

从技术依赖到批判性内化：POA框架下教师主导的人机协同口语教学研究

陈霞

英语学院, 浙江越秀外国语学院, 中国·浙江 绍兴 312000

摘要: 针对全球化背景下高级英语口语教学反馈不足与资源静态化问题, 本研究基于产出导向法 (POA) 理论框架, 探究 POA 视域下 AI 生成式支架对高级英语口语能力的赋能机制。通过 16 周行动研究发现, AI 生成式支架可有效协同提升学习者口语复杂度、准确度与流利度, 推动学生从被动采纳转向批判性内化。教师通过设计深度句法重构、文化转译策略库与批判性交互任务单, 将 AI 从“答案提供者”转化为“认知催化剂”, 构建了“教师—AI—学生”三元协同教学模式, 为 POA 在智能技术环境下的理论创新与口语教学实践提供了实证依据。

关键词: 产出导向法; AI 生成式支架; 高级英语口语; 复杂度准确度流利度 (CAF); 行动研究

From Technical Dependence to Critical Internalization: A Study on Teacher-Led Human-Computer Collaborative Speaking Teaching under the POA Framework

Chen Xia

School of English Studies, Zhejiang Yuexiu University, China Zhejiang Shaoxing 312000

Abstract: Against the backdrop of globalization, advanced English speaking competence has become a core quality for internationally competitive talents. Addressing the issues of insufficient feedback and static resources in current college-level oral English instruction, this study draws upon the theoretical framework of the Production-Oriented Approach (POA) to investigate the empowering mechanism of POA-oriented AI-generated scaffolding on advanced English speaking ability. Through a 16-week action research design, the findings demonstrate that AI-generated scaffolding, guided by the POA principles of precision, gradation, and diversity, effectively facilitates the synergistic development of learners' syntactic complexity, accuracy, and fluency (CAF), while shifting students from passive acceptance of AI output toward critical internalization and strategic expression. By designing deep syntactic reconstruction workshops, a cultural transcreation strategy bank, and critical AI interaction task sheets, the teacher successfully transformed AI from an “answer provider” into a “cognitive catalyst,” thereby constructing a triadic “teacher-AI-student” collaborative teaching model. This study provides empirical evidence for the theoretical innovation of POA within intelligent technology environments and offers practical implications for oral English instruction.

Keywords: Production-oriented approach (POA); AI-generated scaffolding; Advanced English oral proficiency; Complexity, accuracy, fluency (CAF); Action research

0 引言

在全球化进程加速的背景下, 高级英语口语能力已成为衡量国际化人才的关键指标, 不仅要求语言形式的准确与流利, 更强调表达的复杂性 with 思想深度。然而, 当前高校口语教学面临深层挑战, 大班制使教师难以为学生提供持续、个性化的高阶反馈, 导致学生口语输出长期徘徊于“达意”层面; 静态、统一的教学资源无法适配学生个体差异, 制约了语言复杂度与交际效能的协同发展。

生成式人工智能的崛起为破解上述困境提供了新的技术路径。然而, 若缺乏科学的教学理论引领, 技术应用

易陷入“工具理性”窠臼, 或沦为脱离语境的机械操练, 或与课程目标脱节。在此背景下, 根植于中国教育情境的“产出导向法”(POA), 凭借其“学习中心”“学用一体”与“全人教育”的核心原则(文秋芳, 2015), 为构建“教师主导、技术赋能”的教学模式提供了理论框架。POA 所倡导的“精准性、渐进性、多元性”促成理念, 与生成式 AI 的动态响应和个性适配能力高度契合。然而, 如何将这一理论框架与技术潜能转化为可复制、可推广的实证方案, 仍是当前领域亟待探索的前沿议题。

基于此, 本研究聚焦于 POA 视域下 AI 生成式支架

对高级英语口语能力的赋能机制，致力于探究以下核心问题：第一，该支架能否有效协同提升学习者口语产出的复杂度、准确度与流利度（CAF 三维度）？第二，教师的干预策略如何优化支架的教学效能？第三，学习者对该融合模式的教学感知与接受度如何？通过对这些问题的系统考察，本研究旨在推动 POA 在智能技术环境下的理论创新，构建以教师设计为核心、以 AI 为赋能工具、以学生能力发展为主体的三元协同教学模式，为数字化转型时期的口语教学改革提供实证依据。

