

AI智能测评视域下英语师范生教学设计能力的 诊断与干预策略

王 蕾*, 陈 愉

鞍山师范学院外国语学院, 辽宁 鞍山

收稿日期: 2026年8月4日; 录用日期: 2026年9月9日; 发布日期: 2026年9月17日

摘 要

本研究基于AI智能测评, 构建了英语师范生教学设计能力诊断体系, 并通过闭环干预模式验证了其有效性。研究采用准实验设计, 将189名师范生随机分为三组: A组(人机协同组, 接收AI报告 + 教师解读)、B组(传统教师反馈组)、C组(纯AI反馈组)。结果表明, A组在目标设计、内容选择、活动安排及评价方法四个维度上均取得极其显著的提升($p < 0.001$), 其效果显著优于仅接收教师反馈的B组($p > 0.05$)和仅接收AI报告的C组。方差分析进一步证实, A组在所有维度上均表现最佳, 而C组在目标设计与评价设计上显著优于B组。该模式通过AI提供的客观诊断与教师的支架式引导, 有效提升了师范生的教学设计能力, 尤其对低起点学生具有良好补偿效应, 为AI赋能教师教育提供了实践路径。

关键词

AI智能测评, 师范生, 英语教学设计能力, 诊断与干预, 教师教育

Diagnosis and Intervention Strategies for Pre-Service English Teachers' Instructional Design Competency from the Perspective of AI-Driven Assessment

Lei Wang*, Yu Chen

School of Foreign Languages, Anshan Normal University, Anshan Liaoning

Received: August 4, 2026; accepted: September 9, 2026; published: September 17, 2026

*通讯作者。

文章引用: 王蕾, 陈愉. AI 智能测评视域下英语师范生教学设计能力的诊断与干预策略[J]. 创新教育研究, 2026, 14(9): 403-411. DOI: 10.12677/ces.2026.149703

Abstract

This study constructs a diagnostic system for the teaching design ability of English student teachers based on AI intelligent assessment and verifies its effectiveness through a closed-loop intervention model. Using a quasi-experimental design, 189 student teachers were randomly assigned to three groups: Group A (human-AI collaboration, receiving AI reports + teacher interpretation), Group B (traditional teacher feedback), and Group C (AI-only feedback). The results show that Group A achieved extremely significant improvements in all four dimensions of teaching design ability ($p < 0.001$), significantly outperforming Group B ($p > 0.05$) and Group C. Further ANOVA analysis confirms that Group A performed best across all dimensions, while Group C significantly outperformed Group B in target design and evaluation design. This model effectively enhances student teachers' teaching design ability by combining objective AI diagnostics with teacher-led scaffolding, particularly benefiting low-achieving students, and provides a practical pathway for AI-empowered teacher education.

Keywords

AI-Driven Assessment, Pre-Service Teachers, English Instructional Design Competency, Diagnosis and Intervention, Teacher Education

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. 研究背景与意义

师范生的教学设计能力是其专业素养的核心支柱, 直接关乎未来课堂教学质量与职业发展高度。《中学教师专业标准(试行)》¹将其置于专业能力首位, 凸显了基础性地位。然而, 当前培养面临评价过度依赖指导教师经验和个性化反馈滞后的严峻挑战。近年来, 基于自然语言处理与知识图谱的人工智能技术为教师教育评价注入新动能[1][2]。

本研究聚焦 AI 智能测评视域, 旨在构建师范生中学英语教学设计能力诊断指标并设计精准干预策略, 解决能力短板难量化、干预路径缺乏适配性等关键问题。现实意义在于以智能测评驱动能力精准提升, 优化教师教育课程实效; 理论价值在于探索数据驱动评价新范式, 推动评价由经验判断向循证诊断转型, 深化 AI 与教师教育融合的理论认知。

2. 文献综述

英语师范生教学设计能力的培养长期沿用“课程学习 + 微格实训 + 教学实习”的传统路径, 评价多依赖指导教师基于通用量表的经验性评分。针对英语学科需兼顾语言能力与育人价值的特点, 《普通高中英语课程标准日常修订版(2017年版 2025年修订)》强调主题意义探究与学习活动观[3], 对师范生的教学设计能力提出了更高要求[3]传统评价在细颗粒度指标上缺乏客观量规, 难以揭示学科内容整合等隐性短板。

