

生成式人工智能赋能EFL课堂教学测评的公平性问题

——基于师生双视角的研究

王静*, 董子涵, 戴文蕙, 张泽涵

中国矿业大学(北京)文法学院, 北京

收稿日期: 2026年4月6日; 录用日期: 2026年5月8日; 发布日期: 2026年5月14日

摘要

近年来生成式人工智能(GenAI)在教育测评领域广泛应用, 测评的公平性面临新的挑战。本研究采用混合研究方法, 通过对209名高校教师和637名学生的问卷调查与深度访谈, 探讨了GenAI赋能EFL教学测评场景下, 教师和学生对测评公平性的认知现状、群体差异及影响因素。研究发现: 1) 师生对GenAI介入测评的公平性整体持审慎态度, 尤其在“AI辅助与独立完成作业采用同一评分标准”的情景中, 师生感知的公平风险最高; 2) 师生在公平性关注点上存在显著差异, 教师更关注评分规则的形式公平与技术中立, 而学生更在意资源获取的机会平等与竞争公平; 3) 评分标准的适配性、AI工具的可及性以及个体技术素养是制约测评公平性的核心因素。研究结果为构建数智时代测评公平、规范教学测评体系提供了实证依据。

关键词

生成式人工智能, EFL课堂教学测评, 公平性, 师生双视角, 影响因素

Research on the Fairness of Generative AI-Empowered EFL Classroom Assessment

—From Dual Perspectives of Teachers and Students

Jing Wang*, Zihan Dong, Wenhui Dai, Zehan Zhang

School of Law and Humanities, China University of Mining and Technology (Beijing), Beijing

Received: April 6, 2026; accepted: May 8, 2026; published: May 14, 2026

*通讯作者。

文章引用: 王静, 董子涵, 戴文蕙, 张泽涵. 生成式人工智能赋能 EFL 课堂教学测评的公平性问题[J]. 教育进展, 2026, 16(5): 826-834. DOI: 10.12677/ae.2026.165926

Abstract

In recent years, Generative Artificial Intelligence (GenAI) has been widely applied in the field of educational assessment, bringing new challenges to assessment fairness. This study employs a mixed-methods approach—comprising a questionnaire survey of 209 university teachers and 637 students, alongside in-depth interviews—to explore the current state of cognition, group differences, and influencing factors regarding assessment fairness in GenAI-empowered EFL (English as a Foreign Language) teaching and assessment scenarios. The study findings indicate that: 1) Both teachers and students generally maintain a cautious attitude toward the intervention of GenAI in assessment; the perceived fairness risk is highest in the scenario where “the same scoring criteria are applied to both AI-assisted and independently completed assignments”. 2) Significant differences exist between teachers and students regarding their focus on fairness: teachers are more concerned with the formal fairness and technical neutrality of scoring rules, while students care more about equal opportunity in resource acquisition and competitive fairness. 3) The core factors impacting assessment fairness include the adaptability of scoring standards, the accessibility of AI tools, and individual technical literacy. The results of this research provide empirical evidence for constructing a fair assessment framework and a standardized teaching assessment system in the digital-intelligence era.

Keywords

Generative Artificial Intelligence (GenAI), EFL Classroom Assessment, Fairness, Dual Perspectives of Teachers and Students, Impacting Factors

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

课堂教学测评作为连接教学过程与学习效果的核心纽带，其公平性是教育公平的重要具象化体现，直接关乎教育目标的达成与学生的全面发展。尤其是在高校 EFL (English as a Foreign Language) 教学场景中，课堂教学测评不仅承担着对学生语言能力进行量化评估的基础功能，更涉及跨文化语境下的机会平等、人格尊重与成长激励。因此，其公平性内涵既包含传统意义上评分标准的一致性、测评过程的透明度、结果评定的公正性[1]，也涵盖对不同文化背景、不同学习基础学生的包容性与差异化关怀[2]。长期以来，测评的公平性一直是教育领域的核心议题，刘秀梅等[3]的研究发现公平且符合伦理规范的测评行为能够保护学生自尊心、激发学习兴趣，而失范行为则可能挫伤学生信心、损害教育生态，凸显了测评公平性在教学实践中的重要性。