1 文献综述

生成式人工智能正深刻重塑二语口语教学生态。Du 与 Daniel（2024）对 AI 聊天机器人用于 EFL 口语教学的系统综述表明，AI 工具通过可及性、重复性与低焦虑环境，在提升流利度与交际信心方面成效显著。国内研究亦证实，以 ChatGPT 为代表的生成式 AI 在丰富学习者词汇多样性、拓展句法结构方面展现出潜力（吴坚豪等，2024；葛楠、郑春萍，2025；孔蕾，2026，王华，2026）。然而，技术的增益呈现维度分化：吴坚豪等（2024）发现 AIGC 辅助下学习者的口语复杂度与流利度有所提升，但准确度并未同步发展，呈现出 CAF 三维度的失衡态势。徐锦芬、陈聪（2018）进一步提示，若无教学层面的认知规约，单纯增加“技术复杂度”可能加剧 CAF 子系统间的资源竞争。于涵静、戴炜栋（2019）基于复杂动态系统理论将其解读为“复杂性—准确性在发展初期呈竞争性生长”，为“跛脚鸭”效应提供了本体论解释。这些发现共同指向一个核心问题：AI 的“赋能”并不自动等同于“增效”，其教学价值的释放依赖于恰当的规约设计。

当前主流的“学生自主调用”模式虽提升练习灵活性，却常与系统教学目标脱节。Wang 等（2024）的三臂对照实验显示，人机协同下的口语训练虽能显著提升意愿沟通（WTC）与降低焦虑，但口语测试成绩本身未现组间显著差异——这一“情感增益稳、产出增益弱”的分化，恰折射出自主调用模式下 AI 对话易流于“可说”却难达“说得好”的局限，对文化负载词、语境深度的阐释尤显表层（吴坚豪等，2024；邹娟娟，2025）。情感维度上，Hong 与 Lee（2026）的分级反馈研究亦提示，焦虑降低仅出现在新手组，中高级学习者从“随意用 AI”中获得的情感红利有限。上述线索共同表明，教师主导的支架设计是调节技术边界的关键，亦是人机协同从理念走向实证的核心缺口。产出导向法的“精准性、渐进性、多元性”促成理念（文秋芳，2015；邱琳，2020）为规约 AI 使用提供了

可操作框架，其与“学习者—机器—教师”三维协同（邹娟娟，2025）及人机互动协商能力 HAINC（文秋芳、梁茂成，2024）的衔接，也为人机协同的“教师主导”立场提供理论支撑。但既有“教师—AI—学生”协同探讨缺乏在高阶课堂中系统检验教师主导的 AI 支架如何协同驱动 CAF 三维度发展的实证证据，尤其在 POA 规约下，如何将“句法薄弱、文化转译生硬、AI 交互浅表化”等高级口语瓶颈打通，仍是待填空白。

2 理论框架与研究设计

2.1 理论框架

生成式 AI 在口语教学中的应用面临双重困境：学生易形成“技术依赖”而导致口语表达有流利无深度，教师角色则在缺乏教学规约的情境下趋于边缘化。POA 的“精准性、渐进性、多元性”促成理念为化解这一矛盾提供了理论锚点。据此，本研究对“POA 视域下 AI 生成式支架”作出如下操作性界定：它是指教师为达成特定产出目标、基于 POA 促成原则设计的，引导学生与 AI 进行批判性交互的结构化任务与策略总和，其核心特征为教师主导设计、批判性交互与过程性迭代，将 AI 定位为“认知催化剂”而非“答案提供者”。在此基础上，本研究进一步探究该支架对高级英语口语能力的赋能机制。

