

基于教育智能体自学的县中学习机制构建 与干预研究

——以高一数学为例

肖忠远^{1*}, 郑富^{2#}

¹北京全品乐衡国际教育管理研究院, 北京

²河南大学生命教育研究中心, 河南 开封

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

随着县域普通高中自习课对高质量学习支持与过程治理需求的增加, 教育智能体在课堂中的应用受到关注。然而, 县域普通高中场景普遍存在低质量交互、过程不可见与策略性应付等问题, 限制了智能体赋能自学的实际成效。针对上述需求, 本研究提出一种“教育专用智能体 + 学习过程治理”的县域普通高中自学支持方案, 构建“目标设定 - 交互学习 - 反向输出评价”的闭环学习机制, 并配套分阶段的反作弊与质量纠偏规则, 以实现大班条件下较为可执行、可审计的课堂治理。研究采用准实验对比与过程数据追踪相结合的方法, 在高一两个平行班的有效样本($n = 86$)中开展为期6周干预, 以数学一模与二模成绩评估成效, 并结合系统对话日志、教师干预记录与学生反向输出文本分析交互行为与阶段演进特征。结果显示时间 \times 组别交互显著($F = 6.145, p = 0.015, \eta^2 = 0.068$), 同时提炼出“新接触 - 适应 - 策略性规避 - 引导修正 - 良好习惯”的五阶段演进模型与关键干预点。研究为县域普通高中人工智能辅助自学提供了具有一定可操作性的实施思路与初步实证参考。

关键词

教育智能体, 县域普通高中(县中), 自主学习, 过程治理与反作弊, 五阶段演进模型

*第一作者。

#通讯作者。

Construction and Intervention Study of Learning Mechanisms for County High School Students Based on Educational Agent Self-Study

—A Case Study of Grade 10 Mathematics

Zhongyuan Xiao^{1*}, Fu Zheng^{2#}

¹Beijing Quanpin Leheng International Education Management Research Institute, Beijing

²Life Education Research Center, Henan University, Kaifeng Henan

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

Abstract

With the growing demand for high-quality learning support and process governance in self-study periods at county-level ordinary high schools, the application of educational agents in classrooms has attracted increasing attention. However, such settings commonly suffer from low-quality interactions, invisible learning processes, and strategic compliance behaviors, which limit the practical effectiveness of agent-supported self-study. To address these challenges, this study proposes a self-study support scheme that combines education-specific agents with learning-process governance, constructing a closed-loop mechanism of “goal setting - interactive learning - reverse-output evaluation”, together with phased anti-cheating and quality-correction rules to support relatively executable and auditable classroom governance under large-class conditions. Using a quasi-experimental comparison combined with process-data tracking, a 6-week intervention was conducted with an effective sample of 86 students from two parallel Grade 10 classes. Mathematics Mock Exam I and II scores were used to assess outcomes, while system dialogue logs, teacher intervention records, and students' reverse-output texts were analyzed to examine interaction behaviors and stage evolution. Results showed a significant Time \times Group interaction ($F = 6.145$, $p = 0.015$, $\eta^2 = 0.068$). A five-stage evolution model of “initial contact-adaptation-strategic avoidance-guided revision-good habits” and corresponding intervention points were identified. The findings provide a practically relevant implementation approach and preliminary empirical evidence for AI-assisted self-study in county-level ordinary high schools.

