

AIGC 赋能大学英语口语“教-学-评”一体化模式构建与实证研究

宋西雅, 陈芳芳

温州商学院, 浙江温州, 中国

【摘要】立足传统大学英语口语教学中跨时空交互壁垒、真实语境匮乏及过程性评价滞后等多重结构性困境, 基于互动假说、输出假说与情感过滤理论, 构建以生成式人工智能为交互核心、以智慧教学云平台为管理载体的“教—学—评”一体化教学模式。该模式旨在通过定制分级教学语料、深化人机协同探究机制及建构即时增值评价闭环, 消除传统教学场域中教、学、评相割裂的结构性壁垒。为验证模式有效性, 采用准实验设计范式, 以 65 名非英语专业本科生为被试样本, 实施为期 16 周的教学干预。T 检验结果显示, 实验组在口语流利度与准确度维度上均显著优于对照组, 且流利度提升效应最为明显; 回归分析进一步证实, 大模型提供的即时评价反馈有效性对学生的人机互动投入度具有显著正向预测作用。研究表明, 该模式能有效降低口语焦虑, 通过“以评促学”机制显著提升学生口语综合应用能力, 为人工智能时代口语教学改革提供实证依据。

【关键词】生成式人工智能 (AIGC); 大学英语口语; “教—学—评”一体化; 人机协同

【基金项目】温州商学院课堂教学模式改革项目 (2024jxmg22)

1. 引言

随着全球教育数字化改革深入实施, 生成式人工智能 (AIGC) 技术的加速迭代, 正改变高等教育数字化发展生态与实施路径。AIGC 凭借其强大自然语言理解、多模态内容生成及智能交互能力, 正在重塑外语教学的底层逻辑与生态系统。长期以来, 大学英语口语教学面临“高耗低效”困境, 传统课堂多采取“教师讲授+学生被动操练”模式, 语言输入与输出严重失衡, 学生缺乏真实语境交互, 导致“哑巴英语”现象频发。尤为关键的是, 教学评价往往滞后于学习过程, 教师难以在大班授课环境下针对每位学生提供即时、个性化反馈, 致使“教、学、评”三环节相互割裂, 严重制约学生口语应用能力提升[1]。

“教—学—评”一体化理念强调教学目标、学习活动与评价反馈的一致性, 主张评价应贯穿学习全过程[2]。AIGC 技术的介入为实现这一理念提供关键支撑, 它不仅能生成分级教学资源以辅助精准教学, 更能作为全天候智能语伴提供沉浸式交互环境, 并基于语音识别技术给予秒级诊断反馈, 从而构建起“以评促学、以评促教”的教学闭环。基于此, 旨在构建 AIGC 赋能下大学英语口语“教—学—评”一体化教学模式, 并通过准实验研究及数据量化分析, 验证该模式在提升学生口语流利度、准确度及复杂度方面的实效, 探究其内在作用机制, 以期为人工智能时代大学

英语教学改革提供实证依据与实践路径。

2. 理论基础与模式构建

2.1 理论基础

构建的 AIGC 赋能口语教学模式主要依托于“认知—情感—交互”三维一体理论支撑体系, 具体包含以下三大核心理论:

第一, 互动假说与认知交互。单纯的可理解性输入并非二语习得的充分条件, 语言习得发生于有意义的互动中[3]。传统大班授课的互动往往缺乏真实信息差, 而 AIGC 能够基于上下文提供高质量、具有不可预测性的“意义协商”。AIGC 根据学生语言水平动态调整对话难度, 创造高频交互环境, 使学生在持续的意义协商中习得规则。

第二, 输出假说与形式加工。Swain 强调“强制性输出”的关键作用, 指出产出语言迫使学习者从单纯的“语义加工”转向句法层面的“形式加工”[4]。该研究模式中, AIGC 的即时纠错机制能够通过“强制性重述”环节, 促使学生对自身产生的中介语漏洞进行“有意识的注意”, 实现从显性知识向隐性知识的转化。

第三, 情感过滤假说与低焦虑阈值。Krashen 认为, 学习者的情感状态直接影响语言输入的吸收[5]。AIGC 的“非评判性”特质有效剥离社会评价带来的“面子焦虑”, 从而大幅降低情感过滤, 为口语输出扫清心理障碍。