2.2 研究范式与整体设计

本研究采用“基于设计的研究（DBR）+ 准实验比较”的复合研究范式。DBR 框架指导实验组的两轮迭代教学改进，准实验设计通过设置对照组来分离 AI 生成式支架的净效应。研究共持续 16 周，分为两轮迭代（第一轮：第 1—3 周；第二轮：第 4—16 周）。实验组（n=30）接受 POA 视域下 AI 生成式支架教学，对照组（n=30）接受传统 POA 口语教学（相同产出任务与课时，不使用 AI 支架）。两组学生均为英语专业大三学生，入学英语成绩无显著差异（ $t=0.43$, $p>0.05$ ）。

为清晰呈现两轮干预的效果演变，本研究设置三个数据采集时间点：第一轮开始前（第 1 周）进行前测（T1），采集两组口语测试数据作为基线；第一轮结束后（第 3 周末）进行中测（T2），采集两组口语测试及实验组 AI 交互日志，用于效果评估与问题诊断；第二轮结束后（第 16 周末）进行后测（T3），采集两组口语测试、实验组 AI 交互日志及刺激性回忆访谈，用于整体效果检验。

2.3 两轮干预的迭代逻辑

第一轮干预（第 1—3 周）围绕 POA 闭环展开，教师构建了与产出任务“实地讲解：走进龙井茶村”高度契合

的 AI 语言支架库。第一轮结束后,通过中测(T2)口语语料的 CAF 分析及 AI 交互日志编码,诊断出三项核心问题:句法能力系统性薄弱(平均 T 单位长度 8.45 ± 1.23 ,从句/T 单位比率 0.38 ± 0.08)、AI 交互认知层次浅表化(词汇替换类建议采纳率 78% 但词汇密度无显著提升)、跨文化转译策略性知识缺失(文化负载词解释类支架采纳率低于 30%)。第二轮干预(第 4—16 周)基于上述诊断,重点引入深度句法重构工作坊、文化转译策略库、批判性 AI 交互任务单三项改进策略,教学重心从“促进 AI 使用”转向“引导批判性内化”。

2.4 数据分析框架

本研究的数据分析遵循两条并行路径。在整体效果检验层面,以 T1 和 T3 的两组口语语料为数据来源,采用独立样本 t 检验比较两组在 CAF 各指标上的前后测增量差异,相应结果呈现在表 1 与表 2 中。在迭代效果追踪层面,以实验组 T1、T2、T3 三个时间点的数据为来源,比较第一轮(T1→T2)与第二轮(T2→T3)在 AI 交互质量与认知策略运用上的变化,相应结果呈现在表 3 中。两条路径互为补充,共同构成本研究从效果验证到机制解释的完整证据链。

3 教学目标设定与促成活动设计

遵循 POA 以产出任务为依据设定教学目标的原则,本研究基于核心任务“实地讲解:走进龙井茶村”,设定了交际与语言双重教学目标。交际目标包括三个递进层次:鉴别与描述龙井茶村核心要素、阐释其文化价值并与杭州茶文化传统关联、对不同类型游客提出个性化游览建议。语言目标涵盖词汇(茶文化主题学术词汇)、句法(定语从句、状语从句、非谓动词等复杂句式)、语篇(逻辑连接词构建完整语篇结构)与策略(功能对等、情境类比等文化转译策略)四个维度。

在明确上述分层目标的基础上,本研究依据 POA“精准性、渐进性、多元性”原则规划了系统性的促成活动体系。精准性要求 AI 支架直指学生表达瓶颈:针对句法能力薄弱问题,设计“深度句法重构工作坊”和“AI 生成句式升级库”;针对文化转译生硬问题,构建“文化转译策略库”,提供功能对等与诗意象等修辞取向的 AI 提示模板。渐进性强调将复杂产出任务分解为“描述事实→阐释观点→比较文化→提出建议”的序列化子任务,为每个阶段配备高度适配的输入材料。多元性通过整合多模态资源实现,AI 生成式支架超越文本提示,整合语音示范与逻辑可视化提示;同时构建“教师课堂引导—AI 个性化训练—