Keywords

Educational Agent, County-Level General High Schools (County High Schools), Self-Directed Learning, Process Governance and Anti-Cheating, Five-Stage Evolution Model

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. 研究背景与问题提出

2025 年底, 教育部基础教育司负责人在《县域普通高中振兴行动计划》答记者问中指出, 我国县域

普通高中占据普通高中的“半壁江山”，有近 60% 的普通高中学生在县中就读，是县域基础教育的龙头 [1]，这表明县域普通高中在我国高中阶段教育体系中具有基础性地位。这一判断清晰表明：县域普通高中（以下简称“县中”）不仅是我国高中阶段教育的主体承载，更直接关联教育公平、人才培养质量与县域发展韧性。早在 2021 年 12 月，教育部等九部门印发《“十四五”县域普通高中发展提升行动计划》，强调“县域普通高中在推进教育高质量发展和乡村振兴战略中承担着重要使命” [2]，这是新时代促进教育公平、统筹城乡教育发展的重要战略举措。2025 年 9 月发布的《县域普通高中振兴行动计划》进一步明确提出“提高人工智能在教育教学中的应用效能”，为县域高中在资源不足、师资结构不稳定与学习支持系统薄弱的现实条件下探索新路径提供了政策指向。

本研究以河北省某县域高中高一学生为对象，围绕“基于教育专用智能体的教师引导自学行为干预”开展准实验研究：通过实验班与对照班的对比，考察学生学习策略、自主学习表现与数学成绩的变化，并基于过程性对话日志与访谈材料，追踪学生与教育智能体交互行为的演变轨迹。在此基础上，研究进一步提出并检验县中学生“新接触 - 适应 - 策略性规避 - 引导修正 - 良好习惯”五阶段自学发展模型，尝试精准界定不同阶段的典型问题及教师可执行的干预策略，从而为“教育智能体驱动的县中振兴行动”提供可复制、可评估、可落地的实证参考。

1.1. 县域普通高中发展面临的现实困境

县中是指在我国县、县级市、旗等县级行政区内举办、以实施普通高中教育为主要任务的学校，是高中阶段教育的重要组成部分。县中作为农村学生接受高中教育的重要场域，既承载着广大农村学生通过教育实现阶层流动与能力跃升的现实期待，也构成县域教育生态系统的关键节点：对内引领县域内义务教育与高中教育的衔接运转，对外与县域经济、文化与人口流动深度耦合、相互影响。

近年来，随着城镇化进程和大城市“超级中学”的涌现，使许多县中在面向清北输送优秀生源上逐渐失去优势，并普遍陷入了资金不足、优质师资和优秀生源大量流失、教学质量下降的困境 [3]。这种困境具有显著的循环结构：优秀生源与优秀教师向外流动，使县中教学质量与升学竞争力下降；而教学质量与升学竞争力下降又进一步削弱县中吸引优秀生源与优秀教师的能力，最终导致县中内部学习支持系统弱化、学生自学能力与学习策略发展受限，形成“低支持 - 低效率 - 低结果”的结构性堵点。对县中振兴而言，仅依赖增加投入或强化管理并不足以在短期内扭转教学质量，关键在于：在县中现实条件约束下，能否构建一套可复制的学习支持机制，使学生在相对有限的外部资源下依然获得高质量的自学支架与即时反馈，从而提升学习效率与学习结果。

1.2. 教育智能体应用的现实难点

尼尔森(Nilsson)提出，人工智能是关于知识的科学，即研究知识的表示、知识的获取和运用。卢格将人工智能定义为计算机科学的一个分支，它是关于智能行为自动化的研究。融合了人工智能定义与人类专家知识及行为的人工智能系统，被称为“智能体” [4]。教育智能体则是面向教育场景、围绕教学与学习目标进行定制的智能系统，通常包含对教师教学的辅助、对学生学习的指导以及对师生沟通与反馈机制的支持。与传统“静态资源”不同，教育智能体的价值不只在提供答案，而在于提供围绕学习目标的可交互支架：通过追问、提示、诊断与反馈，促进学生概念理解、策略生成与元认知监控。

然而，在县中真实教学场景中，教育智能体要发挥效能，面临至少三类现实难点。第一，教育智能体的输出质量在很大程度上依赖学生输入质量。学生若能清晰表达学习目标、已知条件、问题本质与知识结构，智能体更可能输出高质量内容；反之，当学生需求模糊、提问无序、过度依赖、缺乏规则意识或存在关键概念偏差时，交互很容易走向低质量输出甚至误导性输出。第二，生成式人工智能存在“幻觉”

