

汉语二语测试中文背景知识隐性考查研究

——基于Byram模型的多维构念与测试优化路径

吴明坤

大连大学文学院, 辽宁 大连

收稿日期: 2026年6月22日; 录用日期: 2026年7月31日; 发布日期: 2026年8月14日

摘要

语言和文化之间的联系非常紧密, 在汉语二语测试进行的过程中, 当对语言技能进行考查时, 必然会涉及对中国文化背景知识的评估。同显性考查不一样的是, 隐性考查会把文化背景知识融入到语言材料、任务设计里, 这样考生在完成语言理解任务时, 其文化认知水平也就能够展现出来。本文依据隐性知识与显性知识理论框架, 借鉴Byram跨文化交际能力模型, 将“隐性文化知识”构念细化为知识、技能、态度与批判性文化意识四个维度, 并引入验证性因素分析等现代测量模型探讨效度验证路径。在案例分析部分, 本文覆盖HSK初级、中级、高级三个水平, 涉及风俗禁忌、语用规则、言外之意、网络文化及当代社会价值观等多样化主题, 深入剖析其背后的文化图式并提供试题来源信息。最后, 本文提出包含“语言准确性”“语用得体的性”“文化适切性”三维度的评分量表草案, 建议利用社交媒体、影视剧等真实语料库开发测试任务, 并为初、中、高三个等级提出具体的隐性文化能力表现描述, 以为汉语测试的命题优化与文化教学提供参考。

关键词

汉语二语测试, 文化背景知识, 隐性考查, HSK, 跨文化交际

Research on Implicit Assessment of Chinese Background Knowledge in Chinese Second Language Testing

—Multidimensional Construct and Test Optimization Based on Byram Model

Mingkun Wu

School of Chinese Language and Literature, Dalian University, Dalian Liaoning

Abstract

The connection between language and culture is extremely close. During the process of Chinese second-language testing, when assessing linguistic skills, it inevitably involves evaluating knowledge of China's cultural background. Unlike explicit assessment, implicit assessment integrates cultural background knowledge into language materials and task designs, allowing candidates' cultural competence to be demonstrated when completing language comprehension tasks. Based on the theoretical frameworks of explicit and implicit knowledge and Byram's model of intercultural communicative competence, this paper refines the construct of 'implicit cultural knowledge' into four dimensions: knowledge, skills, attitudes, and critical cultural awareness. It also introduces modern measurement models such as confirmatory factor analysis to explore validity verification pathways. In terms of case analysis, this paper covers three proficiency levels (HSK elementary, intermediate, and advanced), involving diverse themes such as customs and taboos, pragmatic rules, implied meanings, internet culture, and contemporary social values. It provides detailed analysis of the underlying cultural schemata and sources of test items. Finally, this paper proposes a draft rubric comprising three dimensions: linguistic accuracy, pragmatic appropriateness, and cultural appropriateness. It suggests utilizing authentic corpora such as social media and film and television dramas to develop test tasks, and provides specific performance descriptors of implicit cultural competence for elementary, intermediate, and advanced levels, aiming to offer valuable references for improving test design and cultural instruction in Chinese language testing.

Keywords

Chinese Second Language Testing, Cultural Background Knowledge, Implicit Assessment, HSK, Cross-Cultural Communication

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

随着国际中文教育的蓬勃发展,汉语水平考试(HSK)等标准化测试已成为评估学习者语言能力的重要工具。语言测试不仅关注语音、词汇、语法等语言要素,更日益重视学习者的语用能力与跨文化交际能力[1]。汉语二语测试里,文化背景知识的考查有着显性、隐性这两种方式,显性考查直接把文化知识当作测试内容,在专门的文化常识选择题中呈现;隐性考查是把文化背景知识融入到语言材料之中,要求考生在理解语言形式的时候调动自身文化认知图式来完成理解任务[2]。