小组协作探究—个人批判性反思”的互动路径,其中“批判性 AI 交互任务单”通过要求学生记录互动过程与修改理由,将隐性认知过程显性化,推动学生从被动采纳转向批判性内化。

4 两轮干预:从“脚手架”到“认知引擎”的迭代

本研究在高级英语口语课程中开展了两学期、两轮的行动研究,以“实地讲解:走进龙井茶村(A Tour Guide to Chinese Longjing Tea Village)”为核心产出任务。第一轮侧重搭建基础语言支架,初步探索 AI 的应用效能;第二轮则基于诊断出的问题进行策略性重构,推动 AI 从“脚手架”向“认知引擎”转化。

本节的数据呈现遵循两条并行线索。第一条线索聚焦整体效果检验:以第一轮开始前的口语测试(T1)和第二轮结束后的口语测试(T3)为时间节点,通过表 1 和表 2 报告实验组与对照组在 CAF 三维度上的前后测差异,回答“AI 生成式支架是否有效”的问题。第二条线索聚焦迭代过程追踪:以第一轮结束后(T2)和第二轮结束后(T3)为时间节点,通过表 3 报告实验组在 AI 交互质量与认知策略运用上的变化,回答“教师干预策略如何优化支架效能”的问题。需要特别说明的是,表 1/2 与表 3 的测量指标分属两个不同的数据体系:前者来自口语语料的 CAF 分析(词汇多样性、句法复杂度、准确度、流利度等),后者来自 AI 交互日志与批判性任务单的编码分析(批判性采纳率、深度修订次数、元语言评论字数等)。两套指标虽然在内容上不同,但在时间点上存在对应关系——表 3 的“第二轮结束时(T3)”数据与表 1/2 的“后测(T3)”数据在时间上一致,但测量维度不同。两者共同服务于“效果验证+机制解释”的研究目标,形成从“是否有效”到“为何有效”的完整证据链。

4.1 第一轮:初步尝试与问题诊断

第一轮干预持续 3 周,围绕 POA 闭环展开。教师构建了与“龙井茶村”主题高度契合的语言支架库,采用“课堂引导—AI 训练—成果固化”的三位一体模式。然而,随着教学推进,三个核心问题逐渐浮现。

首先是句法能力的系统性薄弱。前测数据显示,实验组平均 T 单位长度为 8.45 ± 1.23 ,从句与 T 单位比率仅为 0.38 ± 0.08 ,远低于高水平口语产出的常规阈值,多数学生口语产出停留在简单句堆砌层面。其次是 AI 交互认知层次的浅表化。词汇替换类建议采纳率虽高达 78%,但词汇密度指标未呈现显著提升,部分学生产出呈现模板化特征,

疑似直接复制 AI 输出而未加消化。最后是跨文化转译策略性知识的缺失。学生在处理“龙井”“炒青”等文化负载概念时普遍采取字面翻译策略，对文化负载词解释类支架的采纳率低于 30%。

上述诊断结果表明，后续干预不能仅满足于促使学习者“使用”AI，而应通过结构化任务设计，引导其进行批判性、选择性的采纳，从而将 AI 从“提供答案”转化为“催化内化”。

4.2 第二轮：策略性重构与认知赋能

在第二轮教学实施过程中，本研究基于第一轮研究的发现，对高级英语口语课程的同一核心任务进行了深化实施。在数据收集方面，在延续首轮口语语料与 AI 交互日志采集的基础上，本轮实现了三个重要突破，主要是新增的“批判性 AI 交互任务单”为分析学生的认知加工深度提供了直接证据；“刺激性回忆访谈”有效揭示了学生决策背后的思维过程；而系统化的“教师干预日志”则完整记录了教师根据课堂动态进行的实时调整，为解析教师主导作用的具体效能提供了过程性数据支持。