此外, 尽管 AIGC 在现阶段无法完全复现真实人际交往中复杂的情感纠葛与社会文化场域, 但其在提供高频次、高度可控、极度容错的“拟态”交际环境方面, 恰好弥补了传统课堂的局限。并非主张人机互动可以完全替代人际交往, 而是将 AIGC 定位为“智能认知载体”, 为学习者提供一个从内部心理演练向外部真实交际过渡的关键缓冲带。

2.2 “教—学—评”一体化教学模式设计

构建的“教—学—评”一体化模式旨在利用 AIGC 技术打破传统教学时空壁垒, 以国产大语言模型“豆包”作为智能生成与交互核心, 以“学习通”作为教学管理与数据留存载体, 实现教学资源动态生成、学习过程深度交互及评价反馈即时精准, 如图 1 所示。

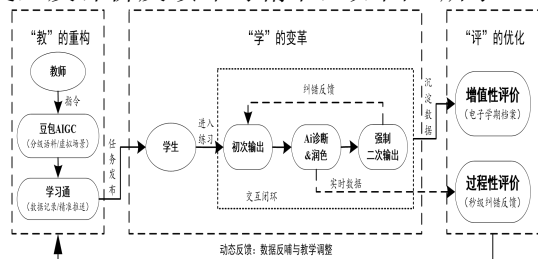


图 1. “教-学-评”一体化教学模式图

一是“教”的重构, 注重资源生成与场景创设。AIGC 技术重构教师备课与资源供给方式。针对传统教材内容固化、更新滞后痛点, 教师利用豆包 APP 依据学生语言水平产出分级、个性化口语语料, 并通过学习通平台精准推送至学生端, 实现教学内容定制化。同时, 借助豆包情境构建能力, 创设雅思口语、职场面试、跨文化交流等多元虚拟仿真场景, 有效解决课堂教学真实语境缺失问题, 增强语言输入真实性与丰富度。

二是“学”的变革, 强调人机协同与个性化训练。学习方式由被动接受转向人机协同探究。一方面, 通过人机对话机制, 学生利用豆包“口语陪练”智能体开展全天候、开放式口语交流, 在无社会评价压力的低阈值环境中有效克服“开口焦虑”, 提升交际意愿。另一方面, 建立“自主纠错”深度学习机制, 学生依据豆包提供的润色建议及优化表达, 对原始口语产出进行修正并完成二次输出, 促使语言训练从单纯“流利度”操练向“准确性”与“复杂性”提升转变。

三是“评”的优化, 突出即时反馈与数据驱动。评价体系实现从结果导向向过程导向转型。在过程性评价层面, 依托豆包内置自动语音识别与自然语言处理技术, 实时捕捉

发音偏差、语法谬误及逻辑断层, 提供秒级诊断反馈, 确保纠错时效性。在增值性评价层面, 整合豆包生成的对话反馈记录与学习通教学管理数据, 建立电子学情档案, 全周期追踪记录学生从期初至期末词汇密度、句法复杂度等关键指标变化, 绘制能力增值曲线, 为精准教学决策提供数据支撑。

3. 研究设计与实施

3.1 实验对象

为验证前文构建的“豆包+学习通”一体化 AIGC 赋能教学模式实效, 选取某高校非英语专业大二年级两个平行自然班作为研究对象, 共计 65 人。其中, 设定 A 班为实验组 (N=32), B 班为对照组 (N=33), 两组学生在年龄、性别比例及教育背景方面基本一致。实验开展前, 统一组织口语水平前测, 测试题型及评分标准对标雅思口语考试。利用 SPSS 软件对前测成绩进行独立样本 T 检验, 结果显示两组学生口语总分及分项成绩差异均无统计学意义 ($p>0.05$), 表明两组受试者英语口语起始水平处于同一基线, 满足实验同质性要求。

