

测验 Q 矩阵中属性指定、选择和验证方法

宋丽红

(江西师范大学 初等教育学院 江西 南昌 330022)

摘要: 在认知诊断评估中,构建正确测验 Q 矩阵十分关键但比较困难。本文在简要介绍认知诊断评估之后,从认知诊断模型视角,重点阐述和评价了知名模型或框架下(规则空间模型、属性层级模型、融合模型、回归树、确定性输入噪声与门模型、证据中心设计和认知设计系统)测验 Q 矩阵中属性指定、选择和验证的方法,指出了各方法的优势和劣势。最后,归纳了测验 Q 矩阵构建的一般性步骤和方法,并讨论了未来可能研究方向。

关键词: 认知诊断评估;测验 Q 矩阵;属性指定;属性选择;属性验证

中图分类号: G449 **文献标识码:** A **文章编号:** 1000-579(2017)01-0080-08

Attribute Specification, Selection and Proof Methods of in the Test Q - matrix

SONG Lihong

(Elementary Education College, Jiangxi Normal University, Nanchang, Jiangxi 330022, China)

Abstract: In cognitive diagnostic assessment, to build up the right test Q - matrix is crucial but difficult. From the perspective of the cognitive diagnostic model, the paper elaborates and evaluates attribute specification, selection and proof techniques in the test Q - matrix under the famous models or frameworks(Rule Space model, attribute hierarchy model, fusion model, return tree, deterministic input noisy AND - gate model, evidence center design and cognitive design system), points out the advantages and disadvantages of each method, summarizes the general steps and methods of the test Q - matrix construction, and discusses possible future research directions.

Key words: cognitive diagnostic assessment; test Q - matrix; attribute specification; attribute selection; attribute proof

1 引言

认知诊断评估用于测量被试在知识结构和加工技能(以下简称属性)上的掌握情况。知识结构包含陈述性知识和程序性知识,而加工技能包括处理这些信息所需的转换和策略。认知取向的测量学重

收稿日期:2016-04-10

基金项目:国家自然科学基金项目(编号:31500909,31360237,31160203,30860084);全国教育科学规划教育部重点课题(编号:DHA150285);教育部人文社会科学研究青年基金项目(编号:13YJC880060);江西省自然科学基金项目(编号:20161BAB212044);江西省教育科学 2013 年度一般课题(编号:13YB032);江西省教育厅科技计划项目(编号:GJJ13207)

作者简介:宋丽红(1981-),女,江西新干人,博士,江西师范大学讲师。研究方向为教育和心理测量。

点旨在探索人类在特定领域的认知过程和结构,及其对人类个别差异的理解和分析。简而言之,作为心理测量学最新发展方向之一的认知诊断评估的兴起,标志着心理测量学开始更深层次地探讨人类个体差异。认知诊断评估的终极目标,是服务于学习和学习进程的评估,而不是对学习结果的评估。

要实现上述目的,认知诊断评估的实施必然有别于传统的教育或心理测量的实施。纵观已有研究,研究者和实践者在不同的心理计量模型下就诊断评估过程和实施进行了大量探索和讨论,如规则空间模型、属性层级模型、证据中心设计和认知设计系统方法。从这些文献的探讨中不难发现,认知诊断评估的实现需要广泛纳取各行各业专家智慧,需要在测验使用者、测验开发者、认知心理学家、学科领域专家、心理计量学者的相互密切协作下进行。正如统计手册(第26卷,心理测量卷)中第31章所讨论的,实现认知诊断评估至少包括6个步骤:(1)描述认知诊断的目的;(2)构建诊断的潜在技能模型或认知模型;(3)开发和分析评估任务;(4)匹配与整合认知模型与计量模型;(5)选择模型参数估计方法并评价估计结果;(6)报告诊断结果。还有研究者对以上步骤进行扩充,提出了9个步骤^[1]。总的来说,大多数认知诊断评估的实施都包含1-3和6步,无论采用哪种认知诊断模型,只是所采用的方法有些不同。尤其是第2步,构建诊断的潜在技能模型或认知模型,是实现认知诊断评估目的的一个基本且十分关键的步骤,也是认知诊断评估理论区别于传统测量理论的根本所在。