为全面评估 POA 视域下 AI 生成式支架的教学效果，

本研究从 CAF 量化数据、过程性表现、主观感知和社会文化维度四个层面进行系统分析。

4.2.1 AI 生成式支架对 CAF 三维度的协同提升效果

通过对 60 名学习者（实验组 30 人，对照组 30 人）前后测数据的统计分析，获得以下结果。所有数据分析均采用 SPSS 26.0 进行，连续变量经 Shapiro-Wilk 检验符合正态分布（ $p > 0.05$ ），组间比较满足方差齐性假设（Levene 检验 $p > 0.05$ ），见表 1。

实验组在词汇和句法复杂度各指标上均呈现显著提升（ $p < 0.01$ ），组间比较显示实验组各指标的增益幅度显著大于对照组（ t 值均达显著水平）。这表明“深度句法重构工作坊”和“文化转译策略库”有效提升了学生的语言复杂度。

实验组在准确度和流利度各指标上均显示出大效应量的提升（ $d = 1.16 \sim 1.95$ ），特别是在错误修正频率的降低和语速提升方面表现突出。相比之下，对照组的效应量较小（ $d = 0.27 \sim 0.75$ ），仅在流利度指标上达到中等至接近大效应。值得注意的是，对照组在语速（+10.3 音节 / 分， $d = 0.66$ ）和发音时间比（+0.05， $d = 0.75$ ）上也表现出一定

表1 口语复杂度前测（T1）与后测（T3）组间与组内对比分析（N=60）

测量维度	具体指标	组别	前测均值±SD	后测均值±SD	组内t值	组间t值
词汇复杂度	词汇多样性(LV)	实验组	0.42 ± 0.07	0.61 ± 0.08	5.32**	12.45**
		对照组	0.41 ± 0.06	0.46 ± 0.07	1.87	
	词汇密度(LD)	实验组	0.39 ± 0.05	0.51 ± 0.06	4.87**	9.63**
		对照组	0.38 ± 0.04	0.41 ± 0.05	1.24	
句法复杂度	平均T单位长度	实验组	8.45 ± 1.23	13.52 ± 1.85	6.45**	18.32**
		对照组	8.52 ± 1.31	9.76 ± 1.42	2.01	
	从句/T单位比率	实验组	0.38 ± 0.08	0.71 ± 0.12	5.92**	15.76**
		对照组	0.39 ± 0.07	0.45 ± 0.09	1.78	

注：采用独立样本t检验进行组间比较，配对样本t检验进行组内比较；所有连续变量经Shapiro-Wilk检验符合正态分布（ $p > 0.05$ ），满足方差齐性（Levene检验 $p > 0.05$ ）；* $p < 0.05$ ，** $p < 0.01$ 。

表2 准确度与流利度前测（T1）与后测（T3）双重测量结果（N=60）

CAF维度	测量指标	组别	前测均值±SD	后测均值±SD	增益值	合并SD	d [95% CI]
准确度	无错误T单位比	实验组	0.67 ± 0.09	0.78 ± 0.10	+0.11	0.095	1.16 [0.80, 1.52]
		对照组	0.66 ± 0.08	0.70 ± 0.09	+0.04	0.085	0.47 [0.13, 0.81]
	错误修正频率	实验组	12.5 ± 3.2	8.3 ± 2.5	-4.2	2.89	1.45 [1.07, 1.84]
		对照组	12.8 ± 3.5	11.9 ± 3.1	-0.9	3.33	0.27 [-0.07, 0.61]
流利度	语速(音节/分)	实验组	106.8 ± 15.3	135.2 ± 18.7	+28.4	17.22	1.65 [1.24, 2.06]
		对照组	105.3 ± 14.8	115.6 ± 16.2	+10.3	15.50	0.66 [0.31, 1.00]
	发音时间比	实验组	0.53 ± 0.07	0.69 ± 0.09	+0.16	0.082	1.95 [1.50, 2.40]
		对照组	0.52 ± 0.06	0.57 ± 0.07	+0.05	0.066	0.75 [0.39, 1.10]

