

AI 测评体系的标准化建构下国际中文教育数据驱动 - 标准反哺机制

王丽璇

连云港师范学院, 江苏 连云港 222000

DOI: 10.61369/ETR.2026030021

摘要 : 人工智能生成内容 (AIGC) 的兴起推动了新一轮智能技术革命, 也正在推进教育体系的改革与发展。国际中文教育正经历深刻的数字化变革。伴随着国际中文教育数字化、智能化进程的不断加快, 教学测评体系的建设需要有更高的质量要求与标准化需求。面对当前教学与评价系统之间脱节、教学反馈滞后等问题, 本文以 AI 测评体系的标准化建构为出发点, 提出构建数据驱动的测教学体系, 探究测评结果如何反哺教学内容与评价标准, 形成标准更新的机制。文章梳理了 AI 测评在语音识别、语义理解、写作反馈等方面的关键技术路径, 指出现有国际中文测评体系在标准一致性、数据利用和反馈机制上存在的不足。为此本文结合教学实际与生成式人工智能的能力, 提出由标准导向、智能评分、数据采集、反馈生成与资源重构组成的 AI 测评体系建构模型。本文强调建立标准、测评、反馈、再标准化的反哺机制, 有助于提升测评系统的可持续性, 也为国际中文教育智能化转型提供了可行路径。

关键词 : AI 测评体系; 标准化建构; 国际中文教育; 数据驱动; 标准反哺机制

The Data-Driven - Standard Feedback Mechanism for International Chinese Language Education under the Standardized Construction of the Ai Assessment System

Wang Lixuan

Lianyungang Normal University, Lianyungang, Jiangsu 222000

Abstract : The rise of AI-generated Content (AIGC) has driven a new round of intelligent technology revolution and is also promoting the reform and development of the education system. International Chinese language education is undergoing profound digital transformation. With the continuous acceleration of the digitalization and intelligence of international Chinese language education, the construction of the teaching assessment system needs to have higher quality requirements and standardization demands. Facing the current problems such as the disconnection between teaching and evaluation systems and the lag in teaching feedback, this paper takes the standardized construction of the AI assessment system as the starting point, proposes to build a data-driven assessment teaching system, explores how assessment results can feed back into teaching content and evaluation standards, and forms a mechanism for standard updates. This article sorts out the key technical paths of AI assessment in aspects such as speech recognition, semantic understanding, and writing feedback, and points out the deficiencies of the existing international Chinese assessment system in terms of standard consistency, data utilization, and feedback mechanisms. For this reason, this paper, in combination with teaching practice and the capabilities of generative artificial intelligence, proposes an AI evaluation system construction model composed of standard orientation, intelligent scoring, data collection, feedback generation and resource reconstruction. This article emphasizes the establishment of a feedback mechanism involving standards, assessment, feedback, and re-standardization, which is conducive to enhancing the sustainability of the assessment system and also provides a feasible path for the intelligent transformation of international Chinese language education.

Keywords : AI assessment system; standardized construction; international Chinese education; data-driven; standard feedback mechanism

引言

在人工智能技术呈指数级发展的当代语境下全球教育生态正经历着从范式革新到体系重构的深刻变革^[1, 2]。国际中文教育作为承载中华文明传播使命、促进跨文化交流的重要载体，智能化转型已成为顺应时代发展的必然选择^[3]。截至 2024 年全球已有 180 多个国家和地区开展中文教学，学习者规模突破 2 亿人，庞大的教育需求与差异化的学习特征对教学质量保障体系提出了更高要求。作为教育质量监测与改进的核心环节教学测评体系在国际中文教育数字化进程中扮演着关键角色^[4, 5]。

传统测评方式基于人工评分与经验判断存在评分主观性强、反馈周期冗长、难以适配大规模个性化教学需求等固有缺陷。以 HSK（汉语水平考试）为例，线下笔试模式虽具有权威性但在即时反馈与动态评估方面存在明显不足，无法满足数字化时代学习者对学习效果的即时感知需求^[6]。AI 测评体系凭借算法的高效性、数据处理的精准性以及评价的客观性为破解上述难题提供了创新解决方案^[7]。然而当前 AI 测评体系在国际中文教育领域的应用仍面临诸多瓶颈，如测评标准缺乏统一规范、数据价值未充分挖掘、反馈机制难以满足个性化学习需求等，这些问题严重制约着国际中文教育智能化转型的高质量发展。因此构建标准化的 AI 测评体系并建立与之适配的数据驱动标准反哺机制已成为推动国际中文教育现代化进程的重要课题。

一、国际中文教育 AI 测评体系的技术演进与现实困境剖析

（一）AI 测评核心技术的多维突破与应用路径

国际中文教育 AI 测评体系的技术架构依托于人工智能领域的前沿技术成果，形成了语音识别、语义理解、写作反馈三大核心技术模块协同运作的格局^[8]。在语音识别层面基于深度学习的端到端模型，如 Transformer 架构与 Wav2Vec 2.0，通过对海量中文语音数据的无监督预训练实现了对学习者发音特征的精准捕捉与分析。以科大讯飞的中文语音评测系统为例，采用动态时间规整（DTW）算法能够将学习者发音与标准语音库进行毫秒级的时序对齐，从而定位发音部位、声调调值等维度的错误并生成可视化的发音改进方案。