AI 测评技术为突破此困境提供了新路径。钟博维(2025)基于大语言模型构建了四层架构的智能评价

¹http://www.moe.gov.cn/srcsite/A10/s6991/201209/t20120913_145603.html

模型，验证了个性化学习对实训效果的提升作用[1]；刘晓红等(2026)开发的多智能体系统亦证实了精准干预职前教师能力的价值[2]。但现有研究仍存局限：一是通用模型对英语学科特性表征不足，对“语言学习活动观”等内容测评效度有限；二是诊断结果多停留于分数，缺乏可操作的个性化干预衔接。

本研究植根于形成性评价与教师专业发展阶段理论。形成性评价强调将评价嵌入学习过程，通过持续反馈促进能力建构[4]。Berliner (1988)的专长发展阶段理论揭示，处于新手过渡期的师范生需借助诊断性反馈与支架式指导实现行为内化[5]。蒋宇瑛等(2022)提出的基于大数据的精准诊断和施策[6]，也为“诊断-支架-迭代”闭环提供了实践参照。

本研究承袭智能技术赋能教师教育评价的脉络，构建融合英语学科特征的评价指标体系。将形成性评价理念与能力发展阶段理论深度嵌入诊断逻辑，据此设计数据驱动的个性化干预策略，力图为学科化、可追踪的 AI 测评干预研究提供范例。

3. 研究设计与方法

本研究选取东北地区 L 省 A 市某地方师范院校 2023 级英语专业三年级全体师范生共 189 人为对象。为深入探究人工智能工具在英语师范生教学设计能力反馈中的具体效能及“人机协同”模式的独特价值，本研究采用准实验设计，将研究对象随机分配至 A、B、C 三组，由三名教师分别任课，并实施三种不同的教学干预模式。各组样本量分别为：A 组 63 人，B 组 64 人，C 组 62 人。具体实验实施过程(见表 1)：

Table 1. Experimental group design
表 1. 实验组别设计表

组别	名称	样本量(N)	干预描述
A	人机协同实验组	63	接收 AI 报告 + 教师解读课
B	传统教师反馈组	64	仅接收教师人工反馈(无 AI)
C	纯 AI 反馈组	62	仅接收 AI 报告(无教师解读)

大学三年级的英语专业师范生已经系统修读教育学、心理学及学科教学论等前序课程，正在参与《中学英语教学设计》课程，具备基础理论素养但缺乏实际教学经验，正处于由“知”向“行”转化的关键期。选取该样本能有效控制实习经验等无关变量，清晰呈现 AI 干预的真实效果，且地方师范院校培养质量直接关联区域基础教育师资水平，具有典型现实观照价值。

研究采用混合方法设计，收集并处理量化和质性两种数据。量化层面，由任课教师收集 AI 评定的各维度分数，以衡量能力变化。质性数据方面，由于研究者同时是实验班的任课教师，承担双重角色，可能带来实验者效应、干预实施的不一致性、数据收集与处理的偏倚和伦理与角色冲突等可能带来的潜在偏倚，故质性数据的采集和编码都由非任课教师完成。该教师对三个不同组别的学生共 6 人，分别来自于 A、B、C 三组，每组随机抽取两人进行半结构化访谈，以理解学生感知与运用 AI 建议的过程。量、质数据三角互证增强了结果解释力。

课程结束前，收集 A、B、C 三组学生最终教学设计作为后测数据，全程历时 14 周。数据分析方面，量化部分采用配对样本 t 检验和单因素方差分析进行差异对比，质性数据采用建构主义扎根理论三级编码[7]，提炼反馈的模式与策略特征。即编码过程并非对参与者真实体验的客观提取，而是研究者在访谈互动、文本细读与理论敏感性之间不断往返的解释性建构。开放编码阶段，研究者首先将访谈转录稿逐行拆解为行动导向的初始概念。主轴编码阶段，通过“条件-互动-结果”矩阵将初始概念聚类为过程性范畴。选择性编码阶段，以“内化”为核心范畴，将三组差异建构为同一过程的不同变体(variations)，