注：前测（T1）于第一轮教学开始前采集，后测（T3）于第二轮教学结束后采集。效应量 $d = (M_{后} - M_{前}) / SD_{合并}$ ， $SD_{合并} = \sqrt{[(SD_{前}^2 + SD_{后}^2) / 2]}$ 。95% CI基于非中心t分布计算（Cumming, 2012）。判断标准： $d = 0.2$ （小效应）、 $d = 0.5$ （中等效应）、 $d = 0.8$ （大效应）（Cohen, 1988）。组内比较采用配对样本t检验，组间比较采用独立样本t检验。所有变量经Shapiro-Wilk检验符合正态分布（ $p > 0.05$ ），满足方差齐性（Levene检验 $p > 0.05$ ）。

提升。这一现象可从两个方面解释。其一，对照组接受的常规 POA 口语教学本身包含充分的产出练习机会，学生经过 16 周的学习后流利度自然增长，属正常的成熟效应与教学效应。其二，对照组的提升幅度虽然可观，但显著小于实验组 ($d \approx 1.2 \sim 2.0$)，表明 AI 生成式支架在流利度提升上具有额外的增益效应，而非单纯替代传统教学效果。这一对比进一步强化了实验组 CAF 协同发展的归因可信度：实验组的提升不能简单归因于时间推移或一般教学效果，而应归因于 AI 生成式支架的介入。综合表 1 与表 2 的数据，可以得出结论：AI 生成式支架在提升语言流利度的同时，并未牺牲准确度，而是实现了 CAF 三维度的协同发展。

4.2.2 教师干预策略的优化路径与教学效能

为揭示教师干预策略如何优化 AI 生成式支架的教学效能，本研究结合过程性数据、教师反思日志与同行观察，从“认知脚手架”与“元认知催化”两个维度进行了深入分析，见表 3。

过程性数据显示（见表 3，T2 vs T3），学生的批判性采纳率与元语言评论字数在第二轮干预后显著提升，表明“批判性 AI 交互任务单”与“深度句法重构工作坊”等策略有效引导学生从被动接受 AI 输出转向主动进行语言分析与策略选择。上述指标（批判性采纳率、深度修订次数、元语言评论字数）可参照王华、张雅茜（2025）提出的 AI 技术媒介任务“行为—认知—主体”参与三维框架解读：深度修订与元语言评论的增长，既反映行为层面的互动频次提升，更体现认知参与（对语言形式的主动监控）与主体参与（对 AI 反馈的批判性评估与规划）的同步深化。

教师反思日志精准揭示了这一转变的教学作用机制。“批判性 AI 交互任务单”中“记录原始提问—AI 回应—采纳与否—修改理由”的四步结构，与文秋芳、梁茂成（2024）所提人机互动协商能力（HAINC）中“分析反

馈—调整策略”的循环要素形成呼应，将隐性的认知过程显性化。教师日记记录到，学生的修改理由从初期的“因为 AI 说这样更好”逐渐转变为“选择版本二是因为其非谓语句词开头能更生动地呈现画面感”。这一转变表明，教学干预通过强制显性化认知过程，成功将学生与 AI 的互动质量从“机械采纳”提升至“策略性内化”。同行教师课堂观察为此提供了外部视角。观察报告指出，在“深度句法重构工作坊”中，学生能围绕 AI 生成的多个句式变体展开“高质量的元语言讨论”，例如深入比较定语从句与同位语结构在语篇功能上的细微差别。这一观察结果与量化数据中句法复杂度的显著提升，以及过程性数据中“元语言评论字数”的大幅增长形成了有力呼应，共同印证了教师设计的认知脚手架在催化深度学习中的关键作用。

4.2.3 学习者对融合模式的教学感知与接受度

通过对刺激性回忆访谈、反思札记及 AI 交互日志的三角验证分析，本研究发现学生对“教师—AI—学生”三元协同模式表现出高度认同。在认知自主性方面，学生普遍认为“批判性任务单”迫使其主动思考而非被动修改，从而建立起更自主的语言学习与优化能力。AI 交互日志也为此提供了客观佐证：第二轮中学生对 AI 输出的批判性采纳率显著提升，机械复制行为大幅下降，修订理由从笼统的“更好听”转变为对句式功能与修辞效果的明确阐述。

