

Whether Examinees' Reading Abilities are in Line with Examination Structure —From the Perspective of Cognitive Diagnosis

Duan Huiqiong Jiang Yuwen*

School of Foreign Languages, Nanchang Hangkong University, Nanchang

Abstract: In this study, eight cognitive factors that stimulate the reading process were identified by cognitive diagnosis, and the probability of each cognitive factor being mastered by the examinee was calculated. Combining with expert opinions, this paper analyzes the structural bias of reading test, and explores the relationship between the examination structure and students' reading ability. It is pointed out that teachers should attach importance to the cognitive factor that has a low presence in the examination and carry out special training for students. In order to improve the "examination structure mechanism", this study designed a cognitive diagnostic table in order to comprehensively improve students' reading level.

Key words: Cognitive diagnosis; Reading ability; Exam structure; Cognitive factor

Received: 2020-07-20; Accepted: 2020-07-29; Published: 2020-09-17

阅读测试内容与考生阅读能力一致性研究

——基于认知诊断视角

段惠琼 江钰文*

南昌航空大学外国语学院, 南昌

邮箱: 648429296@qq.com

摘 要: 为探究阅读考试内容是否考察考生阅读能力, 基于认知诊断理论对 332 名大学生进行四级阅读测试, 从有声思维报告和四级阅读测试要求中析出阅读过程中出现的 8 个阅读认知因子, 构建 Q 矩阵, 对考生阅读能力因子进行分析, 计算出考生对每一认知因子的掌握概率。结合专家意见, 指出教师应重视考试结构中出现较少的认知因子, 对学生进行专项训练。为完善考试结构机制, 设计了认知诊断信息表, 旨在综合提高学生的阅读水平。

关键词: 认知诊断; 阅读能力; 考试结构; 认知因子

收稿日期: 2020-07-20; 录用日期: 2020-07-29; 发表日期: 2020-09-17

Copyright © 2020 by author(s) and SciScan Publishing Limited

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

<https://creativecommons.org/licenses/by/4.0/>



1 引言

英语阅读贯穿于整个英语学习过程中,是英语习得者必须掌握的基本技能之一。然而在英语学习过程中,每一位英语学习者的阅读水平或阅读能力存在明显差异。在现阶段的教学过程中,考试是一个可行的评估方法,是测试考生英语阅读能力的主要手段。考生英语阅读能力通过单一终结性分数呈现出来——考生在阅读考试中得分越高,则越肯定其阅读能力;反之,则认为该考生阅读能力有所欠缺。尽管上述说法具有一定真实性,却也存在着片面性。该结论强调了学生的宏观阅读能力,把阅读能力视作最小单位,忽略了考生在微观层面上的一系列内部心理认知过程。考生的心理认知过程都有哪些?如何通过考试检测出学生的微观认知过程?考试结构是否能真实有效测试出考生的阅读能力,即考生阅读能力与考试结构是否相同?上述问题值得探究。

国内外有许多关于考生阅读能力的研究,但大多数是基于CTT理论(经典测验理论)[1]进行研究,这样的研究结果必然是宏观的。为了探寻考生微观认知过程,国外在阅读认知诊断评估取得一些进展,但国内相关研究还比较少。Kasai(1997)和Jang(2005)通过考生在TOEFL考试中的表现,界定出几种认知属性,包括词汇、语法、句子之间关系等属性[2][3]。但其结果却缺乏针对性,在一定程度上不符合大学生英语阅读水平情况。Wang和Gill(2011)基于属性层级方法对SAT考生的认知技能进行认知分析[4]。其结果得出多采用1/0的二分法。该方法只能判断学生掌握或未掌握,至于掌握多少却不得知晓。

综上所述,本研究基于认知诊断视角,构建Q矩阵,对考生阅读能力因子进行分析;结合树回归分析的方法对考试的信度进行分析,以求得一个更为准确的测验结果。同时,运用G-DINA模型,测算出考生对认知因子的掌握概率。该研究不仅分析了考生阅读过程中的认知因子,还要探寻考试结构是否科学有效。旨在为学生、老师及家长提供一个更为有效的考试机制,其最终目的是帮助学生发现自我问题,进行自我改进,最终实现自我提升。

