

人工智能时代英语教师测评素养的挑战与策略

李欣雨, 杨香玲

天水师范大学外国语学院, 甘肃 天水

收稿日期: 2025年11月25日; 录用日期: 2025年12月24日; 发布日期: 2025年12月31日

摘要

人工智能(AI)正深刻重塑英语教学测评的形态, 同时也对英语教师测评素养提出了全新要求。本文系统阐释了AI时代英语教师测评素养的新内涵, 主要包括AI技术应用能力、数据驱动决策能力和智能伦理审视能力, 并深入剖析了其在技术适应、数据素养、理念更新与伦理风险等方面面临的挑战。为此, 本文提出构建融合型分层培训体系、推动测评理念深层转型、深化人机协同测评实践、完善制度保障与激励机制和强化数据伦理与人文关怀五方面策略, 旨在为构建“AI增强型测评素养”框架提供理论与实践参考, 从而推动英语教育测评体系的发展, 最终实现技术赋能下“以评促学”的高质量英语教学。

关键词

人工智能, 英语教师, 测评素养, 挑战与策略

Challenges and Strategies for English Teachers' Assessment Literacy in the Artificial Intelligence Era

Xinyu Li, Xiangling Yang

School of Foreign Languages, Tianshui Normal University, Tianshui Gansu

Received: November 25, 2025; accepted: December 24, 2025; published: December 31, 2025

Abstract

Artificial Intelligence (AI) is profoundly reshaping the landscape of English language assessment, while simultaneously raising new demands for teachers' assessment literacy. This study systematically explores the new connotations of English teachers' assessment literacy in the AI era, encompassing AI tool application capability, data-driven decision-making ability, and intelligent ethical scrutiny capacity. It also further identifies the challenges faced in areas such as technological adap-

tation, data literacy, conceptual updating, and ethical risks. Building on this, five strategic directions are proposed: establishing a tiered training system, promoting the deep transformation of assessment concepts, advancing human-AI collaborative evaluation, refining institutional support mechanisms, and strengthening ethical awareness with humanistic care in data applications. The aim is to provide theoretical and practical references for constructing an “AI-enhanced assessment literacy” framework, thereby advancing the development of the English education assessment system and ultimately realizing high-quality, technology-enabled English teaching grounded in the principle of “assessment for learning”.

Keywords

Artificial Intelligence, English Teachers, Assessment Literacy, Challenges and Strategies

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

外语教师测评素养是教师专业能力的重要组成部分,是实现“以评促学”和“以评促教”的重要基础[1]。自 Stiggins [2]提出这一概念以来,其核心内涵不断丰富与发展。Brindley [3]和 Davies [4]分别提出五要素和三要素模型,均指出测评素养包含知识、技能和原则三个基本维度。Fulcher [5]进一步提出“实践-原则-语境”三维模型,强调教师除掌握测试设计与开发的知识技能外,还应理解测评的社会政治语境并遵循伦理规范。当前,AI 技术正驱动教育测评范式发生深刻变革。2025 年教育部办公厅印发的《关于组织实施数字化赋能教师发展行动的通知》明确将教师数字素养建设置于战略高度,要求实现“经过 3 至 5 年努力,教师数字素养全面提升,熟练应用数字化手段开展教育教学成为新常态”[6]。在此背景下,传统依赖经验与有限数据的测评模式已难以满足智能测评、实时反馈与个性化学习分析的需求。因而,构建契合智能教育生态的教师测评素养新体系势在必行。

然而,现有研究多集中于理论阐述[7][8],针对智能环境下教师测评素养转型的系统性与实证性研究仍相对薄弱。国内少数学者虽探讨了大数据时代英语教师测评素养面临的困境与对策[9][10],但尚缺乏以 AI 为核心的测评素养框架研究。本文拟构建涵盖“技术应用-数据决策-伦理审视”三要素的“AI 增强型测评素养”框架,系统分析当前教师在智能测评实践中遇到的实际困难,进而提出整合性的发展策略,以期为推进英语教师测评素养的转型提供理论支撑与实践指引。