测验或评估中所涉及的潜在技能模型或认知模型常采用一个Q矩阵来刻画。^[2]Q矩阵是一个由1和0两种元素构成的矩阵(还有采用整数形式的Q矩阵)。其各元素1或0用以描述测验项目是否测量了某个认知加工技能、知识或策略(即属性),元素为1表示正确作答项目要求掌握该属性,否则不要求。通过这种表述,Q矩阵实现了对项目属性间关系的形式化描述。换言之,要构建诊断的潜在技能模型或认知模型,就是要构建测验Q矩阵,主要包括对Q矩阵元素进行指定和验证。

本文旨在对国内外近30年来Q矩阵构建的相关研究进行系统地研读和归纳,主要基于以下几点考虑:(1)如何构建测验Q矩阵是绝大多数认知诊断方法要解决的一项重要内容,它连接项目与潜在特质,是结构表征的关键,也是构想效度的主要证据;(2)众多研究显示正确构建测验Q矩阵比较困难和复杂;^[3-4](3)错误构建测验Q矩阵会带来严重的后果,^[5]比如项目参数估计不准确和被试分类不准确等;(4)Q矩阵的属性指定和验证方法成为近年来关注的焦点,因研究和应用特定认知诊断方法的差异,Q矩阵构建方法各有不同。有待就各种测验Q矩阵构建方法进行系统介绍与梳理,但目前关于Q矩阵构建的系统介绍较少。^[6]相比之下,国内外关于认知诊断模型研究的介绍和综述则已经比较系统和深入。

鉴于测验Q矩阵的构建,尤其是属性验证方法常与特定的认知诊断模型有关,本文从认知诊断模型视角切入,按照各认知诊断模型或分类方法下属性指定、选择和验证进行分类叙述,旨在归纳总结和评价测验Q矩阵中属性指定、选择和验证的方法,为认知诊断评估实践者提供参考。涉及的模型或方法主要包括:规则空间模型;属性层级模型;融合模型;回归树;确定性输入噪音与门模型;证据中心设计和认知设计系统。

2 Q矩阵指定、选择和验证方法

2.1 规则空间模型下属性指定、选择和验证方法

规则空间模型是发展较早的认知诊断模型之一。在运用规则空间方法进行认知诊断评估时,首先由学科专家定义目标任务中的属性,并编写或选择测量这些属性集的项目;然后指定项目属性,即指定项目与属性关联Q矩阵;确定知识状态和建立分类空间,对被试进行分类。规则空间模型的应用相对较多,如用于分数减法测验、代数测验、学术能力倾向测验、大学英语入学考试和第三版国际数学和科学学业测试。^[7-8]国内也有不少研究者关注规则空间模型分类方法及其在各学科领域的诊断应用。规则空间模型下属性指定、选择和验证属性的具体步骤如下:

(1) 确定可能影响作答反应的初始属性集。主要依靠文献调查、认知理论、专家经验、测验开发实

践和作答者自省等信息。

(2) 编码可能影响作答反应的项目特征。尽可能从低层次详细编码项目特征,如低频词、复杂的语法等。可能遇到的问题是,有的项目特征不易给出操作定义,甚至两个编码实际上是对同一属性进行的编码,必须删掉其中一个。这样就需要两个评定者确定每个项目必要的信息,并在评定不一致时,由评定者讨论决定。在编码之前,必须有详细的编码手册、解释和示例。编码后,还要估计出属性编码的信度。对于信度计算,对连续属性可计算皮尔逊相关系数 $\rho - 1$ 属性可计算编码一致所占的百分比。