等风险：其生成内容可能表述流畅但事实或推理不可靠[5]，而高中数学具有强逻辑性与高可验证性，一旦产生错误步骤或概念混淆，会对学生认知结构造成明显干扰。虽然相关研究指出幻觉与深层逻辑缺失问题仍需持续解决[6]，但通过检索增强生成(RAG)将权威知识库与生成式 AI 结合，可在一定程度上抑制幻觉、提升输出可验证性[6]，这为县中数学场景提供了技术可行性。第三，课堂管理与过程性评价存在“不可见性”难题：在一个常态班级约 50 人的情境下，即便实现人手一台设备与智能体交互，教师也很难实时、完整地看到每位学生与智能体的对话过程，更难在同一时间对大量交互进行即时判断与反馈，从而造成“表面个性化、实则黑箱化”的教学风险。

1.3. 研究问题与研究思路

围绕上述现实堵点与应用难点，本研究提出以下研究问题：

问题 1：在县域普通高中高一数学教学中引入教育专用智能体，并配套教师引导与规则支架，是否与学生数学学习成效和自主学习表现(含学习策略与元认知监控)的改善相关？

问题 2：县中学生在教育智能体支持下的自学行为，是否呈现“新接触 - 适应 - 策略性规避 - 引导修正 - 良好习惯”的阶段演进？各阶段的典型问题与关键干预点是什么，教师如何通过可执行策略实现从“低质量交互”到“高质量自学”的转化？

为回答上述问题，研究采用“准实验对比 + 过程数据追踪 + 质性补充解释”的总体思路：以两个平行班为实验组与对照组，开展为期 6 周的教学干预，通过前后测数学一模及二模测验评估干预效果，同时基于系统后台自动记录的对话日志与部分学生访谈，对交互行为与对话质量进行分析，并用于佐证与补充阐释阶段演进模型与提炼可复制的干预策略。

本研究的预期贡献主要体现在三个方面：第一，在县中情境下提出并结构化呈现“师 - 智 - 生协同”的自学机制框架，将教育智能体从“答疑工具”转化为“学习支架系统”；第二，构建并检验县中学生教育智能体自学的“五阶段演进模型”，为教师识别学生使用风险、实施阶段性干预提供可操作的判定依据；第三，基于实验对照与过程性证据，给出具有一定操作性的干预策略与实施路径，为县域普通高中人工智能应用效能提升提供经验与初步实证参考。

2. 基于教育智能体的县域自学机制构建

县中场景下的教育智能体应用，不能停留在“学生自由使用”的工具层面，而必须被设计为一套能够在大班规模与资源限制条件下运行的学习机制系统。本研究提出“师 - 智 - 生协同的自学机制”，以“目标驱动 - 交互支架 - 反向输出 - 评价校准 - 过程治理”为核心链条，把自习课转化为可复制、可评估、可干预的学习闭环，并进一步提出“新接触 - 适应 - 策略性规避 - 引导修正 - 良好习惯”的五阶段演进模型。

2.1. 基于本地模型与本地权威知识的可信输出约束

高中数学智能体要在县中真实场景中发挥稳定效能，技术层面的关键不在于开放搜索与无边界生成，而在于“信息源治理与输出结构治理”，故采用本地大模型(DeepSeek 671B)+ 专业编辑校稿的本地教材精炼知识作为权威底座，并通过“信息源控制 - 结构化输出 - 质控闭环”形成三层可信约束。

在构建智能体引导下的教学系统时，首先需要对信息源进行严格控制。在教学场景中，系统关闭联网搜索功能，避免其在生成过程中调取低质量、非本地口径或信息污染严重的材料。此外，所有智能体的回答应限定在校本权威知识库中进行检索与生成，其信息来源必须包括本地教材、教师讲义、题型归纳库与易错点数据库等可追溯材料，从而确保输出内容具备可验证与可复盘的标准。