2. AI 时代教师测评素养的核心内涵与范式转变

2.1. AI 时代下英语教师测评素养内涵拓展

为适应智能化教学大环境,本研究主张从传统的“知识-技能-原则”框架,过渡到以“技术应用、数据决策、伦理审视”为核心的“AI 增强型测评素养”框架。

2.1.1. AI 技术应用能力

AI 技术应用能力要求教师不仅能够理解智能测评工具的基本原理,熟练掌握其操作方法,还需能够根据具体的测评目标选择最恰当的工具类型。他们应善于依托 AI 工具开展多模态数据的收集与深入分析,并可以借助相关技术生成个性化的习题库、设计自适应的测试路径等,最终让智能技术充分融入形

成性评价的全流程。该能力的核心在于教师与 AI 系统有效协作, 实现测评过程专业化与智能化的融合。例如在口语测评中, 教师可以借助 AI 工具进行发音、流利度等基础指标的测评工作, 把主要精力放在对学生交际策略、文化得体性等深层能力的评价上, 在提升评价效率的同时兼顾评价内容的人文底蕴。

2.1.2. 数据驱动决策能力

数据驱动决策能力要求教师能高效获取、分析并运用学生在写作、口语等学习过程中产生的多维度数据[11], 并凭借这些数据做出科学的教学判断, 推动测评范式从依赖主观经验向基于语言证据的科学决策转变。对写作修改轨迹、口语流利度等过程性数据的分析可帮助教师诊断学生在学习过程中采用的认知策略及面临的困难[12], 结合学生实际情况设计针对性的教学干预活动, 最终实现数据驱动的精准教学模式。

2.1.3. 智能伦理审视意识

智能伦理审视意识是教师在运用 AI 工具开展测评工作时, 能够主动察觉潜在伦理问题并自觉开展批判性反思的一种态度。这要求教师不能局限于技术操作, 而应当建立起一套系统的伦理认知框架, 识别并预防数据隐私泄露、算法偏见以及文化歧视等潜在风险, 在技术应用中始终保持教育者的主体地位与专业判断。教师在利用技术提升测评效率的同时, 要始终关注测评活动的公平性与人文性, 确保 AI 技术应用始终服务于学生的全面发展。

2.2. AI 时代英语教师测评范式转变

人工智能正在驱动外语测评模式展开系统性变革, 这一变化主要体现在四个方面: 提升语言测评效率与客观性; 促进形成性评价个性化; 推动测评依据由经验向数据转变; 推进真实情境测评实践。

2.2.1. 提升语言测评效率与客观性

在传统测评中, 评分者的主观判断以及繁重的工作量是限制测评信度和效度的关键因素。以 ChatGPT 为代表的大语言模型, 可有效地改进传统语言考试开发时效率低下、成本较高等弊端提供更为全面且灵活的语言测评方案[13]。AI 依据预设的评分算法严格执行评分标准, 其评估过程可摆脱人为因素的干扰, 持续输出稳定的评分结果, 保证测评工作较高的客观性与一致性[14], 显著提高大规模高风险考试的信度与公平性。比如, 多邻国测试这类自适应考试可依据考生前一题的作答状况, 实时调整后续题目的难度, 凭借更少的题目数量以及更短的测试时间, 准确评估考生的实际语言能力。同时, AI 可承担客观题评分、作文语法问题检查等一些可量化的评分任务, 让教师从大量的重复性工作中解脱出来, 把更多精力投入到评估学生作答中那些 AI 难以涉及的思维深度、逻辑严密性以及表达创造性等高层次能力上, 促使教师角色从“测评执行者”向“学习引导者”转变。

