling Linking. *Applied Measurement in Education*, 2024, 37(2), 89-108.
- [5] 杨涛, 杨婷婷, 辛涛. 题库优化设计的回顾与展望[J]. *心理与行为研究*, 2012 (3): 154-160.
- [6] 修美丽, 柴省三, 鲍杰. 自适性语言测试题库 CTT 与 IRT 参数估计对比研究[J]. *首都师范大学学报(社会科学版)*, 2024 (3): 180-189.
- [7] Reckase, M. D. Item pool design for computerized adaptive tests. Paper presented at the Annual Meeting of the National Council on Measurement in Education (NCME), Chicago, IL. 2003.
- [8] de Ayala, R. J. The theory and practice of item response theory. New York: The Guilford Press. 2009.
- [9] Reckase, M. D. Designing item pools to optimize the functioning of a computerized adaptive test. *Psychological Test and Assessment Modeling*, 2010, 52(2), 127-141.
- [10] He, W., & Reckase, M.D. Item pool design for an operational variable-length computerized adaptive test. *Educational and Psychological Measurement*, 2014, 74(3), 473-494.
- [11] Mao, L. Designing p-optimal item pools for multidimensional computerized adaptive testing [Unpublished doctoral dissertation]. Michigan State University. 2014.
- [12] Zhou, X. Optimal item pool design for computerized adaptive tests with polytomous items using GPCM. *Psychological Test and Assessment Modeling*, 2014, 56 (3), 255-274.
- [13] Gu, L., & Reckase, M. D. Designing optimal item pools for computerized adaptive tests with Symptom-Hetter exposure control. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. 2007 .
- [14] Reckase, M. D. *Multidimensional Item Response Theory*. New York: Springer Dordrecht Heidelberg London. 2009.
- [15] Pastor, D. A., Dodd, B. G., & Chang, H.H. A comparison of item selection techniques and exposure control mechanisms in CATs using the generalized partial credit model. *Applied Psychological Measurement*, 2002, 26(2), 147-163.
- [16] van der Linden, W. J., Veldkamp, B. P., and Reese, L. M. An integer programming approach to item pool design. *Applied Psychological Measurement*, 2000, 24(2), 139-150.

**版权声明:** ©2024 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。  
<http://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**