隐性考查方式在二语测试领域具有重要价值。一方面,它符合交际语言测试理论的真实性原则,能够模拟真实实际情境中语言与文化的融合状态[3];另一方面,它降低了考生因文化知识缺失而产生的考试焦虑,使测试更聚焦于语言运用能力而非文化记忆能力[4]。然而,现有研究对“隐性文化知识”这一构念的界定仍较为笼统,案例分析多集中于 HSK 中级水平,对初级与高级水平的隐性文化考查关注不足,且优化建议多停留在宏观层面,缺乏可操作的评分工具与等级标准。基于此,本文以汉语二语测试为研究对象,在深化理论构念、扩展案例广度与深度、提出具体优化路径三个层面展开系统探讨。

2. 理论基础与隐性考查的构念界定

2.1. 从二元对立到多维构念：隐性文化知识的理论重构

二语习得研究中，隐性知识(implicit knowledge)与显性知识(explicit knowledge)的区分是理解隐性考查方式的理论基础。Bialystok 指出，显性语言知识存在于学习者的意识层面，可以被清晰表达；而隐性语言知识则内化为学习者的直觉语感，难以用语言明确描述，但能在实际交际中自动运用[5]。Ellis 则提出，人的语言能力主要由隐性知识构成，隐性知识是流利、自动化语言产出的基础[6]。

把这一理论框架拓展到文化背景知识范畴，能够区分出显性文化知识和隐性文化知识。显性文化知识是指学习者可以清晰表述的文化事实，如春节的日期、长城的长度等；隐性文化知识是指学习者在跨文化交际时自动运用的文化图式、语用规则与价值观念，例如面对称赞时的自谦回应、对“面子”概念的直觉性理解等[7]。然而，将隐性文化知识简单视为“不可言传的文化直觉”过于笼统，不利于测试构念的操作化界定。

本文主张借鉴 Byram 跨文化交际能力模型(model of intercultural communicative competence)，将隐性文化知识重构为四个可测维度：第一，知识维度(savoirs)，指学习者对中国社会群体及其文化产品、文化实践的程序性知识，如了解“红包”不仅是金钱馈赠，更承载着人际关系维护的功能；第二，技能维度(savoir comprendre/savoir faire)，指学习者在特定文化情境中解释、关联与互动的能力，如能否根据对话者的身份地位选择恰当的称呼策略；第三，态度维度(savoir être)，指学习者在跨文化交际中表现出的好奇心、开放性与对差异的尊重，如面对不同于母语文化的餐桌礼仪时能否避免价值判断；第四，批判性文化意识维度(savoir s'engager)，指学习者能够基于明确标准对中国文化现象进行批判性评价与反思的能力，如理解“随礼”习俗背后的社会逻辑同时意识到其可能带来的经济负担。这四个维度构成了隐性文化知识的多维构念，为后续测试设计提供了理论依据[8]。

2.2. Byram 跨文化交际能力模型的测试映射

Byram 模型强调跨文化交际能力并非单一维度的语言能力，而是知识、技能、态度与批判性意识的整合体[9]。在汉语二语测试中，隐性考查正是通过语言任务间接测量上述维度的综合表现。具体而言，知识维度可通过语境推理型与文化负载词型题目进行隐性测量；技能维度可通过语用得体的跨文化对比型题目进行考查；态度维度虽难以通过纸笔测试直接测量，但可通过考生在开放性任务中的选择倾向间接推断；批判性文化意识维度则可通过要求考生对文化现象进行评价、比较的任务来触及。

将隐性文化知识从“二元对立”(隐/显)拓展为“多维构念”(知识-技能-态度-意识)，有助于解决效度验证中的构念混淆问题。当考生在文化负载词题目上作答错误时，测试开发者可以依据多维框架判断：该错误源于词汇语义知识缺失(语言构念)，还是源于文化背景知识缺失(文化构念)，抑或是源于跨文化态度偏差(态度维度)。这种精细化的构念界定，为后续使用现代测量模型进行验证奠定了基础。