2.2.2. 促进形成性评价个性化

传统课堂的形成性评价受制于时效与规模, 难以提供及时、个性化的反馈。例如, 在大班制教学中, 常见的口语评价可能仅限于课堂提问、背诵课文等, 难以客观、量化地评估每个学生的发音、语调、流利度等真实口语水平。AI 通过充当学习过程的“显微镜”, 极大地强化了形成性评价[15]。在学情诊断上, AI 能通过全程数据采集构建个体知识图谱, 可视化地展示学生的知识掌握情况与联系。在实时反馈上, 智能语音测评系统可即时识别发音偏差并给予纠音建议; 口语实时反馈系统能基于持续数据动态优化学习路径, 形成一个数据驱动的学习闭环。这些应用共同使“测评即学习”的理念得以真正实现。

2.2.3. 推动测评依据由经验向数据转变

人工智能技术正促使测评依据由“经验驱动”向“数据驱动”转变, 大数据分析、学习行为追踪以及

知识图谱建构等 AI 技术, 可全面且多维度地采集并分析学生的学习过程数据。AI 可持续记录学生在英语学习中的口语练习频率、写作常见错误、文本阅读速度等行为数据, 并根据这些数据生成直观的学习路径轨迹图。教师可依托这些数据准确把握班级整体学情与学生个体差异, 学生则能够通过反馈明晰自身学习状况, 让教学决策建立在坚实的证据基础上。

2.2.4. 推进真实情境测评实践

传统的语言测评设计往往脱离真实语境, 其内容大多是像单项填空、命题作文这类为考试服务的抽象题目, 这很容易致使学生出现“高分低能”。虚拟现实技术则为创设高度仿真的语言运用场景创造了条件, 教师可利用这类技术灵活构建丰富多样且贴近现实的语言运用场景, 与学生展开多轮次开放式对话。比如 ChatGPT 等大语言模型可模拟餐厅点餐等多种日常情景, 和学生进行真实有效的口语互动, 还可以提供充足的情绪价值以及个性化反馈, 缓解学生口语焦虑, 提高学生口语练习的兴趣与主动性[16]。这类基于真实任务的测评既能有效提升学生的跨文化交际能力、应变能力与语用能力, 也能显著优化测评的生态效度。

3. AI 时代英语教师测评素养面临的挑战

在人工智能重塑教育生态的浪潮中, 外语教师测评素养发展机遇与挑战并存。当前面临的诸多挑战涉及技术、数据素养、评价理念、工具限制以及情感和伦理问题等多个维度, 亟待得到系统性的审视与解决。

3.1. AI 技术壁垒

英语教师在应用 AI 测评工具时普遍面临技术理解与实践应用的双重困境。在技术理解方面, 不少教师对自然语言处理、机器学习等关键原理的了解程度有限, 无法深入掌握智能测评系统的运行原理以及评分逻辑, 当面对有争议的评分结果时, 难以有效地证实其合理性, 也难以给学生提供合理的解释。在写作自动评分中, 教师如果不了解系统是如何通过算法模型评估作文质量的, 便无法判断其评分结果的可靠性与局限性。

即便选定了合适的工具, 教师在技术应用中仍然存在系统集成障碍, 包括不同测评平台数据格式不兼容、API 接口调用繁杂等实际困难, 导致教师难以将智能测评工具顺利融入日常教学流程。比如, 有些工具支持自定义评分参数, 但教师由于缺少相关专业背景知识, 只能采用系统默认的设置, 造成测评结果和具体教学目标出现偏差。这种技术理解与实践应用的双重困境不仅会限制 AI 测评工具效能的发挥, 也阻碍了 AI 测评技术在推动教学评价朝着更科学、精准方向转变中的潜在价值的充分释放。因此, 提升教师的技术理解与系统集成能力, 是实现智能测评有效落地的前提。

3.2. 数据素养困境

当前英语教师数据素养存在系统性不足。在数据意识方面, 众多教师对于教育数据的认识相对浅显, 大多只是将其应用于如记录成绩等基础功能, 并没有充分认识到其在学习过程诊断以及发展趋势预测方面所具有的价值[17][18]。例如, 智能口语测评系统可提供语速变化曲线、停顿分布热力图等多维度分析指标, 然而在实际的教学当中, 教师大多时候仅仅关注总体分数, 缺少对这些蕴含着丰富教学价值的专业参数的解读能力, 无法精准识别出学生口语表达中的具体问题。