近年来，GenAI (如 ChatGPT、DeepSeek 等) 技术的快速迭代与普及，为教育领域带来了革命性变革。在高校 EFL 教学中，GenAI 已广泛应用于课程内容生成、教学测评、个性化辅导、语言纠错等环节，凭借其高效性、交互性优势显著优化了教学流程与学习体验[4] [5]。

GenAI 的技术特性决定了其介入测评环节必然对传统公平性内涵产生冲击与拓展。现有研究多聚焦 GenAI 在教学内容供给、教学模式创新等方面的赋能作用[6]，对其介入课堂教学测评环节的系统性探索相对匮乏，尤其缺乏对测评公平性这一核心伦理议题的深入关注[7]。目前，GenAI 与课堂教学测评的融合已呈现不可逆转的趋势，其在口语流利度实时评分、写作自动评价、测评标准智能化制定等场景的应

用,正逐步重塑传统测评的运作逻辑与实践形态,但这一融合过程并非毫无风险。上述潜在风险若得不到及时且有效回应,不仅会削弱测评结果的客观性与公信力,更可能违背教育公平的核心准则,阻碍 GenAI 在教育领域“智能向善”的发展与运用[8]。

教师与学生作为课堂教学测评的双主体,其对 GenAI 赋能测评公平性的感知与认知直接影响技术应用的接受度与实践效果。鉴于此,本研究从师生双视角出发,以高校 EFL 课堂测评为具体场景,采用问卷调查与半结构化访谈法进行实证数据收集,系统探讨 GenAI 介入 EFL 课堂教学测评背景下师生群体分别感知的公平性挑战、双方在公平性认知的异同及影响公平性实现的关键因素。本研究关注深化教育公平理论在智能技术场景中的适用性,为构建 GenAI 赋能课堂教学测评的公平性保障体系提供实证支撑,研究结果同时供高校 EFL 教师优化测评实践、教育管理部门制定相关测评政策借鉴,以期推动 GenAI 与课堂教学测评的伦理化、公平化融合。

2. 文献综述

2.1. GenAI 与 EFL 课堂教学测评的融合演进

GenAI 与 EFL 课堂教学测评的融合,已从技术探索、实践应用逐步进入伦理反思的深化阶段。早期研究聚焦技术可行性,顾景倩[4]以讯飞星火为例,验证了 GenAI 在大学英语教学测评中的辅助价值,其在口语流利度评分、写作语法纠错等场景的应用可显著提升测评效率。随着应用深化,后续研究开始关注技术融合中的实践困境,王佑镁等[9]指出,GenAI 在测评应用中兼具技术赋能与伦理风险,生成内容可靠性不足、评价过程“算法黑箱”、教学适配薄弱等问题,对测评科学性构成显著挑战。

在 EFL 课堂场景中,GenAI 的融合呈现显著的场景依赖与能力分层特征。一方面,其在词汇测评、语法纠错等标准化评估中表现出高效优势;另一方面,在跨文化交际能力、批判性思维等高阶能力测评中,GenAI 缺乏对语境与人文因素的把握,易引发测评偏差。孙婷婷[6]的数智化转型研究显示,当前 GenAI 与外语教学测评的融合仍停留在表层工具应用,深度融合机制尚未形成,且公平性这一核心伦理问题尚未得到充分关注与系统性回应。

2.2. 课堂教学测评公平性的理论内涵与维度建构

课堂教学测评公平性的内涵随教育理念与技术发展不断拓展。现有研究中,Green [1]将测评公平性界定为评分标准一致、测评过程透明、测评结果客观公正三大核心维度。Fan 等[2]基于中国高等教育语境的研究进一步补充,强调测评公平还应包括对不同文化背景、不同学习基础学生的包容性。

在智能技术赋能背景下,测评公平的理论框架得以拓展与升级。本研究以 Messick [10]提出的统一效度与测评公平整合框架为核心理论基础,结合算法公平理论与语言测评公平研究成果,构建适配 GenAI 环境的 EFL 测评公平分析框架:机会公平、过程公平与结果公平。