3.2 实验过程

实验周期为一学期 (16 周), 两组均由同一教师授课。对照组沿用传统“教师讲授+小组讨论+课后录音”模式; 实验组实施 AIGC 赋能的闭环教学模式。为最大程度剥离教师期望偏倚以及霍桑效应对实验结果造成的潜在干扰, 实施严格变量控制策略。两组受试者的课内外总计干预学时、核心教学目标及考核任务量保持一致。实验组的 AIGC 辅助训练被精准定位为课外人机协同演练, 而该教师在课堂实体教学中, 对两组学生投入的指导时长与情感支持均保持严格一致。通过“课内同质、课外异构”的设计, 确保口语能力的差异主要源于技术赋能。

3.3 数据来源

遵循多源数据三角互证原则, 采用量化测试与问卷调查相结合方式, 全面评估教学模式有效性。

(1) 量化指标: 口语能力多维测评

选取雅思口语真题作为素材, 评分标准依据 CAF 理论框架[6], 即复杂度、准确度、流利度, 构建“智能测评+人工复核”评分机制。此外, 为排除“学习时间投入”的混淆变量干扰, 通过后台日志比对显示: 实验组每周课外平均在线人机交互时长为 2.5 小时, 对照组每周生生互动及作业录音折合平均时长约

为 2.4 小时。经检验，两组课外练习时间投入无显著统计学差异 ($p>0.05$)。

(2) 调查问卷: AIGC 辅助口语学习体验量表

问卷设计基于技术接受模型(TAM)[7], 重点考察学生对 AIGC 技术在口语教学中的“感知有用性”和“感知易用性”, 并参考 Dizon (2020) 关于智能语音助手对二语口语发展影响的研究[8], 借鉴其关于“口语交际焦虑”、“学习自信心”及“人机协同教学效果”的测量维度, 结合“教-学-评”一体化模式特点进行本土化改编与操作化定义。问卷共包含 15 个题项, 采用李克特五级计分制(1=完全不同意, 5=完全同意)。问卷结构严格对应“教-学-评”三个核心环节, 旨在精准捕捉 AIGC 赋能下的全流程学习体验。

一是教学资源满意度, 主要考察 AIGC 在资源生成与场景创设方面的效能。借鉴姜军(2024)关于“模拟场景精细度”与“口语内容贴合度”的指标[9], 具体设置 4 个题项, 重点测量学生对 AIGC 生成的分级语料是否符合自身水平, 虚拟职场或生活场景是否具有真实感与沉浸感, 以及教学内容是否满足个性化需求。

二是人机互动投入度, 聚焦于学习过程中的心理状态与行为投入。参考吴玲娟(2025)的研究发现“人机互动能有效降低社会评价压力从而缓解焦虑”, 具体设置 6 个题项。重点考察学生在与 AI 语伴对话时的“低焦虑体验”、在无压力环境下的“交际意愿”提升程度, 以及在人机协同任务中的持续专注度。

三是评价反馈有效性, 旨在评估 AI 智能评价对学习效果的促进作用。基于输出驱动假设, 设置 5 个题项, 重点考察 AI 提供的秒级纠错对学生进行“二次修正”的帮助程度, 以及基于“豆包+学习通”生成的电子学情档案对激发学生持续学习动机的感知有效性。

4.数据分析与结果

4.1 数据收集情况与描述性统计分析

于实验周期结束后(第 16 周), 向实验组学生发放《AIGC 赋能口语“教—学—评”体验调查问卷》。问卷通过“问卷星”平台以匿名形式在线发布, 共发放问卷 32 份。剔除答题时间过短及规律性作答的无效问卷后, 最终回收有效问卷 30 份, 有效回收率为 93.75%。数据样本量满足统计学分析的基本要求。运用 SPSS 26.0 对问卷各维度均值与标准差进行描述性统计分析, 结果见表 1。

表 1.问卷各维度描述性统计结果 (N=30)

变量维度	题项数	均值 (M)	标准差 (SD)	描述
教学资源满意度	4	4.21	0.65	学生对 AIGC 生成的个性化语料认可度较高
人机互动投入度	6	4.35	0.58	学生在无压力环境下的参与意愿最强
评价反馈有效性	5	4.18	0.72	即时反馈对纠错具有明显的促进作用
问卷总体	15	4.25	0.62	总体处于“比较满意”至“非常满意”水平