2 认知因子析出及研究过程

2.1 认知因子的假设

本研究结合教育部 2017 年出台的《大学英语教学指南》和大学生英语四级阅读测试要求，再依据大学生阅读教学中的实际情况及自身经验析出了 9 种认知因子。根据前人的研究，将这 9 种因子分成“文本内阅读”和“文本外阅读”两个范畴；又将文本内阅读分为“语言知识”和“阅读策略”两个层面 [5] [6] [7] [8] [9] [10] [11] [12] [13]。请见表 1。

因子 A1、A2、A3、A4 属于文本内阅读范围内有关语言知识认知成分，A5、A6、A7、A8 是关于阅读策略技能的认知成分。上述八种认知因子都属于文本范围内阅读。而 A9 属于文本外的阅读范畴，指的是考生根据自身知识或感觉做出判断。

表 1 英语阅读中的认知因子

Table 1 Cognitive factors in English reading

类别	语言框架	认知因子	名称	定义
文本内阅读	语言知识	A1	词汇基础知识（词汇与语法）	通过句法、语法特点等识别特定单词或短语词性以及固定搭配的认识
		A2	词意	通过上下文线索、文本分析确定词的字面意义或深层含义
		A3	句意	识别句子结构（如倒装、并列、转折、让步）推出句意；通过语法、句法、标点、修辞等，理清句子之间关系
		A4	段落大意	理解段落大意
	阅读策略	A5	提取匹配信息	识别问题、选项、文本中相关的词汇、句子、段落信息及同义表达
		A6	剔除不合理选项	正确理解问题及选项对应的短文信息，并准确排除错误选项的能力
		A7	测试 / 分析事实和细节	识别文本相关句子意思，进行逻辑推断，判断事实，分辨细节
		A8	综合概况主旨、作者态度观点	根据关键词，进行推断主题思想及支持句，判断作者观点态度
文本外阅读		A9	文本外推测和认知能力	根据自身背景知识或直觉做出判断

2.2 研究过程

为了证实上述认知因子,本研究以2016年大学英语四级为考试材料,采取有声思维方式,从定性角度分析认知因子的有效性。接着构建Q矩阵,使用SPSS中的分层回归方式测试考试的信度,从定量的角度进行验证,证明结果的准确性。

2.2.1 考试材料

采用2016年大学英语四级中的阅读理解部分为试题,合计四篇阅读,共30个选项。由于本研究最终目的是测试考试结构的合理性,所以并未自行设计考试题型。本次测验的 α 系数为0.766,奇偶分半信度为0.763,表明此次测验的可靠性良好。根据大学生四级考试要求,规定学生要在40分钟内完成试题。考试结构和分值如下表2所示。

表2 测试结构

Table 2 The test structure

试卷结构	测试内容	测试题型	题号	分值	考试时间
阅读理解	词汇理解	选词填空	26 ~ 35	0.5分/题	40分钟
	长篇阅读	匹配	36 ~ 45	1分/题	
	仔细阅读	多项选择	46 ~ 55	2分/题	

2.2.2 参与测试对象

从南昌某大学2018级大一学生中抽取332名学生进行考试,并且依据考试成绩划分了三个等级。考试成绩在30分以上为优秀,在25~30分之间为良好,在10~20之间为合格。从这三个等级中平均抽取共32名同学进行有声思维作答。

2.3 研究结果与分析

2.3.1 Q矩阵的构建

根据32份多达33582字和12名来自全国范围的英语和非英语专业研究生3万多字的有声思维,经过学科专家确认审核,从语料中提取认知因子,构建

初始矩阵。本研究有 9 个认知属性，所建构初始 Q 矩阵本应该是一个 10 行，31 列的矩阵表。在研究过程中，有声思维报告被认为是主要证据，因为口头报告或多或少捕获了实时阅读过程。然而，专家的意见也不应被低估。专家认为认知因子 A9 属于文本外阅读，是考生根据自身的知识背景或直觉做出判断的。也就是说，考生可能只是猜测或者可能已经没有时间，却并没有煽动项目解决过程。因此，在 Q 矩阵的建立中，本研究将 A9 剔除。最终本研究构建的 Q 矩阵是一个 9 行，31 列的矩阵表。根据前人研究 [14] [15]，一个属性被测量的次数不应少于 3 次，否则对该属性的诊断误差会较大。剩余的 8 个认知因子，都至少被 3 道题目测量，因此没有因子从 Q 矩阵中剔除。Q 矩阵采用二元计分，矩阵中的数字 1 表示该技能是项目所需要的，而数字 0 表示该技能是项目没有涉及或者不需要的。所建 Q 矩阵如下表 3 所示：