语义理解技术借助预训练语言模型（PLM）的强大表征能力实现了对中文语义的深度解析^[9]。BERT、GPT 系列模型通过双向 Transformer 编码器能够学习中文文本的上下文语义信息，从而对学习者的语言表达进行逻辑连贯性、语义准确性与文化适应性的综合评估。在写作反馈领域基于自然语言处理（NLP）的语法纠错系统与文本生成模型相结合能够识别拼写错误、语法瑕疵还能从篇章结构、修辞运用等维度提出修改建议，助力学习者提升中文书面表达能力^[10]。

（二）现有测评体系的结构性矛盾与发展瓶颈

尽管 AI 技术在国际中文测评中已取得一定应用成果但现有测评体系仍存在显著的结构性矛盾。首先标准一致性问题突出。不同机构开发的 AI 测评工具在评价维度、评分标准与权重设定上差异显著导致测评结果缺乏横向可比性。如部分测评工具侧重语言知识的考查，而另一部分则更关注语言运用能力，这种标准的碎片化严重影响了对学习者语言水平的精准定位。其次数据利用效率亟待提升。当前多数 AI 测评系统仅将数据用于评分计算，未能充分挖掘数据背后的教学价值。大量测评数据蕴含着学习者的认知模式、学习路径、常见错误类型等关键信息，若能通过大数据分析机器学习算法进行深度挖掘将为教学策略优化、课程内

容设计提供重要依据。然而由于数据存储分散、分析技术滞后等原因这些宝贵的数据资源大多处于闲置状态。最后反馈机制的局限性制约着学习效果的提升。现有 AI 测评工具的反馈形式多以分数、等级或简单的文字提示为主，缺乏对错误成因的深度剖析与个性化学习建议。这种“一刀切”的反馈模式无法满足不同水平、不同学习风格学习者的需求难以实现测评对教学的有效促进作用。

二、AI 测评体系标准化建构的理论框架与实践模型

（一）标准导向：构建多维融合的 AI 测评标准体系

AI 测评体系的标准化建构需以国际中文教育通用标准为基石，结合 AI 技术特性构建涵盖语言知识、语言技能、文化素养、学习策略等多维度的测评标准体系。在语言知识维度，细化语音、词汇、语法等子指标并明确各指标的能力层级描述；在语言技能维度，区分听、说、读、写四种技能的评价标准，建立从基础应用到高级交际的能力发展阶梯；在文化素养维度融入中国文化知识、跨文化交际能力等评价要素体现国际中文教育的文化育人功能。同时建立标准动态更新机制至关重要。通过定期收集教学实践数据、跟踪技术发展趋势、参考国际语言测评标准的更新成果对测评标准进行适应性调整。随着人工智能在跨文化交际模拟中的应用可将虚拟场景下的交际能力纳入测评标准，使标准体系始终保持先进性与适用性。

（二）智能评分：打造多模态融合的精准评价系统

基于深度学习的智能评分系统是 AI 测评体系的核心模块。通过构建多层神经网络模型对标注的测评数据进行训练，使系统能够自动识别语法错误、词汇误用、逻辑漏洞等语言问题并依据错误的严重程度进行量化评分。引入多模态信息融合技术，将语音、文本、表情、手势等多维度数据纳入评分依据实现对学习者语言表现的全面评估。在口语测评中结合语音语调、面部表情与肢体语言，更准确地判断学习者的语言交际能力。为提升评分的准确性与稳定性，采用集成学习方法，融合多个评分模型的结

果,降低单一模型的误差。同时建立评分模型的持续优化机制,通过定期更新训练数据、调整模型参数使评分系统适应不同学习群体与教学场景的需求。

(三) 数据采集: 搭建全流程覆盖的数据治理平台

完善的数据采集平台是实现数据驱动教学的基础。该平台需覆盖测评前的学习者特征数据、测评中的过程性数据以及测评后的结果数据形成全流程的数据采集体系。该平台需构建涵盖学习者基本特征、测评过程行为、评估结果反馈的多维数据采集体系。采用分布式存储技术与云计算架构确保海量数据的高效存储与快速处理。建立严格的数据安全管理体系,遵循 GDPR、《个人信息保护法》等相关法规,对学习者数据进行加密存储与脱敏处理保障数据隐私安全。同时构建数据质量监控机制,通过数据清洗、校验、标注等环节确保数据的准确性与完整性。