4.2 问卷信效度检验

为确保测量工具的科学性与数据分析结果的可靠性, 利用 SPSS 26.0 统计软件对回收的 30 份有效问卷进行信度与效度检验, 结果如表 2 所示。

考虑到本准实验研究属于定点干预, 受试样本量具有客观局限性(N=30), 在统计学意义上尚不具备开展大规模探索性因子分析的最佳样本容量条件。因此, 为避免小样本带来的统计偏差, 主要聚焦于问卷内容效度与信度检验。在内容效度层面, 本量表的题项生成并非主观臆断, 而是严格依托于技术接受模型理论框架, 并深度借鉴国际权威期刊发表的成熟量表, 如 Dizon 关于智能语音助手的测量工具, 结合高校外语教学场域特征进行多轮专家审阅与本土化词句适配, 从而确保测量工具能精准捕捉研究预设的核心构念。在信度层面, 运用 SPSS 26.0 对有效问卷数据进行内部一致性检验。分析表明, 总量表的 Cronbach’s α 系数达到 0.912, 教学资源满意度、人机互动投入度及评价反馈有效性三个子维度的 α 系数均处在 0.845 至 0.876 的优良区间内, 说明该测量工具的各项题项具有极高的内部稳定性与逻辑一致性, 获取的数据真实可靠, 完全能够为后续实证分析提供定量支撑。

表 2.问卷信度检验结果 (N=30)

变量维度	题项数	Cronbach’s α
教学资源满意度	4	0.845
人机互动投入度	6	0.876
评价反馈有效性	5	0.853
问卷总体	15	0.912

4.3 教学效果对比分析

利用 SPSS 26.0 对实验组与对照组口语测试成绩进行独立样本 T 检验, 以验证“教—

学—评”一体化模式对提升学生口语能力的实效。

(1) 实验前测数据分析

为确保实验结果的信度，首先对两组学生前测成绩进行差异性检验。如表 3 所示，经独立样本 T 检验，两组学生在口语总分及

各项子维度的显著性概率均满足 $p>0.05$ ，且各项效应量 Cohen's d 值均小于 0.2（属微小效应）。统计结果表明，实验前两组受试者的英语口语水平处于同一基线，满足准实验的同质性前提。

表 3.实验前两组口语前测成绩独立样本 T 检验结果

测评项目	组别	人数 (N)	均值 (Mean)	标准差 (SD)	t 值	p 值	Cohen's d
口语总分 (满分 100)	实验组	32	68.45	5.21	0.535	0.595	0.13
	对照组	33	69.12	4.89			
流利度 (满分 30)	实验组	32	20.15	2.15	0.400	0.690	0.10
	对照组	33	20.36	2.08			
准确度 (满分 40)	实验组	32	24.50	3.12	0.287	0.775	0.07
	对照组	33	24.72	3.05			
复杂度 (满分 30)	实验组	32	23.80	2.85	0.646	0.521	0.16
	对照组	33	24.24	2.64			

(2) 实验后测数据分析

经过一学期教学实验，两组学生口语后测成绩对比情况如表 4 所示。数据表明，实验组口语后测总分 ($M=79.82$) 不仅在统计学意义上极显著高于对照组 ($t=6.211$, $p<0.001$)，更值得注意的是，其效应量 Cohen's

d 值高达 1.54。按照统计学基准 ($d>0.8$ 即为大效应)，这一数据远超常规教学干预的平均影响幅度，证明 AIGC 赋能的“教-学-评”一体化模式在大学英语口语教学中具有极高实际应用价值与颠覆性的效能。各分项指标分析如下：

表 4.实验后两组口语后测成绩独立样本 T 检验结果

测评项目	组别	人数 (N)	均值 (Mean)	标准差 (SD)	t 值	p 值	Cohen's d
口语总分 (满分 100)	实验组	32	79.82	4.55	6.211	<0.001	1.54
	对照组	33	72.35	5.12			
流利度 (满分 30)	实验组	32	25.40	1.85	7.728	<0.001	1.92
	对照组	33	21.55	2.15			
准确度 (满分 40)	实验组	32	28.95	2.65	2.765	0.007	0.69
	对照组	33	27.05	2.88			
复杂度 (满分 30)	实验组	32	25.47	2.15	2.508	0.015	0.62
	对照组	33	23.75	3.25			