表 3 认知因子 Q 矩阵表
Table 3 Q matrix built with cognitive factors

属性	项目编号																																
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	54			
A1	1	1	1	1	1	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1			
A2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A3	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A4	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1			
A5	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A6	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1	1	1	1	0	1	1	0	1			
A7	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	1	0	1	1	1	1	0	0	0	1			
A8	0	0	0	0	0	0	1	1	0	1	1	1	0	1	0	1	0	0	0	1	1	1	0	1	1	1	1	0	0	1			

2.3.2 分层回归分析

本研究使用 SPSS 采用分层回归分析模型从定量的角度验证认知因子的有效程度。上述八个认知因子已经被分为“语言知识”和“阅读策略”两个部分。将“语言知识”中 A1 ~ A4 四个认知因子作为第一层变量带入分层回归方程中，构建模型 I；再将“阅读策略”中的 A5 ~ A8 两个认知因子作为第二层变量进入回归方程 I 中，构建模型 II。

模型 I、II 均属于有效的分层回归模型,其自变量都可用于解释或预测因变量。根据表 4 中的数据,模型 I 的解释方差为 46.75%,模型 II 的解释方差为 78.5%, ΔR^2 为 0.3041, p 值小于 0.5(差异显著)。模型 II 的解释方差大于模型 I,其拟合程度更高,且数值超过 60%,为认知因子的有效性和实用性提供了有力的支撑。

表 4 分层回归模型 I、II 的数据输出

Table 4 Data output from hierarchical regression model I, II

分层回归模型	R	R^2	ΔR^2	p
I	0.6641	0.4675	0.3041	$p < 0.5$
II	0.8692	0.785		

3 学生阅读认知能力和考试结构比较

3.1 大学生认知能力分析

本研究试图探究考生阅读能力与考试结构是否相同,就必须了解学生对每一认知因子的掌握概率,以及考试题型对不同认知因子的考察程度。从定量的角度分析学生阅读认知水平与考试测点结构是否相符。

3.1.1 整体认知水平

根据南昌航空大学 332 名大一学生的测试情况,结合已得的英语阅读认知因子 Q 矩阵,进行认知诊断分析。本研究将数据导入 R 软件中,进行 G-DINA 模型分析。G-DINA 模型拟合程度是项目对的 Fisher 转换观测相关和 Fisher 转换预测相关之间的残差(Residual between the observed and predicted correlation of item pair with the Fisher transformation)以及项目对的观测对数差异比和预测对数差异比之间的残差(Residual between the observed and predicted log-odds ratios of item pair),统计假设以上残差等于 0,那么残差与 0 的差异不显著时,模型与数据拟合程度较好;反之,即残差与 0 的差异显著时,模型与数据拟合程度则

不佳 [16]。本次研究中, $p > 0.14$, 可得 G-DINA 模型与学生作答情况及认知因子 Q 矩阵拟合程度较高。

由图 1 可知, 大学生在阅读过程中对每一认知因子的掌握概率。总体而言, 学生在语言知识层面 (A1 ~ A4) 的表现比在阅读策略 (A4 ~ A5) 的表现更好。在八个认知因子中, 学生对词汇与语法 (A1: 0.78) 的掌握概率最高, 这表明中国大学生在学习英语过程中非常注重语法学习; 从中学到大学, 语法学习始终贯穿其中, 使得我国大学生在语法和词汇上的把握程度较好, 也更熟悉有关语法类的题型。除此之外, 学生在句意 (A3: 0.72) 和段落大意 (A4: 0.78) 的把握上也呈现出一个较为良好的水平。这是由于二语习得者在阅读过程中习惯将第二语言转化为自己所熟悉的母语, 并且结合自身词汇与语法知识, 识别句子结构, 理清句子关系, 最终明白句子及段落意义。在语言知识层面中, 词意 (A2: 0.65) 的掌握概率最低, 也就是说中国学生在推断词意的能力方面稍显欠缺, 因为认知因子 A2 需要被试者根据上下文推断出生僻词汇的意义; 尽管学生对句子意义及段落大意都有较为清晰的理解, 这样的推理认知过程对学生来说依然具有一定难度。正是因为学生在词意 (A2) 上的理解有所偏差, 导致其在分析事实和细节 (A7: 0.55) 上的表现也不甚理想。