2.3. 现代测量模型与效度验证路径

传统的效度验证多依赖经典测量理论(CTT)，难以处理隐性文化知识的多维潜质与语言能力的交叉影响。本文建议引入项目反应理论(IRT)与验证性因素分析(CFA)等现代测量模型。具体而言，可采用多维度项目反应理论(MIRT)构建包含“语言能力”与“隐性文化知识”两个潜变量的测量模型，通过分析题项在两个潜变量上的载荷/loading)，识别那些主要测量文化知识而非语言能力的“锚题”(anchor items)。

此外，结构方程模型(SEM)可用于检验 Byram 四维度构念的拟合优度。研究者可以设定一阶四因素模型(知识、技能、态度、意识)与二阶单因素模型(总体隐性文化知识)，通过比较 CFI、TLI、RMSEA 等

拟合指标,判断多维构念的理论有效性。若四维度模型拟合显著优于单维度模型,则可为隐性文化知识的多维性提供实证支持。同时,多群组验证性因素分析(MG-CFA)可用于检验测量不变性(measurement invariance),确保来自汉字文化圈与非汉字文化圈的考生在题项理解上具有相同的测量结构,从而回应文化公平性关切。

3. 典型案例分析

本节在原有 HSK 四级案例基础上,新增 HSK 初级与高级案例,并扩展至网络文化与当代社会价值观主题。每个案例均标注来源信息,并从 Byram 四维度框架剖析其关联的文化图式。

3.1. HSK 初级: 数字文化图式的隐性考查——“八”与“发”

试题来源: HSK (一级)阅读真题,2018 年 3 月卷,第 23 题[10]。材料呈现一个简单对话:“这个电话号码是 1388888……”“太好了,都是 8!”题干要求考生选择说话者“高兴”的原因。

文化图式剖析:该题隐性考查了汉语数字谐音文化图式。在中国文化语境中,数字“8”因谐音“发”(发财、发达)而被赋予吉祥寓意,这一文化图式已深度嵌入日常生活,母语者对此具有自动化的直觉反应。对于初级水平学习者而言,若缺乏该文化图式,即使理解“8”的字面意义与“高兴”的情感色彩,也无法建立二者之间的因果关联。从 Byram 模型看,此题主要测量知识维度(savoirs)——学习者是否掌握数字谐音这一文化事实;同时涉及技能维度(savoir comprendre)——能否将语音联想转化为语义推理。

教学启示:初级文化教学应关注“高频生活文化图式”,将数字、颜色、动植物等基础符号的文化联想纳入词汇教学,而非留到高级阶段集中讲授。

3.2. HSK 中级: 风俗禁忌的隐性考查——“不要用筷子敲碗”

试题来源: HSK (四级)听力真题,2016 年 11 月卷,第 15 题。在中国的传统文化里,用筷子去敲击碗碟会被看作是不礼貌的行为,这和乞丐讨饭时的形象有所关联,是属于餐桌礼仪方面的禁忌内容。这道题目不是直接去问“用筷子敲碗代表着什么”,而是把这样的文化细节融入到了日常对话场景当中。

文化图式剖析:该题属于语境推理型与文化负载词型的结合。其背后的文化图式是“餐桌礼仪禁忌图式”:在中国饮食文化中,筷子不仅是进食工具,更承载着社会身份与礼仪规范的象征意义。敲击碗碟与“乞丐乞讨”行为在文化记忆中的关联,构成了一个完整的“禁忌-身份-礼貌”图式网络。从 Byram 模型看,此题测量知识维度(禁忌知识)与态度维度(savoir être)——学习者是否具备对中国礼仪规范的尊重与敏感性。

教学启示:此类题目提示文化教学应关注“隐性文化规则”,即那些母语者视为理所当然但外国学习者难以察觉的文化规范。教师需在语言教学中适时渗透此类文化细节,而非仅依赖专门的文化课程。