在此框架下,GenAI 赋能测评的公平性新增算法公平与数据公平两大关键维度:算法公平强调模型去偏、透明与可审计;数据公平强调隐私保护、知情同意与合规使用[9]。由此,测评公平从传统的“程序-结果”结构,拓展为“机会-过程-结果-算法-数据”五位一体的整合公平体系,更贴合智能测评的实践特征。综合现有研究,本研究将 GenAI 赋能课堂教学测评的公平性界定为:在 GenAI 介入的测评全流程中,保障不同学生在测评机会、过程参与、结果评定等方面的平等地位,规避算法偏见、数据滥用等导致的不公平现象,实现技术效能与教育公平的协同。

2.3. GenAI 赋能测评的公平性争议与研究局限

现有研究已识别出 GenAI 赋能教学测评的多重公平性风险。其一,算法偏见导致过程与结果不公。

GenAI 受训练数据偏差与标注者主观倾向影响,可能对非英语母语背景、地域、性别不同的学生产生隐性歧视,刘艳红[11]进一步验证,算法模型的“数据偏见”会直接转化为评分偏差,影响测评结果的公正性。其二,数据隐私泄露侵犯学习者权利公平。王佑镁等[9]的伦理风险管理框架研究显示,GenAI 在测评中收集的学生学术记录、语音信息等隐私数据,存在未经授权使用的风险,侵犯学生的数据主权。其三,学术诚信边界模糊引发竞争公平失衡。谢娟[8]指出,GenAI 辅助下的内容生成可能导致学生学术诚信认定标准混乱,部分学生过度依赖人工智能工具获得不正当优势,破坏测评的公平竞争环境。其四,技术可及性差异与数字鸿沟加剧起点不公。孙婷婷[6]发现,不同学生对 GenAI 工具的获取与使用能力存在差异,设备条件、付费功能、数字素养的分层会直接扩大测评结果差异,形成新的教育不平等。

尽管相关研究已关注 GenAI 测评中的公平性问题,但仍存在三方面明显局限:其一,研究视角单一化,多聚焦教师或技术层面,缺乏对师生双视角的对比分析,忽视公平性认知的主体差异性与互动性;其二,研究场景泛化,针对高校 EFL 课堂的针对性、情境化研究不足,忽视语言教学测评的交际性、跨文化性与过程性特征;其三,研究内容碎片化,多关注单一公平性风险,缺乏成熟理论框架统领,对公平认知差异、群体特征及系统性影响因素的探讨欠缺。基于上述研究缺口,本研究以整合型公平框架为依托,从师生双视角出发,聚焦 EFL 课堂教学测评场景,系统探究 GenAI 融入后师生对公平性的认知特征、群体差异与核心影响因素,以期弥补现有研究不足。

3. 研究设计

3.1. 研究方法

本研究采用混合研究方法。首先,通过在线问卷开展大规模调查,获取师生对 GenAI 赋能测评场景下公平性认知的总体特征;其次,运用半结构化访谈收集质性数据,深入解析问卷结果中出现的低一致性矛盾场景,探讨影响公平性的深层因素,兼具研究的统计广度与分析深度。

3.2. 研究对象

研究所使用的问卷采用“本校 + 外校”、“学生 + 教师”全域覆盖模式:首先覆盖不同层次高校(重点高校与普通高校),然后以线上投放为主、线下辅助推广,联系本校及外校师生共同参与,确保样本具有代表性。教师调研对象均为高校外语系教师,样本覆盖北京市、陕西省、山东省、山西省共 4 个省级行政区,涉及 15 所高校:中央部属重点高校(中国矿业大学(北京)中国地质大学(北京)、中国人民大学、北京航空航天大学、北京邮电大学、北京理工大学、北京化工大学、中国传媒大学、中国人民公安大学、青岛大学、太原理工大学)、北京市属院校(北京服装学院)、省属重点师范院校(首都师范大学、山西师范大学)、地方师范专科院校(吕梁师范专科学校);学生调研对象均为英语专业学生,包括本科生和研究生,主要来源为中国矿业大学(北京),同时涵盖山东省的山东师范大学、安徽省的合肥工业大学,样本以中央部属重点高校为主体、省属重点高校为补充,专业背景统一且院校层次多元。此外,在问卷发放阶段,研究团队严格遵循学术伦理规范,充分确保了所有调查参与者的知情同意权。并于在线问卷的首要位置清晰地标明了本研究的学术目的、具体流程以及数据保密承诺。