在数据伦理方面, 教师的认知水平和实践能力也存在明显欠缺。部分教师在没有获得明确授权的情况下便收集学生的语音数据, 缺乏规范的写作样本管理机制, 对于敏感信息加密存储的要求认识不足, 这些行为都存在侵犯学生隐私的风险[19]; 其次, 教师对算法可能存在的系统性偏见往往缺乏足够警觉。

例如, 以标准英语语料为基础训练的语音识别系统在处理带有非母语特征的口音时, 其识别准确率会出现明显下降; 写作自动评分系统在碰到有特定文化背景的独特表达方式时, 也有可能会出现理解偏差甚至误判。教师如果没有及时察觉这些潜在偏见, 就可能会在无形中加剧测评结果的不公平。

在数据知识和数据能力方面, 部分教师对数据采集的基本原理以及分析方法的认知有限, 很难运用专业工具开展深入的学情分析, 也无法有效地借助数据反馈来优化教学过程、提升教学效果。

3.3. 认知转型冲突

在人工智能赋能英语测评的进程中, 英语教师正面临深层次的认知转型困境。一部分教师仍然固守“测评即考试”的传统观念, 仅把 AI 工具当作自动阅卷的替代品, 未能充分意识到其在实现个性化形成性评价等方面的潜力, 难以真正实现“以评促学”的目标。另一部分教师则过度依赖技术和算法, 把 AI 系统输出的数据视为绝对权威, 主动放弃了在评价学生的思维深度、文化理解及情感态度等方面必不可少的专业判断权。比如在评价有关“春节”的口语表达时, 人工智能可检测出语言形式的正确性, 却难以识别学生表达中所包含的文化认同与个人情感, 而这恰恰是教师应当发挥专业价值的关键之处。这种工具理性和价值理性的不平衡会削弱教师文化解读与批判性思维培养等关键职能, 让教师从教育者变成数据的被动接收者; 一味分析和解读冷冰冰的数据还会强化教师“教书匠”的身份, 不利于培养和提升教师的创新意识以及教学反思能力, 最终可能导致教师沦为算法的附庸。

3.4. 工具与伦理限制

当前 AI 技术在英语测评中的应用仍存在明显边界。在工具层面, AI 虽在教学等多个领域中呈现出强大的能力, 但仍存在一些不可避免的局限性。其一, AI 算法与训练数据所固有的局限性在一定程度上会影响测评的稳定性。同一答案面对不同的提示词或模型版本往往出现分数不一致的情况; 部分 AI 系统还会给出看似明确, 实则完全错误的语法分析与反馈, 进而对教师教学和学生学造成误导。有案例表明, ChatGPT 等大语言模型有时不能准确识别学生写作中的实际错误, 甚至会针对正确表达给出不合理的修改建议。其二, 语言的意义与其文化背景与语境紧密相关, 然而 AI 系统难以识别和解读如幽默、反讽、隐喻、双关等需要依赖深厚文化背景和语境的语言现象, 造成评价结果出现偏差。其三, AI 虽然可以判断一个句子语法是否正确, 却缺乏必要的语用判断能力, 例如学生在口语测试中对教授使用过于随意的称呼时, AI 通常无法像人类教师一样敏锐地识别其中的语用失误。