而非孤立的组间对比。

研究工具采用 Kimi AI 智能助手的“项目功能”，该功能为长线任务的持久化工作空间。教师搭建“英语教学设计评估”项目，嵌入目标设计、内容组织、活动安排、评价方法四维度评分标准。Kimi 在项目内自动遵循指令，输出教学设计评价报告，形成指向明确的诊断。为了保证研究工具的评价效度，在正式实验开始前，研究者抽取了以往教学设计样本 30 份，同时邀请 1 名中学英语教研员和 1 名高校英语课程与教学论专家进行独立评分，然后使用 ICC 计算 AI 评分与 2 专家评分之间的一致性，Fleiss's Kappa 系数为 0.781，说明一致性很好，Kimi AI 的评价效度可靠。

实验遵循准实验设计。前测取自上学期期末教学设计作品，经 AI 诊断生成初始报告。前测表明三组实验参与学生在实验初始阶段英语教学设计能力不存在显著差异， p 值均大于 0.05 (见表 2)。

Table 2. Pre-test comparison of each dimension of instructional design competence between experimental and control groups
表 2. 实验组与对照组教学设计能力各维度前测比较表

维度	总均值	A 组	B 组	C 组	ANOVA p 值
目标设计能力	2.879	2.881	2.957	2.797	0.272
内容选择与组织能力	3.322	3.242	3.43	3.293	0.196
活动安排能力	3.296	3.319	3.341	3.226	0.543
评价方法设计能力	2.93	2.908	2.898	2.985	0.686

干预阶段，A 组学生不仅能够获取基于人工智能技术生成的个性化学习分析报告，还配套实施了教师深度解读机制。即在每份 AI 报告发布后，教师会根据报告内容组织专门的解读课或辅导环节，旨在通过教师的认知引导与情感支持，辅助学生深度理解 AI 反馈信息，促进学生对学习策略的调整与内化。B 组作为传统教学模式的对照，完全由教师依据其教学经验与学生表现提供反馈意见。在此过程中，不引入任何 AI 生成的报告或数据分析工具，以维持常规的师生反馈互动作为基准参照，用于衡量传统人工反馈的效果。C 组学生仅接收与 A 组内容一致的 AI 生成学习报告，但剔除了教师解读课这一干预环节。该组旨在剥离教师引导变量，单纯考察 AI 工具提供的数据驱动反馈对学生学习产生的直接影响。

4. 诊断体系构建

本研究构建的英语教学设计能力诊断指标体系，植根于《普通高中英语课程标准日常修订版（2017 年版 2025 年修订）》倡导的英语学习活动观[3]，融合 Smith-Ragan 教学设计模型的分析、策略与评价三阶段逻辑[8]。体系涵盖四个核心维度：目标设计能力，考察基于主题意义等要素分析制定可检测目标的能力；内容选择与组织能力，检视语篇解读深度与内容组织逻辑；活动安排能力，聚焦学习理解、应用实践、迁移创新三类活动的层次性；评价方法设计能力，判断是否体现“教-学-评”一致性(图 1)。

依据英语教学设计能力诊断指标体系并借鉴了 Hattie 和 Timperley (2007)提出的“任务-过程-自我调节”三层次反馈模型[9]，设计 Kimi AI 智能助手的“项目功能”统一指令语(prompt)：请你以教学专家的身份，从以下四个维度评价每份上传的教学设计，并写出评语。(1) 目标设计能力，即能否基于对主题意义、语篇类型、语言知识、文化内涵、语言技能及学习策略的深层分析，制定可达成、可检测的课时目标；(2) 内容选择与组织能力，考察教学语篇解读的深度、主题意义探究路径的合理性及内容组织的逻辑递进性；(3) 活动安排能力，聚焦学习理解、应用实践、迁移创新三类活动的层次性、关联性与实践性；(4) 评价方法设计能力，检视学习评价活动是否体现“教-学-评”一致性，能否有效反馈学生学习表现。评语输出要求：1) 每个维度 5 分，先写 1~2 条优点，然后写改进意见；2) 将每份评语生成可下载的 word

文档; 3) 依据每份教学设计的评分生成相应的雷达图并嵌入 word 版的评语中。该提示语保证了笼统表扬的大幅减少, 并增加指向明确的操作性指令。例如, 针对目标设计短板, 评价报告会直接显示“请将‘辨识’替换为‘列举(三条原因)’等可观测的行为动词”, 或者直接提示“补充达成条件”, 使师范生即刻获得清晰的修改锚点, 而非仅面对一个抽象的低分。