(3) 构建初始 Q 矩阵。标识和分析项目属性,厘清项目特征与属性的关系,从更为抽象、理论和高层次描述属性,如有的项目特征可被解释为低频词的知识,或根据文本中零散的信息进行推断的能力等。由于规则空间模型中 Q 矩阵元素限定为 0-1 形式,连续特征变量需要 0-1 化,属性 0-1 化可采用难度和属性分的散点图来决定划界值。如果找不出明显的分界,可选择属性分均值,也可以采用下面介绍的分树方法。

(4) 使用规则空间模型和回归分析筛选属性。为了确保测量信度,简化属性时应删除或合并测量项目过少的属性。先采用规则空间模型进行分析,对被试进行分类,同时计算被试分类率和被试属性掌握概率(属性分数)。然后使用属性分数对总分的多元回归做交叉验证,参考回归系数的显著性、决定系数和样本的平均属性掌握率来对各属性进行筛选。在检查了每个属性的统计性质时,需要考虑各属性的理论重要性,以筛选和确定属性。

(5) 探索属性之间交互效应并增加属性。在实际数据中一般不会出现三阶交互作用,若某两个属性存在交互效应,会使项目变难/变易。将产生显著交互效应的两个属性编码成新的属性,确定最后的属性集。

(6) 验证属性。再使用规则空间模型对属性进行验证。通过各种描述统计量,如属性分数与总分的相关,或属性分数对总分的回归系数等,删除解释力小的属性。采用规则空间模型反复分析,直至属性能比较充分地解释作答反应为止。标准主要有被试分类率和决定系数。

规则空间模型下属性指定、选择和验证属性方法的优势:(1) 主要应用于已有测验数据分析,可以挖掘测验数据中的潜在诊断价值。一方面可发挥已有测验的诊断作用,如在分析学生成绩的同时进行诊断分析,可为学生提供具有一定价值的诊断分数报告。另一方面可用于筛选具有诊断功能的测验项目,因为要编写具有诊断功能的项目并标定其属性相对比较难,再加上需要预试筛选等,开发成本较高。(2) 采用被试属性分数对总分的多元回归方法,根据回归系数和决定系数,可对属性进行筛选、合并及其编码交互作用的属性。对交互作用的属性进行编码,可提高方差解释率。从统计上来看,该方法比 Q 矩阵对难度的回归方法更合理,因为后者的样本量(项目数)相对较少。(3) 规则空间模型并不强行对所有被试进行分类,通过计算被试分类率,可以探查是否有可能存在被遗漏的属性。有研究指出,当项目关键属性漏指时,会导致属性掌握概率低估。

规则空间模型下属性指定、选择和验证属性方法的劣势:(1) 规则空间方法分类准确率不是很理想,会对属性选择和验证产生一定的影响。已有研究结果显示,规则空间方法的分类准确率不如属性层级模型下 A 方法或广义距离方法,^[9]而采用规则空间模型得到的属性分数对总分的回归系数等来选择和验证属性,应该会给属性选择和验证带来一定的负面影响。(2) 规则空间模型采取对已有测验进行事后分析,所提取的测验 Q 矩阵未必能区分开所有知识状态,势必也会影响分类的准确率及其影响属性选择和验证。

2.2 属性层级模型下属性指定、选择和验证方法

属性层级模型强调认知属性并非独立存在,而是有层次的网络结构,具有层级结构的认知属性可以更好地反应人类的认知加工。采用属性层级模型对被试进行诊断分类,可以归纳为五个步骤:(1) 指定认知属性和属性层级;(2) 采用 Q 矩阵对属性层级进行形式化表述并确定各项目属性;(3) 产生期望反应模式;(4) 用项目反应理论模型估计得到项目反应概率;(5) 对观察反应模式进行分类。

国内关于属性层级模型的理论或方法研究,主要包括分类方法、属性权重、属性层级验证和多级评分属性层级模型研究等。国内关于属性层级模型的应用研究,主要涉及三段论推理测验、分数加减测验和进位计数制转换测验等。国外对属性层级模型的应用研究,包括对英语^[10]和数学^[11]等进行的诊断分析。