在伦理层面, AI 赋能英语测评面临双重伦理困境, 既包括数据隐私、算法偏见等普遍性问题, 也涉及教育公平、人机互动伦理等教育特有难题。首先, 英语测评过程中会大量收集学生的口语录音、写作样本及在线学习行为轨迹等敏感数据, 这类数据在自动化处理中面临泄露与滥用的风险。其次, 算法偏见在语言测评中尤为突出。以国内广泛应用的“讯飞语言测评系统”为例, 该系统的语音识别模型主要基于标准美式英语以及普通话发音数据训练, 在识别中国学习者常见的“中式英语”发音特征时会存在偏差, 从而导致准确率显著下降。例如, 因为中国学习者易将 /r/ 与 /l/ 发音混淆, 系统可能将“rice”错误识别为“lice”, 导致评分偏低。因此, 教师不能仅停留在使用工具层面, 应主动对讯飞语音测评系统提供的数据进行审查, 有意识地抽样听取那些被系统评为低分的语音片段, 亲自复核其评分合理性, 尤其关注带有地方口音或特定语音习惯的学生样本, 并积累数据向技术供应商反馈, 从而推动算法优化。

AI 赋能的英语测评方式还可能会加剧“数字鸿沟”, 资源匮乏地区的学生因缺乏测评专用设备以及相应的数据素养, 在接触自适应测试等新型测评方法时处于明显劣势地位。教育的本质是充满温度的情感互动与精神培育, 而现阶段的 AI 反馈多为标准化与格式化的输出, 缺乏人类教师特有的情感认同与鼓励性表达, 容易让学生感到被物化, 挫伤学生的学习积极性。

4. AI 时代英语教师测评素养提升的策略

为推动英语教师测评素养的发展, 需建立系统化的发展路径: 构建融合型分层培训体系, 以加强教师能力建设; 教师应主动更新测评理念, 确立人工智能时代的科学评价观, 并积极探索人机协同的测评模式; 教育相关部门应通过制度支持与伦理规范为教师测评素养转型提供保障。只有通过教师自主发展与外部支持的有效结合, 方能实现英语教师测评素养的全面提升与现代化转型。

4.1. 构建融合型分层培训体系

为系统提升 AI 时代英语教师测评素养, 需构建一套涉及职前培训与在职培训, 理论与实践并重的融合型分层培训体系。该体系应结合教师专业发展不同阶段, 设计针对性的培训内容。职前培训阶段, 师范院校应开设《AI 语言测评技术基础》《教育数据科学入门》等核心课程。这些课程既要讲解自然语言处理、机器评分模型等技术原理, 还要通过实操训练来帮助未来教师建立起对 AI 技术的全面认识。比如可组织对比 AI 与人工批改同一篇作文的教学活动, 让未来教师直观体会自然语言处理技术的优势与局限, 建立对技术的批判性认识。同时应配套开设《教育统计学》等基础课程, 通过系统地理论学习与实践训练, 培养未来教师“用数据指导决策的能力”[20][21]。

在职培训应立足课堂教学实际, 构建阶梯式工作坊, 切实回应不同教师群体的差异化需求。初级工作坊以技术应用为核心, 借助实操训练帮助教师掌握智能测评工具的基本操作, 例如讯飞口语测评系统、批改网等平台的操作等, 同时深化主流测评工具在班级管理、任务分配等核心功能方面的应用能力。在实施过程中应考虑到城乡差异和学段差异。例如, 农村学校可侧重组织线下实操工作坊并分发图文指导手册, 城市学校则可引入线上模拟平台对职前教师进行培训和强化。中小学教师应重点掌握如何将工具融入日常教学环节, 而高校教师则需进一步了解不同工具的适用场景与参数设置逻辑。

中级工作坊着重培养教师的数据分析能力, 指导教师利用 Excel、SPSS 等工具深入挖掘测试数据, 对 AI 生成的班级学情报告进行二次分析, 精准找出学生的薄弱环节, 并据此设计分层教学任务。对于硬件与网络条件受限的农村学校, 培训应着重授人以渔, 指导教师如何利用手机等便携设备有效采集离线学习数据(如录音、拍照), 并利用 Excel 等基础工具进行整理与可视化分析。而资源充沛的城市学校则可进一步培训教师利用智慧课堂等平台进行实时学情监测并及时调整教学策略。中小学教师要学习如何将测评数据与中高考的能力要求精准对接, 使测评过程和结果直接服务于中高考; 大学教师应注重从学术写作、口语表达等多个维度综合挖掘数据背后的教学启示, 推动教学内容的深化。