在跨文化阐释能力方面，学生对文化转译策略库表现出高度接纳。刺激性回忆访谈显示，学生的文化转译能力呈现系统性提升：在介绍龙井茶时，不再简单称之为“green tea”，而是将其阐释为“a living artifact of Hangzhou's tea civilization”；在介绍“炒青”工艺时，能精心设计跨文化类比，既建立认知参照又突出文化独特性。这些质性发现与 AI 交互日志中文化转译策略采纳率的提升趋势相互印证，共同表明学生已初步建立起系统的跨文化阐释能力。

表3 AI交互深度与认知策略发展（实验组n=30，T2 vs T3）

观察维度	测量指标	第一轮结束时(T2)均值 ± SD	第二轮结束时(T3)均值 ± SD	提升幅度	t值
AI交互质量	批判性采纳率(%)	31.5 ± 8.2	70.2 ± 9.5	+38.7	6.32**
	深度修订次数(次)	1.7 ± 0.5	3.8 ± 0.9	+2.1	5.87**
策略运用	元语言评论字数(字)	25.3 ± 7.1	82.6 ± 12.3	+57.3	7.45**
	文化转译策略数(种)	1.3 ± 0.4	2.9 ± 0.7	+1.6	5.74**
	句式变体使用数(种)	1.1 ± 0.3	2.7 ± 0.6	+1.6	5.62**

注：T2为第一轮干预结束后（第3周末），T3为第二轮干预结束后（第16周末）。该表数据来源于实验组AI交互日志与批判性任务单编码，与表1/2的CAF口语语料数据分属不同测量维度：表1/2回答“AI生成式支架是否有效”（整体效果），本表回答“教师干预策略如何优化支架效能”（过程机制）。两条线索共同构成从“效果验证”到“机制解释”的完整证据链。**p<0.01。

5 不足与启示

本研究通过两轮行动研究验证了 POA 指导下 AI 生成式支架的有效性,但在实践过程中也暴露出若干值得深入反思的问题。在 AI 交互的“度”的把握上存在现实难题。尽管批判性任务单显著提升了学生的元认知水平,但过程数据显示,仍有约 25% 的学生在第二轮后期表现出对 AI 建议的机械采纳。刺激性回忆访谈发现,这些学生普遍面临“认知超载”——他们既渴望通过 AI 获得即时改进方案,又难以在有限课时内消化多个句式变体背后的修辞逻辑。这一矛盾提示,元认知训练需要比预期更长的浸润期,且教师需设计更精细的“认知阶梯”,帮助学生逐步建立自主判断力。这一发现亦与葛楠、郑春萍(2025)对智能聊天机器人“情感支持功能负面效应”的警示形成互证,即当技术反馈过于“顺滑”时,部分学习者会倾向让渡判断权,批判性 AI 交互能力的培养因而需要比预期更长的浸润周期与更精细的认知阶梯设计。

与此同时,任务复杂度与课时限制之间存在显著张力。本研究的产出任务涉及语言能力、文化转译与数字素养的多重维度,对 16 周的教学周期构成了挑战。有学生指出,连续的任务安排挤压了反思与消化的空间,导致学习过程更像是“赶工”而非探索。教师日志也记载为完成预设内容而压缩了小组讨论时间。这反映出 POA 在处理复杂产出任务时,需要对“促成环节”进行更现实的时序规划,必要时对任务进行“模块化”分解,跨越多个教学周期逐步实现。

基于以上反思,未来研究可从三方面进行优化:开发“渐进式自主权转移”模型,在教学中逐步放宽对 AI 使用的规制;采用“项目式学习”框架,将复杂任务转化为可持续一学期的学习历程;进一步完善个体学习过程的追踪机制,通过超星平台的交互日志持续记录每位学习者的 AI 交互频次、采纳策略与修订路径,为个性化反馈提供数据支撑。