虽然属性层级模型在不同类型的考试和学科中有所应用,但属性层级模型下Q矩阵构建的思路基本是借助构建的层级模型指导诊断测验开发的思路,也有部分研究是对已有测验进行分析。下文对属性层级模型下认知模型和项目属性指定及验证步骤进行归纳。

(1) 初步构建认知模型。主要采用以下三种方式得出认知模型:第一种方式也是最主要的方式,是请有多年教学经验的学科专家根据课程标准讨论决定认知模型。第二种方式可通过文献调查得到初步的认知模型。第三种方式适用于对已有测验进行分析,主要是请专家解答项目样本,并标定各项目测量的属性(如概念、操作、过程和策略),然后对得到的各属性进行分类并根据属性的心理或逻辑上的顺序决定属性层级结构。

(2) 采用口语报告法构造并验证认知模型。通过分析口语报告结果得出项目解决的流程图,表示问题解决中包含的认知成分及其结果,由此推断属性之间的关系。并将流程图中基本的认知加工成分编码成一般的问题解决策略。最后将专家给出的属性描述和每个项目的认知加工流程图进行比较,评价被试报告的认知加工与专家认知模型的匹配程度。另外还有研究对口语报告结果进行编码并报告被试正确作答项目使用属性的频次。

(3) 使用属性层级模型进行计量分析。首先使用层级相合性指标评价认知模型与数据的拟合程度。其次采用属性层级模型分类方法(如神经网络方法)对被试进行分类,再采用属性信度对分数结果的信度进行评价。使用各属性(掌握概率)之间的相关大小评价属性层级关系紧密程度,属性层级结构中邻接属性相关应较大,而距离较远的属性相关应较小。另外可以请教师对学生样本进行诊断,并与认知诊断模型分类结果进行一致性评价等。

属性层级模型下属性指定、选择和验证方法的优势:(1) 属性层级模型下可使用认知模型指导测验开发。一旦测验领域中属性层级结构确定后,测验开发者就能根据属性层级结构,针对性地编写相应属性向量的测验项目,测验开发者就能最大程度上检查项目所测的属性,这时标定项目属性相对容易。(2) 认知模型可用于指导测验编制和组卷。在给定属性层级模型或认知模型下,根据认知诊断测验蓝图设计原则可编制或组成分类准确率更高的测验,^[12]有利于属性验证。(3) 属性层级模型还可以使用属性层级指标评价和选择特定领域的认知模型。(4) 属性层级结构对于诊断补救具有非常特殊的意义,属性层级结构反映的是属性掌握的先后关系,有利于循序渐进式的补救。属性层级结构是否真是如此,可以通过补救的效果进行评价。(5) 属性层级模型还可以用于学习过程的评估,可以建立起学习过程中假设表现水平(与属性掌握模式有一定的关系)与期望反应模式之间的联系,因此可借鉴标准参照测验的标准来标定属性。

属性层级模型下属性指定、选择和验证方法的劣势:(1) 属性层级指标可以用于从多个认知模型或测验Q阵中选择拟合较好的模型,但是尚不能修正拟合不佳的认知模型或测验Q阵。(2) 属性层级模型下还没有结合作答数据的专门的Q阵修正方法。毕竟学科专家或口语报告得到的认知模型和测验Q阵代表性不高,较难反映各种类型的被试的知识结构,尚有进一步提升的空间。(3) 要为各个学科领域和年级构建相应的认知模型,并指导编制相应的认知诊断测验,然后施测并利用测验数据对认知模型和测验Q阵进行验证,短时之内仍然很难完成。