在智能体输出方面, 应强调结构化与可检验性, 而非冗长的篇幅, 针对解题类任务及知识讲解任务, 设计不同的固定结构予以调用。在解题类任务中, 应以“审题要点-解题思路-相关依据”为基本框架展开论述; 在知识讲解任务中, 应以“核心概念-概念比较-常见误区-机制理解”作为组织逻辑。

在保障智能体输出质量的同时, 还需配备专业编辑校稿团队, 并结合本地化教学内容形成数据底座。编辑团队将从教材内容、例题解析与教学易错点入手, 构建权威知识单元, 配套记录其来源与版本信息。此外, 系统后台同步记录交互轮次、提交文本、评分变化与修改痕迹, 供教师进行过程抽检与质量回溯。针对模型输出需进行合规性测试, 明确“输出红线”与回应策略, 例如禁止超纲答复、过度给出结论、引发学生误解的诱导性语言等情形, 并在发现违例时设定标准回应语句。

2.2. “师-智-生”协同的三大自学场域构建

在当前县中教育资源紧张、自习课通常由班主任维持秩序而非学科教师主导的背景下, 若要让教育智能体真正介入并提升学生自学效能, 必须构建一套由机制设计与职责分工共同支撑的协同生态系统。为此, 本研究将“师-智-生”协作体系划分为三类关键场域, 并通过“学习闭环”与“治理闭环”两条逻辑主线, 形成能够在常规教学环境中稳定运行的结构性方案。

在教师与智能体协作的第一个场域中, 教师承担着规则设计与质量标准建构的职能角色。每一节自习的目标需具备可操作性, 即必须限定在 60 分钟内能够完成的合理任务。教师还需为学生与智能体之间的提问与输出交互建立规范, 要求所有提问紧扣学习目标, 输出必须符合目标任务预设的要求。与此同时, 教师通过调取学生与智能体对话数据, 借助系统识别如提问频次、交互时长、轮次与评分等维度, 对低质量交互或风险行为进行阶段性识别与干预。

学生与智能体的交互构成了自学发生的第二场域。在该场域中, 学生围绕自主设定的学习目标, 与智能体展开深度交互, 包括提出问题、获取提示、澄清理解并记录可复用的学习笔记。智能体扮演策略支架的角色, 通过反向追问帮助学生压缩模糊目标为可解的问题。同时, 在“反向输出”环节中, 学生须使用自身语言对所学概念、步骤与策略进行复述, 从而完成一次以自我表达为核心的元认知校正过程。

第三个场域是教师与学生之间构成的治理维度。虽然教师不直接参与学科讲授, 但通过对学习秩序的管理与关键节点的过程抽检, 可以确保学生自学活动符合事前设定的规则框架, 并对出现策略性规避或长期低质量交互的学生进行定向干预。

2.3. 人智结合的三段式自学闭环设计

在当前县中现实条件下, 自习课的价值往往被消解为“时间占位”, 学生的自学行为缺乏机制支撑、教师又难以进行有效过程引导。为了让自习课由松散转向紧迫、由低效转向高效, 本研究将“自学”设计为可执行、可追踪的三段式结构化任务。

第一是目标设定阶段, 学生需在 3 至 5 分钟内明确写出本节课的学习目标。目标的设置需符合三个标准: 一是可在本节课完成, 二是可通过交互与反向输出得到验证, 三是贴合学生当前认知水平。

第二是进入目标驱动的交互阶段, 历时约 15 至 20 分钟。在这一过程中, 学生围绕既定目标与智能体开展深度交互, 包括获取关键概念的澄清、解题步骤的解释、条件的核对以及解法的验证提示。此阶段尤其强调提问的“围绕性”与“聚焦性”, 避免出现无关话题、碎片对话或漫无目的的信息拉取行为。