注：Cohen's $d\approx 0.2$ 为小效应，0.5 为中等效应，0.8 为大效应。

第一，流利度呈现出超大效应跃升 ($d=1.92$)。实验组得分取得突破性增长，在量化层面印证人机协同学习环境的核心优势。豆包提供的全天候、低焦虑交互打破传统课堂的“哑巴英语”困境，高密度的语言输出频率促使学生实现从“不敢说”到“脱口而出”的转变。

第二，准确度取得中高强度实质性改善 ($d=0.69$, $p=0.007$)。该结果有效反驳部分传统观点对“AI 对话仅能提升流利度而忽视规范性”的担忧。数据证实，本模式中建立的“即时纠错—强制重述”评价闭环发挥关键的“脚手架”作用。深度的错误加工与二次输出，有效驱动学生显性语法知识向隐性自动化表

达的转化。

第三，复杂度获得中等强度稳步提升 ($d=0.62$, $p=0.015$)。尽管其提升幅度略逊于流利度，但也达到中等偏上的效应规模。这表明 AIGC 生成的丰富语境与高级词汇确实在逐渐拓宽学生的语料边界，但由于句法结构的重构往往需要经历更漫长内化周期，未来仍需依赖更为长效的追踪干预以实现进一步突破。

4.4“教—学—评”机制影响路径分析

为深入探究一体化模式中“评”对“学”的内在驱动机理，采用线性回归验证 AIGC 智能评价反馈对学生学习行为意愿的预测力。

(1) 变量定义与模型构建

依据“以评促学”及情感过滤理论，将问卷中“评价反馈有效性”维度的均值设定为自变量（X），代表学生对 AI 即时诊断与增值评价的感知有用性；将“人机互动投入度”维度的均值设定为因变量（Y），代表学生在无压力环境下的交际意愿与练习持续性。

(2) 回归模型检验

表 5.模型摘要

模型	R	R ²	调整后 R ²	标准估计误差	Durbin-Watson
1	0.736	0.542	0.525	0.400	1.924

表 6.回归系数分析

模型	非标准化系数 (B)	标准误差	标准化系数 (β)	t 值	Sig.
(常量)	1.872	0.437	-	4.284	<0.001
评价反馈有效性	0.593	0.103	0.736	5.757	<0.001

注：因变量为人机互动投入度；Sig. <0.001。

利用 SPSS 26.0 进行回归运算，模型拟合度及系数分析结果如表 5、表 6 所示。模型判定系数 R^2 为 0.542，调整后 R^2 为 0.525，表明“评价反馈有效性”能够解释“人机互动投入度”变异的 54.2%，模型拟合优度良好。此外，Durbin-Watson 统计量为 1.924，表明样本残差独立，不存在序列相关性，模型构建有效。系数检验结果显示，回归方程的显著性概率 $p<0.001$ ，非标准化回归系数 B 为 0.593，标准化回归系数 β 为 0.736，这表明“评价反馈有效性”对“人机互动投入度”具有极其显著的正向预测作用。

5.讨论

基于数据分析结果，深入剖析 AIGC 赋能“教—学—评”一体化模式在大学英语口语教学中的有效性及其内在驱动机理。

5.1 人机协同赋能：口语流利度与准确度的双维跃升

数据表明，AIGC 干预对口语能力的提升并非均是均质的，而是存在两条清晰的差异化作用路径。一方面，流利度的超大效应跃升源于“低情感过滤”对交际意愿的彻底释放。互动假说指出，足量的语言试错是习得发生的关键。传统大班授课环境极易让学生产生“面子焦虑”，导致产出机会匮乏；而豆包 APP 创设全天候、非评判性的对话环境，有效剥离社会评价压力，使学生敢于不断试错。交

际意愿的大幅提升直接加速语言知识从“陈述性”向“程序性”的自动化转化，这是流利度取得突破的根本原因。另一方面，准确度的实质性改善得益于“强制性二次输出”引发的深度注意。与吴坚豪（2024）关于“AI 对口语准确度提升不明显”的结论不同[10]，单纯的人机自由聊天确实易导致语言石化，但“纠错—重述”机制能有效破解这一困局。根据输出假说，依托“学习通”强化的靶向训练，迫使学生对 AI 的反馈产生“有意识的注意”。这种必须修正并再次输出的压力，促使学生在反复的交互中深度内化句法规则，最终实现流利与准确的协同发展。