高级工作坊应将伦理思辨与教学创新能力的培养置于核心地位。例如, 可以开设“AI 伦理与教育公平”等专题研讨, 组织教师对教学中可能遇到的算法偏见、数据安全等现实问题进行深入探讨, 并通过模拟真实伦理困境, 提升教师的伦理敏感度与应对能力。在此基础上, 可适当引入 Python 等工具的基础应用, 帮助教师根据自身教学需求对现有测评流程进行个性化调整。在具体实施过程中, 需要充分考虑不同地区的实际情况。农村学校可以组织教师围绕典型教学案例进行深度研讨; 而资源较为丰富的城市学校可以与相关科技企业合作, 开展实训项目。不同学段的教师也应有所侧重, 中小学教师更需要提升在复杂教学情境中的伦理判断与决策能力; 高校教师则可以在理解技术原理的基础上, 进一步参与测评工具的优化甚至开发工作, 实现从使用者到共建者的角色转变。

该培训系统兼顾层次性和灵活性, 既能够紧扣教师专业发展的实际需求, 又能切实关照不同地区、不同学段的具体情况, 为教师在人工智能背景下系统提升测评素养提供了一条切实可行的路径。

4.2. 推动测评理念的深层转型

人工智能时代背景下, 英语教师需从根本上重塑测评理念, 确立以“人机协同、素养导向”为核心

的新型评价观。教师应将 AI 工具定位为“测评协作者”与“超级助手”，充分借助其在语法检测、发音诊断等基础性语言技能评估中的优势，集中精力去设计综合性测评任务、解读多维学习数据，为学生的思维过程以及创新表达能力提供专业的发展性指导。

教师的测评理念还需从“知识本位”向“素养本位”转型。沈书生与祝智庭[22]指出“思维比知识重要，问题比答案重要，逻辑比罗列重要”，英语测评应当超越对孤立语言知识的考查，更侧重于设计可激发批判性思维、彰显文化理解与沟通能力的真实情境任务。教师应当把逻辑组织、观点论证以及问题解决等关键思维品质纳入评价维度，依靠技术手段达成从单纯评估语言形式到综合考查语言运用与思维品质的转变，使得测评切实服务于学生综合素养的全面发展。

4.3. 深化人机协同的测评实践

在确立“人机协同、优势互补”测评理念的基础上，教师需进一步探索将这一理念转化为具体实践的有效路径。文秋芳与梁茂成[23]提出的人机互动协商能力模型能够为英语教师提供可操作的理论框架，该模型包含理解 AI、设定目标、发布指令、分析反馈和调整策略五个关键环节，教师可以基于此框架在教学实践中探索多样化的协同模式。例如，英语教师可根据“四步协同评价流程”进行写作评价。首先由 AI 系统对学生作文进行初步筛查，自动识别语法错误、词汇搭配及句式复杂度等表层语言特征；然后教师可基于 AI 生成的诊断报告，重点关注 AI 难以深入评判的逻辑结构、文化内涵等维度并对学生作文中存在的 key 问题进行标注；接着，教师将 AI 提供的评分结果与自身专业判断进行对比，观察两者存在的差异，并记录可能出现偏差的原因。例如 AI 可能对模板化的表达评分偏高，对作文中的创新性表达却不能够充分和有效识别；最后，教师结合 AI 数据和自身经验，形成全面的反馈报告，并据此设计分层教学任务，对不同水平的学生采取不同的教学措施。这样结构清晰、分工明确的协作流程不仅能够提升测评效率，而且能够确保测评始终服务于学生的综合素养。