(四) 反馈生成: 实现个性化定制的智能反馈系统

利用自然语言生成(NLG)技术构建个性化的反馈生成系统。该系统基于智能评分结果与数据采集分析,自动生成包含错误诊断、改进建议、学习资源推荐等内容的反馈报告。通过对学习者错误类型的聚类分析提供针对性的学习策略;根据学习者的学习进度与能力水平推荐适配的学习资源如微课视频、练习题集、拓展阅读材料等。为确保反馈的专业性与准确性建立教师审核机制,允许教师对自动生成的反馈内容进行调整与补充。同时引入情感计算技术使反馈语言更具亲和力与鼓励性,增强学习者的学习动力。此外支持反馈内容的多模态呈现如文字、语音、图表相结合提升反馈的可读性与易用性。

(五) 资源重构: 构建动态适配的教学资源生态

基于测评结果与反馈数据实现教学资源的智能重构与精准推送。通过分析学习者的知识薄弱点与学习需求,运用知识图谱技术将教学资源进行结构化拆解与关联,生成个性化的学习路径。

三、数据驱动标准反哺机制的运行逻辑与实施路径

(一) 数据挖掘与分析: 揭示教学规律的认知引擎

运用大数据分析机器学习算法对采集的测评数据进行深度挖掘,构建教学认知模型。运用大数据分析机器学习算法对采集的测评数据进行深度挖掘,构建教学认知模型,为教学改进提

供科学依据。建立教学质量评估模型,从教师教学行为、课程内容设计、学习支持服务等维度对教学质量进行综合评价。通过对教师授课视频的多模态分析评估教学方法的有效性;基于课程学习数据分析课程内容的难度适配性与知识覆盖度;依据学习者的反馈信息评价学习支持服务的满意度。这些分析结果为教学改进提供科学依据实现数据驱动的教学决策。

(二) 标准优化与更新: 保持体系活力的动态演进

依据数据挖掘与分析的结果建立 AI 测评标准的动态优化机制。对于学习者普遍存在的知识薄弱环节,提高相关测评指标的权重加大考查力度;对于教学实践中证明有效的教学方法与内容转化为测评标准引导教学方向。跟踪国际中文教育发展动态与人工智能技术进步趋势及时引入新的测评理念与方法。关注国际语言测评领域的最新研究成果如欧洲语言共同参考框架(CEFR)的更新内容,借鉴先进的评价理念;结合人工智能技术的新突破如生成式 AI 在语言测评中的应用,创新测评形式与内容确保标准体系始终处于国际领先水平。

(三) 反馈循环机制: 促进持续改进的闭环系统

构建“标准-测评-反馈-再标准化”的闭环反馈机制实现测评体系的持续优化。在每次测评结束后将测评结果与反馈数据进行深入分析,找出标准体系中存在的问题与不足,据此对标准进行修订与完善。优化后的标准再次应用于后续测评形成螺旋上升的改进循环。建立教师、学习者与系统开发者多元参与的协同反馈机制,推动标准反哺机制的有效运行。通过定期召开多方研讨会、开展问卷调查、建立在线反馈平台等方式广泛收集各方意见,推动标准反哺机制的有效运行。

四、结论

本研究围绕 AI 测评体系的标准化建构与国际中文教育数据驱动标准反哺机制取得了以下成果。一是构建了包含标准导向、智能评分、数据采集、反馈生成与资源重构的 AI 测评体系标准化模型,为国际中文教育测评的规范化发展提供了理论框架;二是建立了数据驱动的标准反哺机制实现了测评标准的动态优化与教学质量的持续提升。

参考文献

- [1] 周洪宇,常顺利.生成式人工智能嵌入高等教育的未来图景,潜在风险及其治理[J].现代教育管理,2023,11:1-12.
- [2] 吴砥,李环,陈旭.人工智能通用大模型教育应用影响探析[J].Open Education Research,2023.
- [3] 刁江智.人工智能技术在国际中文课堂上的应用分析[J].Advances in Education,2025,15:1.
- [4] 董良峰,刘媛,朱思达.国际中文教育治理现代化的现实挑战与推进路径[J].Journal of Soochow University Educational Science Edition,2024,12(1).
- [5] 胡姜,池利李,艺杰王, et al.近五年(2019-2024)人工智能背景下国际中文教育研究动态与发展趋势[J].泰国格乐大学国际中文教学期刊,2024,4(2):126-142.
- [6] 王凌王.大学预科班学生的对外汉语教学模式与效果评估[J].现代教育前沿,2025,6(3):100-102.
- [7] 余娅.人工智能赋能高等教育评价[J].Creative Education Studies,2024,12:346.
- [8] 欧志刚,刘玉屏,覃可, et al.人工智能多模态教学资源的生成与评价——基于 AIGC 在国际中文教育的应用[J].Modern Educational Technology,2024,34(9).
- [9] 韩毅,乔林波,李东升, et al.知识增强型预训练语言模型综述[J].Journal of Frontiers of Computer Science & Technology,2022,16(7).
- [10] 张明,卢庆华,黄元忠, et al.自然语言语法纠错的最新进展和挑战[J].Journal of Computer Engineering & Applications,2022,58(6).