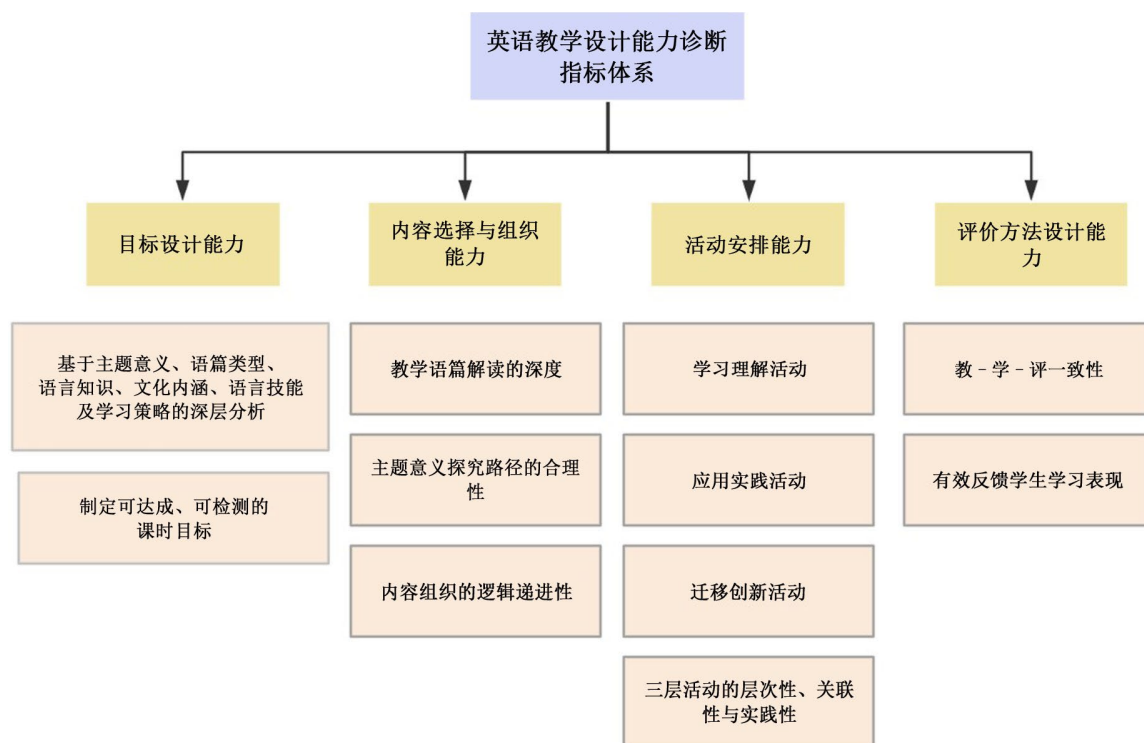


Figure 1. Diagnostic indicator system for English instructional design competence

图 1. 英语教学设计能力诊断指标体系

教师上传学生教学设计文件至 AI 项目后, 系统经结构解析、语义理解、多维判断三层次加工, 输出四维度 5 分制量化评分与描述性质性反馈, 界面同步呈现雷达图与列表分布。系统输出三类数据: 一是各维度独立得分与加权总分等连续变量; 二是标注具体段落指向的“优点”与“改进建议”等结构性文本; 三是动词层级分布、活动数量比例等隐性特征统计数据。测评嵌入 A 组(实验组)和 C 组(仅接收 AI 报告组)《中学英语教学设计 with 技能训练》课程全流程, 设初、中、末三次节点。前测确立基线, 中段支持迭代优化, 后测检验成效。测评定位于学习性评价, 结果仅供引导反思与定向练习, 不计入终结性考核, 确保了评价的低风险与高反馈价值。

5. 能力诊断与问题分析

借助 Kimi “英语教学设计评估”项目, 研究对 189 名师范生期末教学设计进行批量诊断。系统自动解析文本并输出评分与反馈。统计显示, 目标设计($M = 2.88$, $SD = 0.56$)与评价方法设计($M = 2.93$, $SD = 0.61$)得分显著低于内容选择与活动安排维度, 构成能力短板。

质性反馈揭示深层症结在于, 学生多数使用“了解”“掌握”“理解”和“明白”等动词, 缺乏可观测行为动词, 质性访谈中, 超 80% 的学生表示“不知道要在教学目标中写明达成目标所具备的条件”, 导致目标笼统难测, 背离了教学目标需要“可达成、可检测”的基本要求; 教学评价设计上, 学生常出