在访谈环节,研究团队从问卷作答者中,抽取了 26 名师生开展深度访谈。其中包含学生 14 名(5 名本校专业英语本科生、5 名本校专业英语研究生、4 名外校英语专业本科生)和教师 12 名(8 名本校外语系教师、4 名外校外语系教师)。在筛选教师访谈对象时主要依据教师职业特征的多样性,涵盖不同教龄段、不同性别和不同岗位属性(5 名教师兼任行政职务、7 名教师担任教学科研岗位)等维度。访谈全程严格匿名,规范管理录音与笔记,保障受访者隐私安全。

3.3. 研究工具

本研究的核心研究工具为在线问卷和半结构化访谈。二者均围绕 GenAI 在高校语言类课程教学测评公平问题设计, 经过多阶段打磨和验证形成终版。

问卷包含 2 个测评情景, 聚焦测评公平性。在情景编制阶段, 项目组成员分工协作, 结合高校语言类课程教学实际, 围绕课堂汇报、课后作业、期末考核等测评形式, 多次研讨, 初步筛选出 4 个与公平性相关的教学测评情景初始版本。在文献整合阶段, 研究团队以现有文献为框架, 参考 Johnson [2]、刘秀梅[12]和 Fan [2]研究中的问卷结构, 同时根据最新文献中 GenAI 背景下课堂教学测评中的伦理风险, 对 4 个初始情景进行多次调整与优化, 最终确定 2 个典型的教学测评情景。在正式发放问卷前, 研究团队邀请 2 名高校教师、4 名本科生试测问卷初稿, 对问卷内容和语言表述进行核对把关, 以保障问卷科学性。根据试测人员的反馈建议, 研究团队凝练冗余表述, 对模糊情景进行具象化调整, 保证问卷语言表述清晰无歧义。

为深入分析师生群体对测评公平性的差异化认知及影响因素, 研究团队开展了半结构化访谈。访谈过程依循“情景描述 + 合理性判断 + 理由阐释 + 个人建议”的形式, 追问环节遵循“回应式挖掘”原则, 采用“预设问题 + 即时回应”相结合的方式, 既参考提前列出的补充问题, 又根据受访者的回答灵活变通, 深入探究不同观点背后的逻辑和成因。所有问题均为开放式问题, 避免引导性表述, 充分确保访谈对象的表达自主性。研究团队征得受访人同意, 对访谈过程进行录音。

3.4. 数据处理

本研究通过问卷星平台发放在线问卷, 共回收有效问卷 834 份, 其中教师问卷 209 份、学生问卷 637 份。该平台在导出阶段已完成了对原始数据的初次处理, 自动生成了各题项的作答数量和比例。研究团队再次审核, 剔除了无效数据, 确保数据的有效性和可用性。

在访谈结束后, 研究团队对 26 份访谈录音逐一转录、校对, 剔除无效信息后初步形成 26 份访谈文本稿。在编码阶段, 研究团队首先进行预编码, 即文本套色, 用不同颜色标记关键信息如受访人的决策、理由、建议等; 其次结合演绎式编码、归纳式编码和原话编码三种方式进行正式编码; 接着根据两次编码结果分析比对、归纳提炼, 总结出研究主题。在整个编码过程中, 为了保障编码的有效性和严谨性, 研究团队采用两两一组的形式, 各组共同对全部访谈的 20% 进行编码, 经 Cohen's Kappa 系数检验, 编码者之间的一致性在 90% 以上, 其余 80% 的访谈由各小组成员分工进行单独编码。编码过程中的不确定部分通过组内讨论解决。最终形成了完善的编码分析结果, 为后续研究提供支撑。

3.5. 研究局限性

本研究采用方便抽样获取调研数据, 样本在地域分布与院校类型上存在一定侧重, 对结论的普适性构成有限约束。教师样本以北京及部分省份重点院校为主, 部分通过教师社群发放, 样本覆盖范围存在一定集中度; 学生样本来源相对有限。由此, 本研究结论在向更多类型院校、更广学生群体推广时需保持审慎。

4. 研究结果

4.1. 教师和学生对教师行为测评公平性的一致程度分析

通过统计受访教师和学生群体对测评情景 1 和 2 中教师测评行为是否合适的判断比例, 研究团队得出教师和学生对测评公平性判断的一致程度, 见表 1。

表 1 表明教师受访群体对测评情景 1 中教师行为“合适”与“不合适”的判断趋近 50% 与 50%, 表

明受访群体对 GenAI 赋能 EFL 课堂教学测评后的公平性认知内部一致性很低；而学生群体对测评情景 1 和 2 中教师行为“合适”与“不合适”的判断都趋近 50%，表明学生群体对测评公平性认知的内部一致性也很低。