第三是触发反向输出评价阶段, 在课程中段节点, 所有学生统一向智能体发送“我明白了”的指令, 以此触发评价机制。学生需以自身语言对所学内容进行阐释, 智能体依据输出内容进行结构化评分与定点反馈。学生据此反馈进行修正并再次提交, 形成一个“评价-反馈-修正-再评价”的迭代回路, 直到获得满分为止。这一满分设定把学习评价从终结性判定转化为过程校准的内在机制。

2.4. 五阶段自学发展模型

县中学生在教育智能体支持下的自学行为并非线性提升, 而往往呈现阶段性演进。基于县中真实行为特征, 本研究提出的五阶段自学发展模型(见表 1), 除用于识别风险、配置策略与评价阶段性效果之外, 还要把“会不会用 AI”转化为“处于哪个阶段、该用什么策略”, 使教师在大班与非学科讲授条件下仍能实现低成本过程治理。

Table 1. “Five-stage” self-study development model and intervention strategies for county high school educational agent self-study

表 1. 县中教育智能体自学“五阶段”自学发展模型与干预策略

阶段	阶段判定特征(过程证据)	典型问题	关键干预动作 (教师可执行)	阶段退出条件
新接触阶段	目标随意; 轮次少; 评分偏低; 文本与笔记碎片化	把智能体当答案机; 提问无目标	明确目标写法; 规定提问边界; 示范一条高质量流程	目标可完成; 评分开始上升; 能按流程完成反向输出
适应阶段	轮次增加; 评分上升但波动大; 笔记开始成型	追问不深; 验证不足; 总结形式化	强化“自检点”; 要求输出“关键条件 + 关键依据 + 易错点”	评分稳定提升; 能在评价反馈下迭代修正
钻空子阶段	表面完成度高但过程证据异常; 文本呈机械改写/高重复; 评分提升慢或虚高	策略性规避理解; 形式化反向输出	抽检复盘; 要求用新例子解释; 改变输出顺序; 增加“变式检验”	文本出现个人推理痕迹; 评分稳定提高
引导阶段	评分较高且稳定; 达满分迭代次数减少; 轮次与质量匹配	深层概念仍混淆; 迁移不足	投放高阶任务(变式/一题多解); 同伴互评; 个别纠偏	能在新题中调用策略; 自检与纠错常态化
良好习惯阶段	目标稳定可完成; 评分长期满分或快速达标; 输出具迁移性	主要差异由合规与流程转向深度与广度	弱化外控规则; 沉淀范例库; 低频抽检保持系统运行	学习机制内化; 过程证据稳定、风险行为低发

3. 研究设计与实施

3.1. 研究对象与样本

本研究选取河北省某县域高中高一年级两个平行班作为研究对象, 分别为 101 班(实验班)和 103 班(对照班)。实验班在干预期内使用教育智能体开展数学自习学习, 对照班采用常规自习方式。

在成绩分析中的研究依据为可匹配的一模与二模成绩数据, 纳入有效样本共 86 人, 其中实验班 41 人, 对照班 45 人。由于部分学生成绩数据缺失或无法完成前后测匹配, 未纳入量化统计分析。根据前测独立样本 t 检验结果, 两组学生在干预前成绩差异不显著, 具有较好的基线可比性。

3.2. 研究设计

本研究采用准实验设计, 具体为前测 - 后测对照组设计。干预周期为 6 周。实验班在此期间接受“教育智能体 + 教师过程治理”的自习支持干预, 对照班维持常规自习安排。量化效果评估以前测(一模)与后测(二模)数学成绩为依据, 并结合对话日志、教师干预记录与学生反向输出文本, 对学生使用教育智能体的行为特征及其阶段性变化进行分析。