现教学活动与教学目标脱节的错误,“教-学-评”一致性薄弱。活动安排虽得分较高,但三类活动间逻辑关联不足。

究其根源,一是理论向实践转化桥梁缺失,存在“知-行”脱节[10];二是缺乏结构化反思支架,修改依赖直觉致提升浅表化。这提示干预须强化专项训练,并嵌入以证据为导向的反思环节,以精准补齐短板。

6. 干预策略设计与实施

基于首次诊断结果呈现的目标设计(M = 2.88)与评价方法设计(M = 2.93)两大短板, A 组学生(实验组)的任教教师特别设计了干预模块化工作坊。目标设计工作坊采用“案例对比 + 小组重构”模式:教师展示同一语篇的优劣目标实例,引导师范生对照课程标准中“可达成、可检测”的要求[3]进行归因分析,随后以小组为单位改写目标并利用 AI 即时反馈检验改写效果。评价设计工作坊则聚焦“教-学-评”一致性,组织师范生剖析优秀教案的评价量表,自行提炼反馈要素并现场设计评价活动,教师针对共性问题集中点拨。两项工作坊各 4 课时,穿插于课程第 3 至第 10 周,为师范生搭建从理论认知到实践操作的支架。

第 7 周启动中期二次测评。学生提交经首轮反馈修改后的教学设计, AI 系统重新诊断并生成对比报告,直观呈现四维度的分数增减与改进幅度。师范生对比前后两份报告,识别已改善的环节与依然固化的短板,据此进入第二轮针对性修订。学期末提交的教学设计终稿作为后测,由此形成“诊断-反馈-修改-再诊断-再修改”的闭环流程,使能力提升有迹可循、持续累积。

为避免实验组学生机械遵从 AI 建议,每次测评后均设置“反馈解读课”。任课教师先以一份典型设计为例,示范如何将 AI 的质性评价转化为具体的教学设计修改策略,随后引导师范生填写结构化反思日志,逐项回答“原设计的不足、AI 建议了什么、我做了何种修改及依据”。这一流程将 AI 生成的“冷反馈”与教师提供的“热支架”有机融合,推动师范生从“被告诉该改”转向“理解为何这样改”,促进基于证据的自我调节学习[4],实现专业成长的自主化。

7. 研究结果与成效分析

14 周后,对三组英语专业师范生(共 189 人)前后测数据进行配对样本 t 检验(见表 3),结果显示: A 组在目标设计能力、内容选择与组织能力、活动安排能力以及评价方法设计能力这四个维度上,后测成绩均极其显著地高于前测($p < 0.001$),表明该组合干预措施对提升师范生的综合教学设计能力具有强大且全面的效果。相比之下,仅接收教师人工反馈的 B 组,其提升效果则非常有限,仅在活动安排能力上表现出显著提升($p < 0.01$),而在其他三个关键维度上均无统计学上的显著变化($p > 0.05$),这揭示了传统教师反馈在提升教学设计能力方面的局限性。有趣的是,仅接收 AI 报告的 C 组,在所有四个维度上都取得了极其显著的进步($p < 0.001$),证明了 AI 报告本身作为一种教学反馈工具的独立价值。

Table 3. Paired-sample T-test of instructional design competence for pre-test and post-test

表 3. 教学设计能力前后测配对样本 T 检验

维度	组别人数	前测 M ± SD	后测 M ± SD	t	p
目标设计能力	A (N = 63)	2.88 ± 0.57	4.02 ± 0.58	-15.63	<0.001
	B (N = 64)	2.96 ± 0.52	3.05 ± 0.55	-1.31	0.196
	C (N = 62)	2.80 ± 0.57	3.68 ± 0.58	-11.88	<0.001
内容选择与组织能力	A (N = 63)	3.24 ± 0.55	3.92 ± 0.56	-9.60	<0.001
	B (N = 64)	3.43 ± 0.57	3.45 ± 0.58	0.17	>0.05
	C (N = 62)	3.30 ± 0.69	3.58 ± 0.69	-3.30	0.002

续表

活动安排能力	A (N = 63)	3.32 ± 0.56	4.02 ± 0.58	-9.58	<0.001
	B (N = 64)	3.34 ± 0.58	3.57 ± 0.60	3.11	<0.01
	C (N = 62)	3.23 ± 0.70	3.68 ± 0.72	-4.91	<0.001
评价方法设计能力	A (N = 63)	2.91 ± 0.56	3.92 ± 0.58	-13.85	<0.001
	B (N = 64)	2.90 ± 0.73	2.95 ± 0.73	-0.55	0.585
	C (N = 62)	2.96 ± 0.51	3.58 ± 0.52	-8.90	<0.001