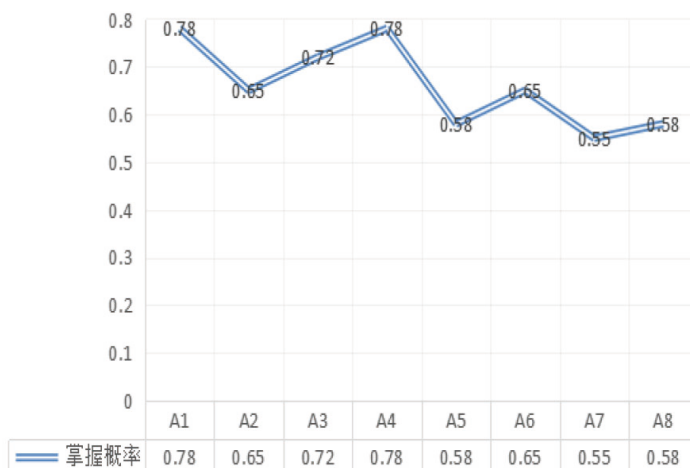


图 1 整体掌握概率

Figure 1 Global mastery probability

考生在推理过程中出现失误,很容易影响其对细节事实的判断;即使词意理解正确,分析事实这一认知加工过程对考生来说也是一大难点。前面提到句意 A3 和段落大意 A4 的掌握程度好,并不意味着词意 A2 的掌握概率就高;同样地,学生在主旨大意和作者态度(A8: 0.58)上也呈现出一个较低的表现水平。原因是把握主旨大意及作者态度判断题型对学生推理、总结等能力有一定要求,此类认知加工过程属于中国大学生阅读过程中的薄弱环节。考生在认知程度较低的题目中得分也相对要高。为了更全面地提高学生阅读能力,在平时的教学过程中,教师应该多煽动相关认知加工过程。

3.1.2 个体阅读认知水平

整体平均认知水平并不能完全代表个体认知水平,相同得分的考生对每一认知因子的掌握情况也不一样。因此,本研究抽取了 A、B、C、D、E 五位得分相同的同学,并用雷达图展现个体认知掌握概率的不同。

从图 2 可以看出这五位同学对 A1、A3、A4、A5 这四种认知因子的掌握程度差别不大,有关阅读策略的认知因子掌握概率则展现出比较大的差异,例如 A6、A7、A8。A2 虽然归于语言知识层面,但是该项因子涉及到推理过程,因此差异也比较大。由此,既可以验证整体水平掌握概率的准确性,同时又能清楚认识到不同个体之间的认知差异。

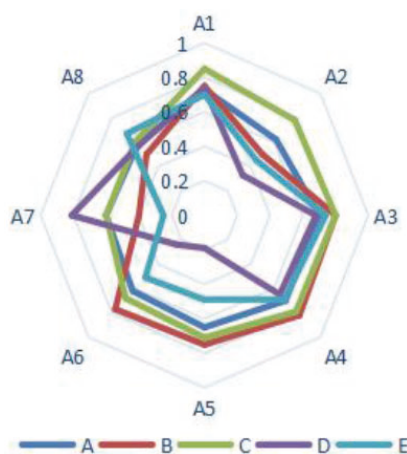


图 2 个体认知水平比较

Figure 2 Comparison of individual cognitive level

3.2 大学四级阅读考点倾向

CET-4 (College English Test Band Four) 作为国家教育部高等教育司主持的全国性英语能力测试,旨在对大学生的英语能力进行客观、准确的测量,其阅读测试部分包含三种题型:选词填空、长篇阅读匹配、阅读理解。以往的研究大部分围绕四级阅读考试的难易程度或者信度的高低展开,却很少有研究分析试题中侧重考察哪一类认知因子,即四级阅读测试对某一类认知因子的倾向程度与考生认知掌握概率有无一定联系。为了探究这一点,本研究请5位专家对四级阅读测试进行认知因子的评估,五位专家有着丰富的大学英语教学经验,教龄均超过10年。如果3位及以上专家认为在某一题阅读过程中涉及某一认知因子(同意率大于等于60%),则认为该题需要运用此项认知因子;若2位及以下专家标注了某一因子(同意率小于等于40%),则认为该题并未检测这一认知因子。在30道阅读题中,认知因子的测量次数统计如图3所示。