4.4. 完善制度保障与激励机制

推动英语教师测评素养转型，需构建多层次协同支持体系。教育行政部门应牵头修订评价政策与课程标准，明确倡导过程性评价、能力导向评价与多元评价方式，推动从“唯分数论”转向关注学生核心素养、学习过程与增值评价，为教师应用多元 AI 测评工具提供政策空间与制度认可；学校需提供稳定可靠的硬件与网络设施，采购与开发安全可靠的 AI 测评工具与数据分析平台，并减轻教师事务性负担，为其学习 AI 技术、实践新方法、分析数据留出时间和空间；校级或区级专业团队需及时响应，为教师解决技术难题，提供专业咨询、方法指导和案例解析，并系统开展培训提升教师测评素养与数据决策能力；学校及教育管理部门还需建立认可与奖励机制，通过设立“智能测评创新奖”、将 AI 测评应用成果纳入评优评先体系等方式，表彰在人机协同测评中成效显著的教师，发挥优秀教师的示范作用，推动教师之间的经验交流，有效激发教师的创新活力，最终实现英语教师测评素养全面提升。

4.5. 强化数据伦理与人文关怀

著名教育家 Martin Buber 认为，教育应超越单纯的知识传授，转向精神生活能力的培养。在智能教育测评深入推进下，英语教师必须意识到语言教学的目的是育人而不是育“器”。首先，英语教师需要辩证看待人工智能的双重影响，既要认识到其在提升测评效率、减轻工作负担方面的显著优势，也要清醒认识其技术局限。教师应始终秉持审慎的批判思维，对人工智能的回答进行核对和验证，加强对 AI 生成的评分、诊断报告及学习路径建议的鉴别与研判，确保答案的正确性，并严格把控其教育伦理指向与科学有效性，使人机协同测评真正促进教学改进与学生发展。人工智能模型在可解释性、真

实性与泛化能力上的表现,以及围绕它们所构建的新型研究方法,均深刻影响着语言测评的效度论证。特别是在高风险测评中,教师必须确保人工智能提供的任何结论都具有充分的效度依据,能够支持后续的教学决策[24]。

其次,测评过程应该始终坚持“以学生为本”。教师应超越冰冷的数据指标,关注影响学生表现的深层因素,包括个体差异和文化背景等,并通过提供及时、具体且富有同理心的积极反馈,帮助学生建立自信心,激发学习兴趣。最后,教师必须成为学生数据安全的守护者,严格遵守数据采集与使用的伦理规范,践行最小化原则,切实保护学生隐私。同时,教师应帮助学生建立对个人数据安全的正确认知,并引导他们以批判的态度看待算法得出的结论。针对 ChatGPT 等工具可能引发的学术诚信问题[25],教师应制定明确的使用规范,坚守教育公平与学术诚信的底线。

5. 结语

人工智能正在不可逆转地改变英语教学测评的生态,但这并非教师角色的终结,而是其测评素养重塑与价值升华的重要契机。在智能技术深刻重塑教育生态的当下,教师需要以开放的姿态迎接创新工具,同时保持批判意识,审慎考察技术应用背后的逻辑与价值取向,将人工智能视为虚拟世界中的“镜像人类”,坚守教育的人文本质。

本研究系统阐释了 AI 时代英语教师测评素养的核心内涵,揭示了其在技术应用、数据决策与伦理审视等方面的拓展,并深入剖析了教师在技术适应、数据素养、认知转型及伦理风险等领域面临的多维挑战。基于此,从培训体系构建、测评理念转型、人机协同深化、制度保障完善及伦理教育强化五个维度提出具体可行的发展策略,形成了系统化的支持方案。只有通过多方共同努力,才能确保 AI 真正服务于英语测评实践,最终实现技术赋能下“人的教育”的最大化,走向人机协同的智慧测评新未来。