5.2 智能评价驱动：学习投入与以评促学的深度耦合

回归分析明确回答智能评价为何能有效解决传统评价滞后与单一的痛点，并发挥强大的前置驱动作用。一方面，秒级诊断缩短“认知修正时延”，构筑消解焦虑的情感安全网。传统期末测试的滞后性使评价失去了纠偏价值，而 AIGC 的即时反馈让学生在产出语言的瞬间就能对比自身中介语与目标语的差距，这种“即错即改”的体验不仅没有引发防御性心理，反而作为一种积极的情感支架，能有效降低外语焦虑。另一方面，增值性评价激发持续的成就动机，实质性打通生态闭环。在一体化模式中，“评”有效转变传统教学中作为“终结性审判”的滞后角色，转变为驱动新一轮“学”的认知锚点。实证数据表明，正是由于评价机制实现从“延时、粗放”向“即时、精准、可视化”的跃升，才促使学生在“学”的环节中投入更高专注度，促成口语学习从“被动应试”向“主动意义建构”蜕变。

6.结语

通过准实验设计，构建并验证基于“豆包+学习通”的 AIGC 赋能大学英语口语“教—学—评”一体化教学模式。研究发现：第一，该模式能显著提升大学生的英语口语水平，尤其在流利度与准确度维度效果显著；第二，AIGC 提供的即时智能反馈与增值评价是驱动学生投入深度学习、降低口语焦虑的关键因素，有效打破传统口语教学中“哑巴英语”与“反馈缺失”困境，研究结果对外语教学改革具有重要启示。人工智能时代，教师的角色需从单纯的知识传授者转型为人机协同的设计者或数据驱动的分析师。在“教”的层面，教师不应被动依赖技术，而应利用 AIGC 生成个性化、分级化的教学语料，重构教学内

容生态；在“学”的层面，应引导学生建立人机协同学习观，利用 AI 作为基础进行深度探究，而非简单的答案索取；在“评”的层面，教师应充分挖掘“教—学—评”一体化平台的过程性数据，将评价重心从“对学习的评价”转移至“为了学习的评价”，利用可视化数据精准干预教学痛点。

尽管研究证实该模式的有效性，但仍存在一定局限。一是受限于准实验设计，样本量相对较小且来源于同一所高校，结论的普适性有待扩大样本范围进一步检验；二是实验周期为一学期，关于 AIGC 干预对口语能力的长期保持效应尚需纵向追踪研究。未来研究可进一步关注不同语言基础的学习者在人机协同环境下的学习异质性。

参考文献

- [1]黎奇,黎芳,雷野.生成式人工智能赋能外语教学:场景、困境与对策[J/OL].长沙大学学报, 1-10.
- [2]王蕾,李亮.推动核心素养背景下英语课堂教—学—评一体化:意义、理论与方法[J].课程·教材·教法, 2019, 39(5): 114-120.
- [3]LONG M H. Native speaker/non-native speaker conversation and the negotiation of comprehensible input[J]. Applied

Linguistics, 1983, 4(2): 126-141.

- [4]SWAIN M, LAPKIN S. Problems in output and the cognitive processes they generate: A step towards second language learning[J]. Applied Linguistics, 1995, 16(3): 371-391.
- [5]KRASHEN S. Some issues relating to the monitor model[J]. TESOL Quarterly, 1977, 11(3): 309-332.
- [6]HOUSEN A, KUIKEN F. Complexity, accuracy, and fluency in second language acquisition[J]. Applied Linguistics, 2009, 30(4): 461-473.
- [7]DAVIS F D. Perceived usefulness, perceived ease of use, and user acceptance of information technology[J]. MIS Quarterly, 1989, 13(3): 319-340.
- [8]DIZON G. Evaluating intelligent personal assistants for L2 listening and speaking development[J]. Language Learning & Technology, 2020, 24(1): 16-26.
- [9]姜军.AI 赋能英语口语教学路径的思考[J].内蒙古财经大学学报, 2024, 22 (06): 72-76.
- [10]吴坚豪,周婉婷,曹超.生成式人工智能技术赋能口语教学的实证研究[J].中国电化教育, 2024 (04): 105-111.