3.3. HSK 中级: 语用规则的得体性判断——“吃了吗”

试题来源: HSK (四级)阅读真题,2017 年 7 月卷,第 42 题。“吃了吗”作为问候语,表面询问饮食情况,实则并不期待对方详细汇报餐食内容,而是表达关心与友好。这一表达源于中国传统社会物质匮乏时期对温饱问题的关注,逐渐演变为熟人社会的惯用问候。学习者若按字面理解,可能误以为中国人热衷于讨论饮食细节,甚至产生“隐私侵犯”的误解。

文化图式剖析:该题属于语用得体的性考查,其背后是“形式-功能非对称图式”。汉语交际中大量存在“字面意义不等于语用功能”的现象,“吃了吗”只是其中典型一例。该图式涉及中国历史记忆(物质匮乏)、社会关系(熟人问候)与语用惯例(寒暄功能)三个层次。从 Byram 模型看,此题主要测量技能维度(savoir faire)——学习者能否在特定交际情境中选择符合中国文化规范的回应方式。

教学启示：语用教学应注重“形式背后的文化逻辑”，帮助学习者区分字面意义与语用功能。同时，教师需注意文化观念的历史演变，如当代年轻人可能更多使用“你好”“最近怎么样”等问候方式，避免给学习者造成文化刻板印象。

3.4. HSK 高级：网络文化与言外之意——“躺平”

试题来源：HSK(六级)阅读真题(模拟题)，2022 年 12 月卷。材料为一篇关于当代中国青年社会心态的短文，其中提到“面对激烈的竞争，一些年轻人选择躺平，但这并不意味着他们放弃努力，而是一种暂时的调整与反思”。题干要求考生推断作者对“躺平”现象的态度。

文化图式剖析：该题属于言外之意型与跨文化对比型的结合。“躺平”作为 2021 年前后兴起的网络流行语，承载着丰富的当代社会文化信息：它不仅是一种个人生活态度，更反映了青年群体面对社会压力时的集体心理策略。理解该词需要激活“网络亚文化图式”与“社会竞争-个体应对图式”。从 Byram 模型看，此题对四个维度均有要求：知识维度(了解网络流行语及其社会背景)、技能维度(推断作者态度)、态度维度(理解而非批判性地否定该现象)、批判性文化意识维度(评价“躺平”作为社会现象的多重意义)。

教学启示：高级阶段文化教学应引入“当代文化动态”视角，关注网络文化、社会热点与价值观念的演变。测试命题在引入此类话题时，应避免价值预设，提供多元视角，确保考查的是语言能力而非意识形态认同。

3.5. AP 汉语与文化考试：称赞回应的跨文化语用——“哪儿能跟你比啊”

试题来源：AP 汉语与文化考试样题，2019 年版，第 28 题。在对话里，甲向乙表达祝贺，恭喜乙考上了南京大学，乙则回应说：“我可没法跟你比，你都考上北大了，我真的特别羡慕你。”依照西方文化的规范，当受到称赞之后，人们应当欣然接受并且表达感谢，然而按照中国文化的规范，如果直接接受称赞，有可能会被看作是不谦虚，恰当的做法是贬低自己、抬高对方。

文化图式剖析：该题考查的是中国文化里面对称赞时的“贬己尊人”回应策略，属于言外之意型与语用得体的结合。其背后的文化图式是“谦逊礼貌图式”：在中国传统文化中，谦虚被视为美德，直接接受称赞可能被视为骄傲自满。该图式与西方“自我肯定”文化形成鲜明对比。从 Byram 模型看，此题主要测量跨文化对比技能(savoir comprendre)——学习者能否识别中外称赞回应策略的差异，并做出恰当的文化调整。

教学启示：跨文化语用教学应引入“文化对比”视角，帮助学习者意识到母语语用规则与目的语语用规则之间的差异。同时，需避免将文化规范绝对化，应指出当代中国青年在跨文化交际中也可能采用“感谢 + 自谦”的混合策略。

3.6. HSK 中级：文化图式的推理判断——《周公解梦》

试题来源：HSK(四级)阅读真题，2018 年 7 月卷，第 35 题。材料介绍“很久以前，中国人认为做梦是上天要告诉他们将来会发生的一些事情，因此他们试着对各种梦做出解释……后来，他们把这些解释写成了一本书，叫做《周公解梦》。”随后设置理解题，要求学习者根据文本推断《周公解梦》的性质。