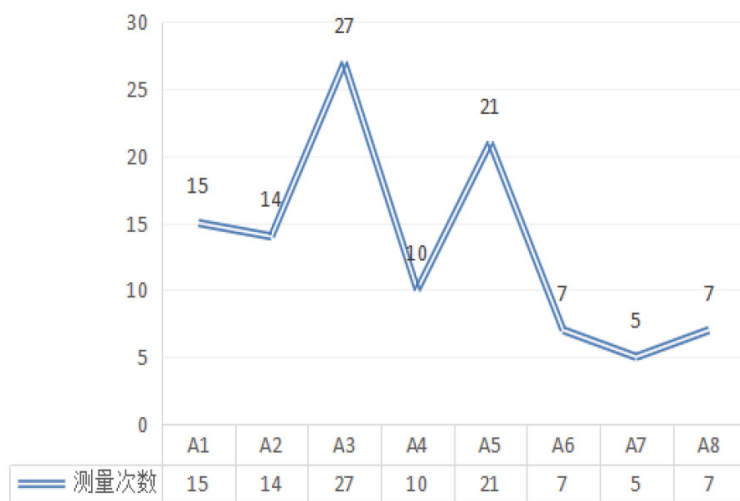


图3 测量次数

Figure 3 Number of measurement

由图3可以看出,认知因子A3(词意)的测量次数最多,高达29次;排在第二的是认知因子A5(提取匹配信息),测量次数为25次;其次为认知因

子 A3, 次数为 15 次。在测试过程中对 A6、A7、A8 认知因子的测量次数较少, 均少于 10 次。由上述数据可以看出, 大学英语四级阅读测试难度适中——相较于复杂的阅读策略而言, 更倾向考察学生的语言基础知识能力。

3.3 比较结果

由上述两小节可知学生认知因子 A1 至 A4 的掌握概率比较好; 而大学英语四级阅读考试对 A1 至 A4 的考察也比较频繁, 次数均在 10 次以上。这四个认知因子考察情况比较吻合。本研究认为这是符合中国国情的数据显示, 中国外语教学中注重语言基础, 例如语法、词汇、句子结构的学习等。这样的重视不但能够给学生打下一个良好语言基础, 同时也能潜移默化地影响着考试结构, 使得考试侧重点偏向于“语言知识”层面。如此一来造成了一个“循环局面”, 即考试越考什么, 学生就学什么。

让学生在阅读测试过程中丢分的往往是涉及 A5 至 A8 这四种认知因子的题型。从图 3 中也可以发现, 排除测量次数达 21 次的 A5 在外, 其余 3 个认知因子被测次数均小于 10。而学生对 A5 这一认知因子的掌握概率仅为 0.58, 可见在阅读过程中, 学生很容易定位至错位信息。而 A6、A7、A8 这三项认知因子在整体掌握概率和测量次数上, 似乎存在一种正相关关系, 即考察次数较少, 学生掌握概率也较低。

综上所述, 考试结构与考生能力同一性高。考试中所侧重考察的认知因在, 学生掌握程度较好; 而考试中测量次数偏低的认知因子, 学生掌握概率较低, 并且在此类题型中, 学生更容易失分。考试结构与考生能力间存在的明显差异性因素, 则成为学生阅读水平低的主要原因之一。

4 启示: 完善阅读测试机制

每一位考生的认知水平都存在差异, 即使是考试得分相同的学生, 其认知结构也存在差异性。从两千多年前的孔子的“因材施教”到如今提倡的个性化教育, 都否定了划一式的教育机制, 强调承认个体差异。考试作为学生学习过程中的必要手段, 理应顺应时代, 发展“个性化考试”。但是, 考试也是学

生能力的检验手段，必须具有划一性；只有这样才能让学生知道在大环境中自身的水平是好是坏，自身的优势在哪里，不足又在哪里。因此，制定个性化试卷是不现实的。如若在考试机制中增加一个必要的环节——制定“考试认知诊断结果分析表”，则能够帮助学生更加明确自身的学习状况。

制定该表有以下三个优点：一是学生能够明确自身水平；二是学生可以认识到阅读过程中存在的不足；三是针对性地帮助学生制定今后学习重点。表格设计如表 5 所示，表格中对学生的每一项认知因子以数据的形式给出，并标明每一项认知因子掌握概率的平均值，这样的设计清楚直观地反映出学生的群体位置以及掌握情况。表格最后设有教师建议栏，学生可以根据教师建议的内容，在今后学习中弥补自身不足。