2.3 融合模型下属性指定、选择和验证方法

融合模型源于统一模型。统一模型通过结合Q矩阵来说明潜在属性空间,并使用类似项目反应理论模型的形式来描述Q矩阵未充分界定的其他因素的影响。在统一模型中由单参数项目反应理论模型来描述剩余的潜在空间,统一模型尝试把认知诊断模型和项目反应理论模型结合起来,用来解释Q

矩阵覆盖每个项目属性的程度。国外,融合模型主要分析英语和数学测验。^[4,13,14]在国内,有研究者采用融合模型分析几何类比推理测验。综合这些文献,可以将融合模型下属性指定、选择和验证的主要步骤归纳如下:

(1) 确定初始项目属性集。考虑到项目所测量内容与所测的加工技能或策略有一定的关系,有必要在口语报告前先进行预分析,如内容分析,以便于口语报告时有针对性地识别和验证特定项目上学生所使用的加工技能。然后进行口语报告,要求学生在答题的过程中或答题后口头报告其思维过程,用于分析提取学生的加工技能和策略,并结合课程标准或学习目标,得到初始的属性集。

(2) 属性选择。通过口语报告法确定的初始属性比较细、比较多,尤其是在题目量不大的情况下,需要将一些过细或不重要的属性合并到对正确作答有直接影响的属性上去。

(3) 专家验证。请学科专家对属性进行验证。给出合并后的属性,要求学科专家独立地对每个项目的属性考察情况进行选择和判断,并对属性的重要性进行排序,同时还要求专家列出可能遗漏的重要属性。将专家标定结果与被试的口语报告结果进行比对,并讨论确定初始 Q 阵。

(4) 统计验证。融合模型常采用逐步简化算法对 Q 阵元素修正。如果一批项目 π 参数太小,可考虑增加属性; r 参数置信区间下限大于 0.9,可删除属性;置信区间太宽,如大于 0.3,可删除不可识别属性; c 参数置信区间下限大于 1.75,可设置 c 参数为 10(相当于删除 c 参数); c 参数置信区间太宽,如大于 0.5,可删除。

融合模型下属性指定、选择和验证方法的优势:(1) 强调理论、专家和作答数据在属性指定、选择和验证过程中的重要作用。利用的信息更加充分和全面,并可以进行相互验证,有利于正确 Q 矩阵的构建。(2) 融合模型下的项目参数可为属性指定、选择和验证方法提供更为丰富的参考信息。如项目参数中的 r 参数,可以反映属性在项目正确作答的重要程度, r 参数可以作为属性选择的参考指标。项目参数中的 c 参数可以提供属性是否存在遗漏的指示信息。(3) 融合模型提供了逐步简化算法,可结合数据对理论或专家 Q 矩阵进行修正。

融合模型下属性指定、选择和验证方法的劣势:(1) 融合模型比较复杂,要达到对项目参数准确估计,对样本量要求比较大。(2) 逐步简化算法伴随着 MCMC 算法进行,速度较慢,对 Q 矩阵进行修正后,是否需要增加链长以达到收敛或获得平稳分布,尚有待探讨。

2.4 基于回归树的属性指定、选择和验证方法

Gao 强调认知心理学和评价结合,以得出被试在任务解决中所使用的知识结构、技能和策略等有价值的信息。^[15] 要达到这一目的,就必须建立起项目难度与项目认知结构的关系。使用回归树进行诊断具有以下特点:(1) 项目特征或属性可为离散型或连续型;(2) 建立的项目难度与属性之间的模型可为线性或非线性模型;(3) 可基于聚类特征曲线和非参数平滑方法将项目难度模型转化成学生能力模型;(4) 可以提供群组水平和个体水平诊断信息。

回归树的应用主要有:使用回归树分析教师资格考试(Praxis I CBT),该考试主要测量阅读、写作和数学属性。认为回归树有助于识别重要的交互效应(如项目内容之间的交互),而这些交互效应可能并没有预先指定出来;使用回归树分析学术能力倾向测验的阅读测验;^[16] 建立托福测试阅读题的难度模型;应用回归树对听力和阅读项目建立项目难度模型;应用回归树对新版托福测验建模;对密歇根大学英语水平测试(MELAB)阅读部分进行分析,使用回归树解释测验项目难度,为专家判断和口语报告得到的认知模型提供证据支持。