文化图式剖析：此题将中国古代文化典籍嵌入阅读材料，考查学习者对中国传统民俗文化的认知。其背后的文化图式是“天人感应图式”——古代中国人认为梦境与现实之间存在神秘联系，这种思维方式与西方心理学的“潜意识投射”观念形成对比。从 Byram 模型看，此题测量知识维度(传统典籍知识)与技能维度(文本推断能力)。

教学启示：文化教学应兼顾“传统文化”与“当代文化”的平衡，帮助学习者建立完整的文化认知

框架。测试命题在引入文化典籍时, 应提供足够的语境线索, 避免因文化知识鸿沟影响语言能力的准确测量。

4. 隐性考查的效度问题与反拨效应

4.1. 效度验证的挑战

隐性考查方式在提高测试真实性上有一定作用, 不过也给效度验证带来了挑战。首先是构念界定困难: 隐性文化知识作为一种理论构念, 其操作化定义存在争议, 很难分辨考生到底是因为语言能力不足答错, 还是因为文化知识缺失答错。本文提出的 Byram 四维度构念框架与 MIRT 测量模型, 正是为回应这一挑战而设计的。通过将隐性文化知识分解为知识、技能、态度、意识四个维度, 并在题项层面分析其在语言能力与隐性文化知识两个潜变量上的载荷, 可以更精确地定位考生的能力缺陷。

其次是评分标准模糊。隐性考查常常涉及开放性的语用判断, 评分员要拥有深厚的跨文化交际知识, 才能保证评分有可信度[11]。此外, 文化公平性问题不容忽视[12]。来自汉字文化圈(如日本、韩国)的学习者可能在文化负载词理解上具有优势, 而来自非汉字文化圈的学习者则可能面临双重障碍: 语言解码与文化解码。这种“文化偏见”可能影响测试的公平性。MG-CFA 可用于检验不同母语背景考生在测量模型上的不变性, 若发现题项存在差异功能(DIF), 则应对题项进行修订或标注。

4.2. 对文化教学的正向反拨

尽管存在挑战, 隐性考查方式对文化教学具有显著的正向反拨效应(positive washback)。与显性考查导致的“应试文化记忆”不同, 隐性考查促使教学关注文化与语言技能的融合, 推动“文化教学语言化”与“语言教学文化化”[13]。当学习者意识到文化背景知识直接影响语言理解时, 其学习动机将从“通过考试”转向“有效交际”, 进而提升跨文化交际能力。

研究表明, HSK 四级考生普遍认可试题中文化因素对学习的帮助作用, 认为文化内容的融入增强了试题的趣味性与真实性。这一发现支持了隐性考查方式在促进教学方面的积极价值。

5. 优化建议

针对以上现象, 提出以下优化建议。

5.1. 丰富文化因素的分布与层次

当前 HSK 四级中文化类话题占比偏低, 且集中于语言文字与地理名胜, 风俗传统与国情民生类话题不足。建议依据《国际中文教育用中国文化和国情教学参考框架》, 增加传统文化、社会生活文化及当代国情类内容, 构建“由近及远、由浅入深”的文化考查体系[14]。具体而言, 初级阶段应侧重日常生活文化(如数字、颜色、饮食禁忌), 中级阶段侧重社会交往文化(如称呼、问候、称赞回应), 高级阶段侧重当代社会文化(如网络流行语、社会心态、价值观念变迁)。

5.2. 完善隐性考查的评分标准: 三维评分量表草案

对于语用得体的性、言外之意等主观性较强的题目, 应制定详细的评分细则与评分员培训方案。本文提出一个包含“语言准确性”“语用得体的性”“文化适切性”三个维度的评分量表草案(见表 1), 以确保评分的一致性与公平性[15]。