Table 1. Consistency between teachers and students in their judgments regarding the fairness of the assessment

表 1. 教师和学生对测评公平性判断的一致程度

情景序号	受访主体	合适占比(%)	受访主体	不合适占比(%)
情景 1	教师	41.63	教师	58.37
	学生	52.90	学生	47.10
情景 2	教师	27.75	教师	72.25
	学生	48.67	学生	51.33

表 1 还表明教师对 GenAI 介入测评的负面感知更为显著，对两个场景中教师做法持“不合适”态度的比例均超半数，其中在情景 2 中达到 72.25% (151 人/209 人)，在情景 1 中为 58.37% (122 人/209 人)，且以测评公平性受损为首要感知。学生对两个场景中教师做法持“不合适”态度的比例均接近半数，在情景 2 中为 51.33% (327 人/637 人)，在情景 1 中为 47.1% (300 人/637 人)，同样以公平性受损为核心，比例略低于教师。

4.2. 影响 GenAI 在测评应用中的公平性实现的关键因素

为了提供更为直观的对比，研究团队选取了问卷调查中受访教师与学生认为最重要的两个影响因素列入表格。见表 2 和表 3。

从表 2 可知，在两个测评情景中，受访教师群体认为 AI 赋能后的 EFL 测评情景中教师测评行为

Table 2. Key factors influencing teachers' perception of fairness after GenAI-enabled EFL classroom teaching assessment

表 2. GenAI 赋能 EFL 课堂教学测评后影响教师公平性认知的关键因素

情景序号	合适		不合适	
	因素	占比(%)	因素	占比(%)
情景 1	发挥技术赋能	81.61	损害测评公平性	82.79
	现实工作应用导向	71.11	能力评估失真	66.39
情景 2	保障评分一致性	82.76	损害测评公平性	88.74
	保持技术中立性	72.41	教育机会不平等	70.86

Table 3. Key factors influencing students' perception of fairness after GenAI-enabled EFL classroom teaching assessment

表 3. GenAI 赋能 EFL 课堂教学测评后影响学生公平性认知的关键因素

情景序号	合适		不合适	
	因素	占比(%)	因素	占比(%)
情景 1	发挥技术赋能	75.67	损害测评公平性	76.33
	现实工作应用导向	70.62	语言能力评估失真	68.33
情景 2	保障评分一致性	72.90	损害测评公平性	81.35
	保持技术中立性	71.29	教育机会不平等	68.20

合适的主要原因包括：保障评分一致性(82.76%)和发挥技术赋能(81.61%)、保持技术中立性(72.41%)和现实工作应用导向(71.11%)；受访教师群体认为 AI 赋能后的测评情景中教师测评行为不合适的主要原因包括：损害测评公平性(情景 2, 88.74；情景 1, 82.79%)、教育机会不平等(70.86%)和能力评估失真(66.39%)。

表 3 表明，在两个测评情景中，受访学生认为 AI 赋能后的 EFL 测评情景中教师测评行为合适的主要原因包括发挥技术赋能(75.67%)、保障评分一致性(72.90%)、保持技术中立(71.29%)和现实工作应用导向(70.62%)；不合适的主要原因有损害测评公平性(情景 2, 81.35%；情景 1, 76.33)、语言能力评估失真(68.33%)和带来教育机会不平等(68.20%)。

对比表 2 和表 3 发现，在测评情景 1 中，教师和学生都将“发挥技术赋能”和“现实工作应用导向”列为首要与次要理由，说明教师认可 GenAI 在提升教学测评效率、衔接职场现实需求方面的技术价值，认为 GenAI 具有赋能课堂教学测评的实践属性。学生也认同 GenAI 对提升语言能力、对接未来职业需求的积极作用。同时，教师和学生都将“损害测评公平性”和“语言能力评估失真”列为主要担忧，说明教师关注 GenAI 介入后，统一评分标准下测评过程公平与结果公平的双重失衡。学生对“能力评估失真”的感知比例(68.33%)略高于教师(66.39%)，反映出学生作为测评参与者，对 GenAI 掩盖自身真实语言能力的体验更为直观。