回归树可以用于筛选测验中重要属性。回归树的分析数据包括项目反应理论模型的项目难度向量和假设的属性与项目的关联矩阵。可以请学科专家标定 Q 阵。回归树主要基于项目之间内在的相关,用来识别出包含重要属性组合的项目类。回归树下属性指定、选择和验证方法的最大优势在于可以用于从多个属性中选择出少数重要的属性,需要与层次聚类方法和多维标定法结合使用,用于识别并选择属性。但是,回归树下属性指定、选择和验证方法,是在既定的测验 Q 阵下进行,尚不能对测验 Q 阵进

行修正。

2.5 确定性输入噪音与门模型下属性指定和验证方法

确定性输入噪音与门模型或约束潜在类别模型,是一种简单且解释性较好的认知诊断模型。对于确定性输入噪音与门模型下 Q 矩阵中属性指定或属性修正方法主要有: δ 经验方法^[17]和推广的 ζ 方法、 γ 方法^[18]、贝叶斯估计方法^[3]、数据驱动方法^[19]、非参数的 Q 矩阵估计方法^[20]和在线标定方法^[21-27]等。还有研究将在线标定方法,如最大似然估计(MLE)、边际最大似然估计(MMLE)方法和交差方法用于测验 Q 矩阵的修改,^[28]结果显示,MMLE 及交差方法表现最好,好于 δ 和 γ 方法。^① 给定原始 Q 阵,MLE、MMLE 和交差方法与 δ 方法的实现步骤相似,只在 EM 算法中增加更新 Q 阵步骤。确定性输入噪音与门模型下属性指定和验证方法研究相对比较独立,鲜有研究对各方法进行综合而系统的比较。

2.6 其他方法

证据中心设计和认知设计系统方法都是原则评估设计的方法,要求在测验设计和分析之前须构建初始的认知模型。证据中心设计和认知设计系统的应用均是建立在已有的认知加工模型基础之上,然后进行测验开发和分析,由于借助已有的认知模型,属性指定、选择和验证难度相对要小些。对属性指定、选择和验证主要采用口语报告、认知分析、回归分析、回归树、层次聚类方法和多维标定法等方法。前已涉及,不再赘述。

3 小结与讨论

在认知诊断评估中,构建正确的测验 Q 矩阵是保证诊断结果具有较高信度和效度的基本前提。本文对国内外近 30 年来相关研究进行了梳理,从几个知名的认知诊断模型和多篇经典的应用类文章出发,较为详细地归纳和评价各模型下 Q 矩阵构建方法。归纳了各模型下 Q 矩阵构建的基本步骤和涉及的定性和定量方法,并指出了各模型下 Q 矩阵构建方法的优劣,以期能为认知诊断测验的 Q 矩阵构建提供借鉴。

尽管各模型下 Q 矩阵构建的具体方法各不相同,但大致上可概括为以下三个步骤。首先,根据文献调查和已有相关理论研究,由学科专家得出初始属性集或认知模型。然后,从定性的角度选择和验证认知模型,如编码项目特征,通过任务分析项目特征与属性的关系,或通过学生口语报告,分析得出学生作答项目时所用到的属性,或通过专家解答样题,依据经验标定项目属性,并辅以内容专家对属性进行评分,选择正确作答项目要求的属性。再从定量的角度选择和验证认知模型,如规则空间模型下的被试分类率和多元回归分析结果作为属性增加、选择或验证的统计依据,属性层级模型采用层级相合性指标选择和评价认知模型,融合模型采用逐步简化算法进行 Q 阵元素修正,层次聚类方法和多维标定法可间接的识别项目特征,回归树主要用于选择重要属性等。