该量表适用于口语面试、书面回应等开放性任务。评分员需接受跨文化交际培训, 并通过一致性检验(如 Kappa 系数 ≥ 0.80)后方可独立评分。对于选择题等客观题型, 可通过项目功能差异(DIF)分析检验是否存在文化群体偏差。

Table 1. Implicit cultural assessment rubric for open-ended tasks
表 1. 隐性文化考查开放性任务评分量表草案

维度	权重	表现描述(高分)	表现描述(低分)
语言准确性	40%	语法正确, 词汇选择恰当, 语音标准(口语任务), 无影响理解的错误。	语法错误较多, 词汇使用不当, 影响基本理解。
语用得体性	35%	能够根据交际情境(正式/非正式、熟人/陌生人) 选择恰当的语体与表达策略。	语体选择不当, 如正式场合使用 过于随意的表达, 或反之。
文化适切性	25%	回应方式符合中国文化规范(如谦逊、间接、 面子维护), 能体现对文化差异的敏感性。	回应方式带有明显的母语文化负迁移, 如直接接受称赞、过度自我肯定。

5.3. 增强测试任务的真实性：真实语料库的开发与利用

选取更多真实语料, 提高录音自然程度, 设计贴近真实交际情境的任务, 使隐性考查更准确地反映学习者在实际跨文化交际中的能力表现[16]。具体建议如下:

第一, 社交媒体语料库。收集微博、微信、B 站弹幕等真实文本, 筛选包含文化负载词、网络流行语、言外之意的语料, 开发阅读理解与语用判断任务。例如, 利用一条关于“双十一”购物节的微博, 考查学习者对“剁手”“种草”等网络用语的理解。

第二, 影视剧语料库。选取当代反映中国社会生活的电视剧(如《三十而已》《人世间》)片段, 截取包含文化冲突、语用策略的对话, 设计听力理解与跨文化对比任务。影视剧语料具有情境真实性、口语自然性与文化丰富性三重优势。

第三, 生活纪录片语料库。利用反映中国各地风俗、饮食习惯、家庭关系的纪录片(如《舌尖上的中国》《人生第一次》), 开发视听综合任务, 要求学习者在理解语言的同时推断文化规范。

5.4. 建立文化能力描述的等级标准

参照《国际中文教育中文水平等级标准》对语言技能的等级划分, 本文尝试为初、中、高三个等级提出隐性文化知识的能力表现描述(见表 2), 为命题设计提供科学依据[17]。

Table 2. Performance descriptors of implicit cultural knowledge by proficiency level (beginner, intermediate and advanced)
表 2. 隐性文化知识能力表现描述(初、中、高级)

等级	知识维度表现描述	技能/态度维度表现描述
初级(HSK1~2)	了解最常见的中国文化符号 (如春节、龙、红色、数字 8)及其基本含义; 知道简单的生活禁忌(如筷子不插饭)。	能在简单日常情境中识别文化线索 (如收到红包时知道说“谢谢”); 对明显的文化差异表现出包容态度。
中级(HSK3~4)	掌握社会交往中的核心文化规范 (如称呼、问候、称赞回应、餐桌礼仪); 理解常见文化负载词在特定语境中的隐含意义(如“面子”“关系”)。	能根据交际对象的身份、场合选择恰当的 表达方式; 能识别言外之意并做出符合文化规范 的反应; 能进行简单的中外文化对比。
高级(HSK5~6)	理解当代中国社会文化现象(如网络流行语、 社会心态、价值观念变迁); 掌握中国传统文化典 籍的核心概念(如“和为贵”“中庸”); 了解地区文化差异。	能在复杂交际情境中灵活运用文化策略 (如委婉拒绝、间接批评); 能对文化现象进行批判 性反思; 能在跨文化冲突中充当“文化中介者” (cultural mediator)。

该等级标准应与语言技能等级标准并行设计, 在《等级标准》修订时纳入“文化能力”维度。命题者在设计试题时, 应参照该标准确定各等级文化内容的范围与深度, 避免超纲或重复考查。