综合来看，AI 诊断报告能够有效提升师范生的教学设计能力；而将 AI 报告与教师的深度解读相结合(A 组)，则能产生更优的协同效应，其提升效果在多个维度上优于仅使用 AI 报告(C 组)，并远超仅依赖教师人工反馈(B 组)。

通过单因素方差分析，我们得以从整体上检验 A、B、C 三组在后测成绩上的差异(见表 4)。结果显示：在目标设计能力、内容选择与组织能力、活动安排能力以及评价方法设计能力这四个维度上，三组之间均存在极其显著的组间差异($p < 0.001$)。进一步的事后检验(LSD)则清晰地揭示了各组间的具体关系：首先，作为实验组的 A 组(AI 报告 + 教师解读)在所有维度上的表现都显著优于 B 组(仅教师反馈)和 C 组(仅 AI 报告)，这再次印证了“AI 报告 + 教师解读”这一组合干预措施具有最佳效果。其次，C 组(仅 AI 报告)在目标设计能力和评价方法设计能力上的表现显著优于 B 组(仅教师反馈)，而在内容选择与组织能力和活动安排能力上，C 组与 B 组之间则无显著差异。综合来看，方差分析的结果不仅验证了三组间存在整体差异，更直观地展示了 A 组的优越性，以及 AI 报告在部分能力维度上对传统教师反馈的超越，为“AI 辅助教学”的有效性提供了强有力的统计证据。

Table 4. One-way ANOVA of instructional design competence across different groups
表 4. 不同组别教学设计能力单因素方差分析结果表

维度	F	df (组间)	df (组内)	p (F)	事后检验(LSD)
目标设计能力	89.58	2	186	<0.001	A > B, A > C, C > B
内容选择与组织能力	21.04	2	186	<0.001	A > B, A > C
活动安排能力	17.88	2	186	<0.001	A > B, A > C
评价方法设计能力	61.92	2	186	<0.001	A > B, A > C, C > B

半结构化访谈结果与量化数据高度吻合。接收到 AI 反馈报告的 A 组和 C 组多数师范生反映，AI 反馈“更具体，直接告诉我将行为动词替换什么”“改进意见都能落地”，促使他们在目标叙写中自觉使用可观测动词。任课教师指出，中期测评后学生修改精准度明显提高，“不再漫无目的地删改”，开始有意识地“追问评价任务是否指向预设目标”。满意度调查显示，87.3%的学生认为 AI 诊断“非常满意或满意”，81.5%表示愿意继续使用此类系统。虽有学生坦言初面对 AI 逐条标注时“感到压力”，但经反馈解读课引导，逐渐将诊断报告视为改进支架而非评判标尺。B 组学生(教师人工反馈组)表示“时常听不懂教师反馈内容，但还不好意思问”。

具体而言，诊断驱动的干预模式在提升师范生英语教学设计能力上更加有效，尤其是目标设计与评价设计两大传统薄弱环节获得实质突破。A 组和 C 组(均接收到 AI 反馈报告)的学生在访谈中均表示：“拿到 AI 评价报告后，一眼就能看清自己四个能力板块的强弱分布，哪些是短板一目了然”。但是，只

有 A 组学生(实验组)明确说明, 参加教学目标设计工作坊时“老师会拿出同一篇语篇的好坏两种课时目标做对比, 带着我们对照 2020 版课标, 从‘可达成、可检测’这个角度分析问题出在哪。之后我们分组改写教学目标, 改完还能用 AI 马上拿到反馈, 看看改得到底怎么样, 不用再全靠自己主观感觉来判断”。这进一步揭示, 这一提升并非源于机械遵从 AI 指令, 而是经由反馈解读、同伴研讨与自我反思的多轮互动, 逐步内化为可迁移的设计自觉。该结果验证了“诊断-支架-迭代”融合模式在教师教育实践中的可行性与有效性[6], 也为化解长期困扰师范院校教学设计评价主观、反馈滞后的难题, 提供了一条以智能技术为辅助、以证据为中心的培养路径。