在简要介绍认知诊断评估的步骤之后,从认知诊断模型视角,重点阐述和评价了各模型下测验 Q 矩阵中属性指定、属性选择和属性验证的方法。Q 矩阵界定是认知诊断评估实践中至关重要的一步,是认知诊断评估理论区别于传统测量理论的根本所在。但由于属性本身定义的多样性,及其潜在性等特点,要准确、客观、高效地构建 Q 矩阵还有较长的路要走。未来 Q 矩阵的研究至少应该着重以下几个方面:

(1) 需要强调认知理论、教育理论和学科专家等在 Q 矩阵构建中的指导作用。需要进行相关文献调查和实证研究,为 Q 矩阵中属性的解释提供更多认知加工证据。当前有研究试图采用完全统计的方法进行属性标定,这样做可以大大降低成本,节约时间,但得出的属性极需赋予实质的含义。在这种情

^① Wang W. - Y., Ding S. - L., & Song L. - H. New Q - matrix validation methods and their sensitivity under the DINA model, Paper presented at the 2013 NCME Annual Meeting, San Francisco, CA.

况下,需要教育测量工作者做出不懈的努力,积极寻求与心理学和学科专家的合作,构建具有理论和经验支持下的测验 Q 矩阵。这应该作为从事教育测量工作者今后的一项重要工作。

(2) 针对性设计和开发用于诊断目的诊断测验。目前为止,多数认知诊断评估都是对已有测验进行分析,其 Q 矩阵是从已有测验翻新出来的。有研究者对这种做法提出了异议,认为不妥。Jang 在探讨像托福这样的标准化测验是否可用于诊断目的时也发现,标准化测验的心理计量本质与诊断评估特征存在明显冲突。多于 20% 的测验项目不能提供诊断信息,且这些项目的使用还会对诊断结果造成负面的影响。^[14]若能与学科专家协作,并在认知模型的指导下设计和开发测验,有针对性设计和开发诊断测验,将有利于属性标定和提高测验的诊断信度和效度。

(3) 有待开发包容各类模型且结合理论和数据的 Q 矩阵指定或验证方法。因为社会各界对于诊断性评估的需求十分迫切和能直接应用于各具体教育测量领域的认知模型较少,不同领域中学生的心理加工过程还不明了,要在各个领域诊断评估中,均要求在认知模型指导下开发诊断测验,在认知诊断评估真正崛起之前条件尚不完全成熟。尽管翻新的方法进行诊断评估,有一定的不足。还是非常有必要开发出相应的方法,能将各种各样的已有测验及其测验数据利用起来,挖掘出有诊断价值的项目和测验 Q 矩阵等。由于测验领域的多样性,从心理学的角度,被试作答过程中属性之间可能适合使用补偿模型,有的情况适合采用非补偿模型。因此,至少应该针对不同的认知假设,开发针对性的 Q 矩阵的探索性和修正性方法。

另外,测试项目反应背后的认知属性的识别是一个相当复杂的问题,属性的确定是一个不断探索和反复验证的过程,只有广泛开展 Q 矩阵的效度证据的收集,如通过实验设计调查和分析诊断结果的实际效果,才能确保 Q 矩阵构建符合真实的认知加工过程,从而保证诊断性评估对学生分类的可靠性和准确性,为后续教学与个性化发展提供宝贵信息。

参考文献:

- [1] 丁树良,罗 芬,汪文义. 认知诊断分类中心的确定[J]. 心理学探新, 2013, 33(5).
- [2] Tatsuoaka K. K. *Cognitive assessment: An introduction to the rule space method* [M]. New York: Taylor & Francis Group, 2009.
- [3] DeCarlo J. T. Recognizing uncertainty in the Q - Matrix via a bayesian extension of the DINA model [J]. *Applied Psychological Measurement* 2012, 36(6).
- [4] Jang E. E. Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion model application to LanguEdge assessment [J]. *Language Testing* 2009, 26(1).
- [5] Rupp A. A. & J. Templin. The effects of Q - matrix misspecification on parameter estimates and classification accuracy in the DINA model [J]. *Educational and Psychological Measurement* 2008, 68(1).
- [6] 汪文义,宋丽红. 教育认知诊断评估理论与技术研究[M]. 北京: 北京师范大学出版社, 2015.
- [7] Buck G., Tatsuoaka K. K. & Kostin J. The subskills of reading: Rule - space analysis of a multiple - choice test of second language reading comprehension [J]. *Language Testing*, 1997, 14(3).
- [8] Buck G. et al. Development, selection and validation of a set of cognitive and linguistic attributes for the SAT - I verbal sentence completion section (RR - 98 - 23) [M]. Princeton, NJ: Educational Testing Services, 1998.
- [9] 孙佳楠,等. 基于 Q 矩阵和广义距离的认知诊断方法[J]. 心理学报, 2011, 43(9).
- [10] Wang C. & M. J. Gierl. Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading [J]. *Journal of Educational Measurement* 2011, 48(2).
- [11] Gierl M. J., Alves C. & Majeau R. T. Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment [J]. *International Journal of Testing* 2010, 10(4).
- [12] 丁树良,杨淑群,汪文义. 可达矩阵在认知诊断测验编制中的重要作用[J]. 江西师范大学学报(自然科学版), 2010, 34(5).
- [13] Hartz S. M. A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality

- [D]. University of Illinois at Urbana – Champaign 2002.
- [14]Jang E. E. A validity narrative Effects of reading skills diagnosis on teaching and learning [D]. University of Illinois at Urbana – Champaign 2005.
- [15]Gao L. Y. Toward a cognitive processing model of MELAB reading test item performance [C]. Spaan Fellow Working Papers in Second or Foreign Language Assessment 4.
- [16]Sheehan K. M. A tree – based approach to proficiency scaling and diagnostic assessment [J]. *Journal of Educational Measurement* 1997 34(4) 2.
- [17]de la Torre J. An empirically based method of Q – matrix validation for the DINA model: Development and applications [J]. *Journal of Educational Measurement* 2008 45(4) .
- [18]涂冬波, 蔡 艳, 戴海琦. 基于 DINA 模型的 Q 矩阵修正方法 [J]. *心理学报* 2012 44(4) .
- [19]Liu J., Xu G. & Ying, Z. Data – driven learning of Q – matrix [J]. *Applied Psychological Measurement* 2012 36(7) .
- [20]Chiu C. – Y. Statistical refinement of the Q – matrix in cognitive diagnosis [J]. *Applied Psychological Measurement* 2013 37 (8) .
- [21]Chen P et al. On – line calibration methods for the DINA model with independent attributes in CA – CAT [J]. *Psychometrika* 2012 77(2) .
- [22]陈 平, 辛 涛. 认知诊断计算机化自适应测验中在线标定方法的开发 [J]. *心理学报* 2011 43(6) .
- [23]陈 平, 辛 涛. 认知诊断计算机化自适应测验中的项目增补 [J]. *心理学报* 2011 43(7) .
- [24]汪文义, 丁树良. 题库结构对原始题在线属性标定准确性之影响研究 [J]. *心理科学* 2012 35(2) 6.
- [25]汪文义, 丁树良, 游晓锋. 计算机化自适应诊断测验中原始题的属性标定 [J]. *心理学报* 2011 43(8) .
- [26]喻晓锋 等. 使用似然比 D2 统计量的题目属性定义方法 [J]. *心理学报* 2015 47(3) .
- [27]喻晓锋 等. 基于作答数据的模型参数和 Q 矩阵联合估计 [J]. *心理学报* 2015 47(2) .
- [28]汪文义, 宋丽红, 丁树良. 基于探索性因素分析的 Q 矩阵标定方法 [J]. *江西师范大学学报(自然科学版)* 2015 39 (2) .

(责任编辑: 余小江)