参考文献

- [1] 许娟,董君. 外语教师测评素养内涵与发展路径探究[J]. 现代英语, 2023(2): 51-54.
- [2] Stiggins, R.J. (1991) Assessment Literacy. *Phi Delta Kappan*, **72**, 534-539.
- [3] Brindley, G. (2001) Language Assessment and Professional Development. Cambridge University Press.
- [4] Davies, A. (2008) Textbook Trends in Teaching Language Testing. *Language Testing*, **25**, 327-347. <https://doi.org/10.1177/0265532208090156>
- [5] Fulcher, G. (2012) Assessment Literacy for the Language Classroom. *Language Assessment Quarterly*, **9**, 113-132. <https://doi.org/10.1080/15434303.2011.642041>
- [6] 中华人民共和国教育部办公厅. 关于组织实施数字化赋能教师发展行动的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A10/s7034/202507/t20250704_1196586.html, 2025-07-04.
- [7] 高旭阳,盛雅乔. 语言教师评价素养: 内涵、框架与展望[J]. 外语研究, 2023, 40(6): 44-51.
- [8] 金艳. 外语教师评价素养发展: 理论框架和路径探索[J]. 外语教育研究前沿, 2018, 1(2): 65-72.
- [9] 池美英,张丹. 大数据驱动下的高校外语教师测评素养提升策略探索[J]. 韩国语教学与研究, 2024(4): 72-76.
- [10] 尹亮宇. 大数据背景下高校外语教师测试与评价素养的发展研究[J]. 重庆电子工程职业学院学报, 2022, 31(4): 93-97.
- [11] 岳敏,何婧,岳菊. 智能时代的在线教师数据素养框架构建研究[J]. 湖北开放大学学报, 2025, 45(3): 3-9.
- [12] 杨现民,骆娇娇,刘雅馨,等. 数据驱动教学: 大数据时代教学范式的新走向[J]. 电化教育研究, 2017, 38(12): 13-20.
- [13] Xi, X. (2023) Advancing Language Assessment with AI and ML-Leaning into AI Is Inevitable, but Can Theory Keep Up? *Language Assessment Quarterly*, **20**, 357-376. <https://doi.org/10.1080/15434303.2023.2291488>
- [14] Kooli, C. (2023) Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability*, **15**, Article 5614. <https://doi.org/10.3390/su15075614>
- [15] 刘建达. 人工智能时代的语言测评: 机遇与挑战[J]. 现代外语, 2024, 47(6): 859-869.

-
- [16] 金银星, 孙凤娟. 基于生成式人工智能的大学生英语口语愉悦情绪干预研究[J]. 外语教育研究前沿, 2025, 8(3): 97-109.
- [17] Yan Z., et al. (2018) Data Literacy in Education: A Systematic Review. *Teaching and Teacher Education*, **75**, 32-44.
- [18] 杨现民, 李新. 中小学教师数据素养的现状、评价及意义[J]. 教师教育学报, 2021, 8(3): 12-20.
- [19] 鹿星南, 高雪薇. 人工智能赋能教育评价改革: 发展态势、风险检视与消解对策[J]. 中国教育学刊, 2023(2): 48-54.
- [20] Kippers, W.B., Poortman, C.L., Schildkamp, K. and Visscher, A.J. (2018) Data Literacy: What Do Educators Learn and Struggle with during a Data Use Intervention? *Studies in Educational Evaluation*, **56**, 21-31. <https://doi.org/10.1016/j.stueduc.2017.11.001>
- [21] 王松丽. “互联网+”课堂变革背景下教师评价素养的发展[J]. 教育测量与评价, 2020(3): 10-14.
- [22] 沈书生, 祝智庭. ChatGPT 类产品: 内在机制及其对学习评价的影响[J]. 中国远程教育, 2023, 43(4): 8-15.
- [23] 文秋芳, 梁茂成. 人机互动协商能力: ChatGPT 与外语教育[J]. 外语教学与研究, 2024, 56(2): 286-296.
- [24] Hao, J., von Davier, A.A., Yaneva, V., Lottridge, S., von Davier, M. and Harris, D.J. (2024) Transforming Assessment: The Impacts and Implications of Large Language Models and Generative AI. *Educational Measurement: Issues and Practice*, **43**, 16-29. <https://doi.org/10.1111/emip.12602>
- [25] Susnjak, T. and McIntosh, T. (2024) ChatGPT: The End of Online Exam Integrity? *Education Sciences*, **14**, Article 656. <https://doi.org/10.3390/educsci14060656>