在测评情景 2，教师和学生都将“保障评分一致性”和“保持技术中立性”列为首要与次要理由，但学生的两项比例均低于教师，反映出学生对形式公平规则的关注度相对较低，教师更看重 AI 介入后评分规则的统一性与中立性，将其视为维护测评形式公平的关键因素。同时，教师和学生都将“损害测评公平性”和“教育机会不平等”列为主要担忧，且两项比例均为所有场景中最高，说明教师清晰认识到 GenAI 工具资源差异带来的起点公平缺失。学生选择的比例与教师接近，体现出学生也能直观感受到资源差异导致的测评结果不公与教育机会失衡。

4.3. 访谈结果分析

访谈结果与问卷数据高度契合，在公平性认知、核心担忧及技术认同上达成统一。教师视角下，受访者提出应按测评利害层级实施差异化管理：高利害考试(如期末考)应坚持人工评分以确保客观真实；低利害测评(如平时作业)可采用“GenAI 辅助 + 教师复核”模式。针对学术诚信，教师建议建立 GenAI 使用声明制度，并由学校统一提供免费工具、明确使用规范，评价应侧重学生原创观点。对于禁令争议，多数教师认为全面禁止不具实操性，主张引导而非“一刀切”。

学生视角下，受访者担忧 GenAI 过度介入会削弱师生互动及教师的个性化指导功能。在程序正义方面，学生强调判定 AI 抄袭应遵循“无罪推定”原则，警惕检测工具的误判风险，并主张赋予学生申诉权。此外，学生反思了教学本质的偏离风险，指出 AI 介入可能导致学生“重语言形式、轻逻辑思维”，违背培养沟通能力与创新精神的根本目标。

5. 成因分析

近年来生成式人工智能(GenAI)在教育测评领域的广泛应用对公平性提出了新挑战。本研究采用混合研究方法，通过对 209 名高校教师和 637 名学生的问卷调查及 26 名师生的深度访谈发现：师生对 GenAI 介入测评整体持审慎态度，尤其在“AI 辅助与独立作业采用同一标准”的情景中感知风险最高。教师视角侧重评分规则的形式公平与利害分级管理，学生则更在意资源获取的机会平等，并主张在学术诚信判定中遵循“无罪推定”与程序正义。研究指出，GenAI 测评公平性问题是算法偏见、制度缺失及数字鸿沟交织的结果。

算法偏见本质是 GenAI 的核心技术局限，根源在于训练数据的偏差。训练数据存在固定的时间截止

点,无法实时更新最新的语言表达,导致 GenAI 模型对语境的理解能力不足,易将学生的创新思路和表达错判为错误答案,形成系统性偏见。

制度缺失是当前 GenAI 在高校 EFL 课堂测评应用中普遍面临的问题。国家、高校层面均未出台统一、明确的应用规范与操作准则,GenAI 测评可使用场景和边界不明确,因此学生的程序性权利难以得到保障,易引发测评公平性争议。

数字鸿沟是造成课堂教学测评不公与机会不均的现实根源。免费版和付费版 GenAI 存在明显差异,造成学生因经济水平不同而产生资源获取的不公问题,且学生使用 GenAI 的能力也存在个体差异性,当前大多数学校既没有禁止学生使用 GenAI,也没有给学生提供免费的统一的 GenAI 平台,加剧了课堂教学测评中的不公平问题。

综上,GenAI 在课堂教学测评中引发的公平性问题,是技术、制度、资源与能力等多方面因素共同作用的结果。

6. 对策建议

鉴于上述成因以及 GenAI 进课堂的重要性,结合课堂教学的实际需求,本研究将从教师、学生两个主体出发,提出针对性的建议,以提升 GenAI 赋能 EFL 课堂测评的公平性。

GenAI 能为教学测评提供个性化支持,且丰富教学资源 and 形式,为此,教师要合理利用 GenAI,在使用 GenAI 自动评分时增加人工复核环节,对学生的主观化作答进行人工评分,发挥技术效率作用的同时保障教学的人文关怀;搭建教师交流平台,邀请一线优秀教师开展 GenAI 赋能高校 EFL 课堂公平性的实操演练,定期开展培训,引导教师形成符合自己课堂的测评公平性保障方案。