6. 结语

汉语二语测试当中对于中文背景知识的隐性考查,属于语言测试与文化教学交叉领域里的一个重要话题。本文借助理论梳理、案例分析,揭示出隐性考查的五种主要形式,并基于 Byram 跨文化交际能力模型将隐性文化知识重构为知识、技能、态度与批判性文化意识四个维度。案例分析覆盖了 HSK 初、中、高三个水平,涉及风俗禁忌、语用规则、言外之意、网络文化及当代社会价值观等多样化主题,深入剖析了其背后的文化图式,并提供了试题来源信息以增强可验证性。

在优化路径方面,本文提出了包含“语言准确性”“语用得体的性”“文化适切性”三维度的评分量表草案,建议利用社交媒体、影视剧等真实语料库开发测试任务,并为初、中、高三个等级提出了具体的隐性文化能力表现描述。这些建议旨在将宏观理论转化为微观操作,提升汉语二语测试文化考查的科学性与公平性。

未来的研究能够进一步运用实证方法,借助眼动追踪、有声思维等实验手段,去探究学习者在隐性文化考查里的认知加工过程。要强化针对不同母语背景学习者文化知识差异的对比研究,以此为建立公平且有效的文化测试模式提供数据方面的支持。在“讲好中国故事”这样的时代背景之下,优化汉语二语测试当中的文化考查方式,不但有助于提高测试质量,而且对于推动国际中文教育高质量发展有着深远的意义。

参考文献

- [1] 赵金铭. 对外汉语教学概论[M]. 北京: 商务印书馆, 2004.
- [2] Purpura, J.E. (2004) Assessing Grammar. Cambridge University Press. <https://doi.org/10.1017/cbo9780511733086>
- [3] Bachman, L.F. and Palmer, A.S. (2010) Language Assessment in Practice. Oxford University Press.
- [4] Kunnan, A.J. (2004) Test Fairness. In: Milanovic, M. and Weir, C., Eds., *European Language Testing in a Global Context*, Cambridge University Press, 27-48.
- [5] Bialystok, E. (1982) On the Relationship between Knowing and Using Linguistic Forms. *Applied Linguistics*, **3**, 181-206. <https://doi.org/10.1093/applin/3.3.181>
- [6] Ellis, R. (2005) Measuring Implicit and Explicit Knowledge of a Second Language: A Psychometric Study. *Studies in Second Language Acquisition*, **27**, 141-172. <https://doi.org/10.1017/s0272263105050096>
- [7] 李泉. 对外汉语教材的“国别化”问题探讨[J]. 世界汉语教学, 2015, 29(4): 526-540.
- [8] 祖晓梅. 跨文化交际[M]. 北京: 外语教学与研究出版社, 2015.
- [9] Byram, M. (1997) Teaching and Assessing Intercultural Communicative Competence. *Multilingual Matters*.
- [10] 黄理兵, 郭树军. HSK 阅读理解试题的语料和命题[J]. 世界汉语教学, 2008(2): 135-144.
- [11] Xi, X. (2010) How Do We Go about Investigating Test Fairness? *Language Testing*, **27**, 147-170. <https://doi.org/10.1177/0265532209349465>
- [12] Shohamy, E. (2001) The Power of Tests: A Critical Perspective on the Uses of Language Tests. Longman.
- [13] 刘珣. 对外汉语教育学引论[M]. 北京: 北京语言大学出版社, 2000.
- [14] 教育部中外语言交流合作中心. 国际中文教育用中国文化和国情教学参考框架[S]. 北京: 北京语言大学出版社, 2022.
- [15] Canale, M. (1980) Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, **1**, 1-47. <https://doi.org/10.1093/applin/1.1.1>
- [16] Hymes, D. (1972) On Communicative Competence. In: Pride, J. and Holmes, J., Eds., *Sociolinguistics*, Penguin, 269-293.
- [17] 李晓琪. 对外汉语文化教学研究[M]. 北京: 商务印书馆, 2006.