6 结语

本研究以 POA“精准性、渐进性、多元性”促成理念为规约,通过两轮迭代行动研究,在高级英语口语课程中检验了教师主导设计的 AI 生成式支架对口语 CAF 三维度协同发展的作用路径,验证了教师通过精准提示词设计、元语言讨论引导与批判性任务单干预,将 AI 从“答案提供者”转化为“认知催化剂”的内在机制。在理论层面,本研究将 POA 的驱动与促成机制从传统课堂延伸至人机协同场景,以 CAF 协同发展的实证结果回应了“技术赋能有

余、教学规约不足”的核心争议。在实践层面,所构建的深度句法重构工作坊、文化转译策略库与批判性 AI 交互任务单等支架工具,为高级口语教学的共性难题提供了可复制的解决方案。

本研究仍存在若干局限。样本规模较小(N=30/组),对中等效应统计检验力仅约 0.38,对照组部分未达显著的指标不宜直接解读为“无差异”;且样本来自同一院校,外部效度有待扩大检验。16 周的研究周期对于元认知养成仍显短暂;“龙井茶村”单一任务限制了结论的迁移性;数据采集未能实时捕捉人机对话中的语音特征与情感反应。未来研究可扩大样本范围、延长干预周期,并拓展学术演讲、跨文化谈判等多元任务类型,以检验 AI 生成式支架在不同语域中的迁移效果,推动人工智能与语言教育的深度融合。

参考文献:

- [1] 葛楠, 郑春萍. 智能聊天机器人支持英语口语教学的系统性综述[J]. 外语教学, 2025, 46(3): 56-63.
- [2] 孔蕾. “人机-人际协同”外语教学模式构建[J]. 外语教学与研究, 2026, 58(1): 108-120.
- [3] 邱琳. “产出导向法”促成环节设计标准例析[J]. 外语教育研究前沿, 2020, 3(2): 12-19.
- [4] 王华, 金艳. 人工智能辅助口语诊断测评中学习者参与反馈研究[J]. 现代外语, 2026, 49(3): 367-379.
- [5] 文秋芳. 构建“产出导向法”理论体系[J]. 外语教学与研究, 2015, 47(4): 547-558+640.
- [6] 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+321.
- [7] 文秋芳, 孙曙光. “产出导向法”驱动场景设计要素例析[J]. 外语教育研究前沿, 2020, 3(2): 4-11+90.
- [8] 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(4): 105-111.
- [9] 徐锦芬, 陈聪. 认知要求对学习者的口语产出质量与注意分配的影响[J]. 外语与外语教学, 2018(6): 42-52+144-145.
- [10] 于涵静, 戴炜栋. 英语学习者口语复杂性、准确性的动态发展研究[J]. 外语与外语教学, 2019(2): 100-110+150.
- [11] 邹娟娟. 生成式人工智能支持的英语口语“自评+自比”人机协同评价研究[J]. 现代英语, 2025(14): 68-71.
- [12] Du, J., & Daniel, B. (2024). Transforming Language Education: A Systematic Review of AI-Powered Chatbots for

English as a Foreign Language Speaking Practice[J]. *Computers and Education: Artificial Intelligence*, 6: 100230. <https://doi.org/10.1016/j.caeai.2024.100230>.

[13] Hong, Y., & Lee, J. (2026). Developing a Generative AI-Based English Speaking Chatbot with Level-Specific Feedback: Effects on Speaking Performance and Anxiety, 34(1), 209-233. *The Korea Association of Yeolin Education*. <https://doi.org/10.18230/tjye.2026.34.1.209>.

[14] Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The Impact of Different Conversational Generative AI Chatbots

on EFL Learners: an Analysis of Willingness to Communicate, Foreign Language Speaking Anxiety, and Self-perceived Communicative Competence. *System*, 127,103533. <https://doi.org/10.1016/j.system.2024.103533>.

基金项目：第二批国家级线上线下混合式一流课程“高级英语视听说”阶段性研究成果。

作者简介：陈霞（1979-），女，中国浙江宁波人，硕士研究生，教授，供职于浙江越秀外国语学院，研究方向：从事区域国别学（非洲方向）和数智赋能外语教学研究，邮箱：20092009@zyufl.edu.cn。