表 5 认知诊断信息表

Table 5 Cognitive diagnostic information sheet

阅读测试		
	考生姓名	考生得分
各项阅读技能掌握情况		平均分
	考生水平	平均水平
A1 词汇与语法		
A2 词意		
A3 句意		
A4 段落大意		
A5 提取匹配信息		
A6 剔除不合理选项		
A7 测试 / 分析事实和细节		
A8 综合概况主旨、作者态度观点		
教师建议		

经研究发现考试结构与考生能力基本相同但仍存在一定偏差。考生对考试测量次数较少的因子掌握概率普遍偏低。因此，教师应该多关注测量测试较少的认知过程，关注学生实际阅读过程中的薄弱不服，对学生进行专项训练。本

研究不再以单一的分数作为评估学生的唯一标准,而是主张一种多维度的认知诊断测评,这对外语教学具有实际意义。

基金项目

江西省研究生创新专项资金项目课题“认知诊断视角下英语阅读能力与测试倾向契合度研究”(YC2019033)。

参考文献

- [1] Spearman C. The proof and measurement of association between two things [J]. American Journal of Psychology, 1987, 100 (3/4): 441-471.
<https://doi.org/10.2307/1422689>
- [2] Kasai M. The rule space model applied to the reading comprehension section of the Test of English as a Foreign Language [D]. University of Illinois at Urbana-Champaign, 1997.
- [3] Jang E E. A validity narrative: effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL [D]. University of Illinois at Urbana-Champaign, 2005.
- [4] Wang C J, Gierl M. Using the attribute hierarchy method to make diagnostic inferences about testees' cognitive skills in critical reading [J]. Journal of Educational Measurement, 2011, 48: 165-187.
<https://doi.org/10.1111/j.1745-3984.2011.00142.x>
- [5] Embretson S E, Wetzel C D. Component latent trait models for paragraph comprehension tests [J]. Applied Psychological Measurement, 1987, 11 (2): 175-193. <https://doi.org/10.1177/014662168701100207>
- [6] Kirsch I S, Mosenthal P B. Understanding document literacy: variables underlying the performance of young adults [J]. ETS Research Report Series, 1990 (2). <https://doi.org/10.2307/747985>
- [7] Buck G, Tatsuoka K, Kostin I. The subskills of reading: Rule-space analysis

- of a multiple-choice test of second language reading comprehension [J]. Language Learning, 1997, 47 (3): 423-266.
<https://doi.org/10.1111/0023-8333.00016>
- [8] Sheehan, Ginther. What do passage based multiple choice verbal reasoning items really measure? An analysis of the cognitive skills underlying performance on the current TOEFL reading section [C]. The annual meeting of the National Council of Measurement in Education, New Orleans, LA, 2000.
- [9] Jang E E. Cognitive diagnostic assessment of l2 reading comprehension ability: validity arguments for fusion model application to [J]. Language Testing, 2009, 26 (1): 31-73. <https://doi.org/10.1177/0265532208097336>
- [10] Gao L, Rogers W T. Use of tree-based regression in the analyses of l2 reading test items [J]. Language Testing, 2011 (28): 77-104.
<https://doi.org/10.1177/0265532210364380>
- [11] 陈慧麟, 陈劲松. G-DINA 认知诊断模型在语言测验中的验证 [J]. 心理科学, 2013, 36 (6): 1470-1475.
- [12] 张玉美, 罗少茜. 基于认知诊断模型的英语阅读测试诊断信息反馈 [J]. 外语测试与教学, 2018 (3): 55-60+64.
- [13] 杜文博, 马晓梅. 基于认知诊断评估的英语阅读诊断模型构建 [J]. 外语教学与研究 2018, 50 (1): 74-88+160-161.
- [14] 涂东波. 项目自动生成的小学儿童数学问题解决认知诊断 CAT 编制研究 [D]. 江西师范大学, 2009.
- [15] 涂东波, 蔡艳, 丁树良. 认知诊断理论、方法与应用 [M]. 北京: 北京师范大学出版社, 2012.
- [16] Chen J, Jimmy D L T, Zhang Z. Relative and absolute fit evaluation in cognitive diagnosis modeling [J]. Journal of Educational Measurement, 2013, 50 (2): 123-140. <https://doi.org/10.1111/j.1745-3984.2012.00185.x>