为保障 GenAI 进课堂的教学公平性,学生要保持对 GenAI 测评结果的敏感性,辩证看待评分结果,如有异议,及时向教师申请人工复核,维护自身权益、保障结果公平;加强学术诚信与规范教育,树立正确的 GenAI 使用观念,避免触碰 GenAI 在学习中的边界,合理利用 GenAI 辅助自主学习。

综上所述,保障 GenAI 赋能 EFL 课堂的测评公平,需师生与学校协同发力。高校应制定使用规范与监管机制,消除数字鸿沟,以多方共治坚守教育公平底线。

基金项目

本文系中国矿业大学(北京)本科生大学生创新训练项目(项目编号:202508007)研究成果。

参考文献

- [1] Green, S.K., Johnson, R.L., Kim, D. and Pope, N.S. (2007) Ethics in Classroom Assessment Practices: Issues and Attitudes. *Teaching and Teacher Education*, **23**, 999-1011. <https://doi.org/10.1016/j.tate.2006.04.042>
- [2] Fan, X., Liu, X. and Johnson, R.L. (2020) A Mixed Method Study of Ethical Issues in Classroom Assessment in Chinese Higher Education. *Asia Pacific Education Review*, **21**, 183-195. <https://doi.org/10.1007/s12564-019-09623-y>
- [3] 刘秀梅, 古明, 国红延, 等. 中学教师在教学测评中的伦理困境研究[J]. 中国教育学刊, 2022(1): 86-91.
- [4] 顾景倩. 浅谈 GenAI 技术在大学英语教学中的应用——以 Spark Desk (讯飞星火)为例[J]. 英语广场, 2024(17): 113-116.
- [5] 徐林林, 胡杰辉, 苏扬. 人工智能辅助学术英语写作的学习者认知及行为研究[J]. 外语界, 2024(3): 78-85.
- [6] 孙婷婷. AI 时代外语教学数智化转型的嬗变与跃迁——基于 CiteSpace 可视化图谱分析[J]. 湖北开放职业学院学报, 2025, 38(13): 157-160.
- [7] 王佑镁, 王欣颖, 柳晨晨. 教育领域 GenAI 应用的伦理风险管理框架研究[J]. 电化教育研究, 2024, 45(10): 28-34+42.
- [8] 谢娟. 人工智能与教育融合创新何以“伦理先行”——兼论 GenAI 教育应用的伦理路径[J]. 现代远程教育研究, 2024, 36(6): 11-19.

- [9] 王佑镁, 王旦, 王海洁, 柳晨晨. 算法公平: 教育人工智能算法偏见的逻辑与治理[J]. 开放教育研究, 2023, 29(5): 37-46.
- [10] Messick, S. (1988) Meaning and Values in Test Validation: The Science and Ethics of Assessment. *ETS Research Report Series*, 1988, i-28. <https://doi.org/10.1002/j.2330-8516.1988.tb00303.x>
- [11] 刘艳红. GenAI 的三大安全风险及法律规制: 以 ChatGPT 为例[J]. 东方法学, 2023(4): 29-43.
- [12] 古明, 刘秀梅, 国红延. 中学英语教师对课堂测评活动中伦理问题的认知水平研究[J]. 基础外语教育, 2024, 26(6): 91-99+111.

附 录

生成式人工智能赋能 EFL 课堂教学测评师生公平性感知调查

情景一: 在《英汉口译》课程期末测评中, 教师允许学生使用 GenAI 工具辅助完成实时口译任务, 学生也可选择独立完成口译任务。为保证成绩可比性, 教师对 GenAI 辅助翻译作品与学生独立翻译作品采用同一评分标准进行评判。请问该教师做法是否合适?

情景二: 在《英语写作》课程教学中, 教师要求学生提交一篇议论文作为平时作业, 允许学生使用 GenAI 进行语言润色, 得分按一定权重计入课程成绩。学生 A 使用付费 GenAI 写作工具润色语言, 得分较高; 学生 B 因经济原因只能使用免费版, 得分较低。考虑到评价标准的统一性, 教师在评分过程中未考虑不同学生对 GenAI 工具的可利用程度。请问该教师做法是否合适?