此外, 依据前测总分将 A 组和 C 组师范生分为高、中、低三组进一步分析发现, 高分组后测总分提升 0.48, 中等组提升 0.53, 低分组提升 0.67。低分组增幅显著大于高分组($t=2.34, p<0.05$)。这一差异表明, 基于 AI 诊断的个性化干预对能力起点较低的师范生具有更强的补偿效应, 有助于缩小群体内部的能力差距。这或许因为低水平学习者的设计文本存在更多可被 AI 明确捕捉的结构性问题, 反馈针对性更强, 改善空间也更大。

然而, 在 A 组和 C 组学生中, 仍存在 2 人在实验前后教学设计能力完全没有提升, 甚至略有下滑的状况。单独访谈中, 2 名同学均声称自己“高考时, 没想学英语, 也没想学师范, 但分儿不够, 被调到了英语师范专业”。这说明他们缺乏对专业的内在兴趣和主动学习的动机, 当面对需要深度投入和持续努力的教学设计课程时, 缺乏内在驱动力的他们难以产生有效的学习行为。另一个原因是被调剂的经验致使他们专业认同薄弱, 即使获得了 AI 诊断和教师干预这些外部支持, 内心没有“我要成为一名好老师”的信念, 外部干预也很难转化为真正的能力提升。

8. 结论与建议

本研究基于 AI 智能测评构建了英语教学设计能力诊断体系, 并通过闭环干预模式验证了其有效性。结果显示, 该模式显著提升了师范生教学设计能力, 尤其对目标设计与评价设计维度改善突出, 且对低起点群体具有良好补偿效应。成效得益于 AI 提供的客观诊断证据与教师引导形成的人机协同合力, 促成了人机协同。然而, 本研究样本仅来自一所地方师范院校, 代表性有限。现有系统诊断指标仅聚焦教学设计四维度, 未涵盖跨学科融合等能力。质性数据仅通过访谈获取, 深度或不足。最后, 本研究证明的“AI+ 教师”干预模式虽然有效, 但其前提是学生具备一定的学习意愿和参与度。对于缺乏动机和认同的学生, 再先进的工具和再精心的指导也可能“对牛弹琴”。

为此建议, 课程应将 AI 诊断嵌入迭代流程, 工具研发需强化学科特色并引入对话功能。教师教育数字化转型应坚持人机协同, 通过 AI 分担重复性诊断释放教师精力, 构建数据驱动与人文关怀融合的培养新范式干预, 拓展诊断评价指标并开发通用工具以深化研究。两个学生的异常数据, 恰恰说明了教学干预的有效性是建立在学生自身主观能动性的基础之上的。同时也在提醒英语教师教育者, 在关注教学方法和评价技术的同时, 必须高度重视学生的专业认同教育和学习动机激发。

基金项目

本研究为鞍山师范学院 2024 年智慧课程建设项目, 项目名称: 《中学英语教学设计与技能训练》, 项目编号: 202407。

参考文献

- [1] 钟博维. 基于大语言模型的师范生教学设计智能评价模型的构建及应用[J]. 高教论坛, 2025(4): 73-80.
- [2] 刘晓红, 朱敏捷, 陈孝然, 叶新燕, 穆肃. 多智能体支持的职前教师教学能力实训: 建构、实施与成效剖析[J]. 中国电化教育, 2026(5): 111-118.

-
- [3] 中华人民共和国教育部. 普通高中英语课程标准日常修订版(2017 年版 2025 年修订) [S]. 北京: 人民教育出版社, 2020.
- [4] Black, P. and Wiliam, D. (1998) Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, **5**, 7-74. <https://doi.org/10.1080/0969595980050102>
- [5] Berliner, D.C. (1988) The Development of Expertise in Pedagogy. American Association of Colleges for Teacher Education.
- [6] 蒋宇瑛, 谢玉晓, 张燕明. 基于教育大数据的教师专业发展策略研究[J]. 教育科学论坛, 2022(1): 63-66.
- [7] Charmaz, K. (2006) Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. Sage Publications.
- [8] Smith, P.L. and Ragan, T.J. (2005) Instructional Design. 3rd Edition, Wiley.
- [9] Hattie, J. and Timperley, H. (2007) The Power of Feedback. *Review of Educational Research*, **77**, 81-112. <https://doi.org/10.3102/003465430298487>
- [10] 刘晓玫. 知行合一能致远[J]. 教师视界, 2020(35): 35.