3.3. 干预措施

实验班干预措施: 在每节自习课中, 学生按照“目标设定 - 交互学习 - 触发反向输出评价”的三段

式流程使用教育智能体。教师在课前明确学习目标要求, 课中进行过程监督与个别指导, 课后调取系统后台数据进行质量评估与阶段性干预。为确保大班场景下干预执行一致、过程可追踪, 本研究同步引入分阶段的过程治理与反作弊规则(见表 2), 将关键节点、识别指标、处置规则与留痕方式固化为课堂制度。

对照班措施: 采用传统自习模式, 学生可自由选择学习内容与方法, 教师维持课堂秩序但不提供学科指导。

Table 2. Process governance and anti-cheating rules (early-middle-late stages)

表 2. 过程治理与反作弊规则表(早期 - 中期 - 晚期)

早期(制度建立期): 立规矩 + 留痕 + 断作弊				
治理模块	触发节点	可观察证据/识别指标	处置规则	留痕方式
测评节点制度化	学生触发“我明白了”进入评价节点	明确“测评节点”开始, 要求进入“反向输出 - 修正 - 再提交”闭环	进入测评节点后, 禁止继续绕开输出的追问式聊天; 偏离流程记为“无效对话/未完成”; 当堂要求回到闭环完成一次有效提交	后台记录: 触发时间、提交次数、评分变化; 教师端记录: 偏离类型与处理结果
复制粘贴反作弊(红线)	测评/点评节点	出现粘贴痕迹、文本高度一致、整段搬运	一经识别: 直接判定该次对话无效; 当堂要求重做: 必须用“自己的数学语言”重述; 纳入班级层面通报与重点抽检名单	无效对话名单(含时间、任务、学生编号); 后台相似度/粘贴行为标记
“改路径/搬运思路得分”反作弊	得到高分/满分后复核	逻辑路径异常跳步、与智能体原始输出高度同构但缺少个人解释痕迹	即便满分: 若识别为“修改智能体路径/搬运非本人思路”, 仍判无效; 要求重做: 补“关键步骤依据+自检点”; 必要时追加“变式检验/口头追问”	评分记录 + 对话路径审计; 抽检记录(追问题/变式题)
风险行为与质量维度标签化	全程(交互/测评/修正)	风险: 抄写/复制粘贴/搬运/改路径; 质量: 逻辑、数学语言、知识补充完整性、学习态度	风险标签触发→进入“人工复核/抽检”; 累计触发 ≥ 2 次→进入重点名单; 质量维度低于阈值→必须二次提交直至达标	后台标签(风险/质量) + 重点名单(每日/每周汇总)
技术保障治理	登录失败/无法进入系统	设备/账号异常导致无法学习	走“教师 - 研究员 - 管理员”处理链; 未解决前不计入“偷懒未上线”; 解决后补做当次任务	工单记录: 问题 - 处理人 - 解决时间 - 补做情况
修订留痕规则	二次提交/补充说明	学生修改后未标注修改点	二次提交必须包含“以下为我修改/补充”标记; 无标记视为不可追溯, 不计有效迭代; 反复无标记者列入重点抽检	对话日志对比: 修改前/后差异; 教师端标注“未留痕”
上线监控	每日/每节课结束	未上线、上线时长异常短	先核验技术原因; 确认行为未达标→判定“未完成当日任务”; 纳入次日重点关注并进行补做	登录记录、在线时长; 未上线名单与补做记录
中期(质量提升期): 抓质量维度 + 强对比举例 + 促迭代达标				
治理模块	触发节点	可观察证据/识别指标	处置规则	留痕方式
比较能力治理	测评节点评分后(未达标)	输出仅描述概念, 缺少“对照对象/差异点/适用场景”	要求补写“参照系 + 差异点 + 适用场景”; 至少给出 1 个反例/对照情境; 补完再提交评分	评分反馈字段: 缺陷 = 比较缺失; 二次提交文本
举例能力治理	测评节点评分后(未达标)	无具体例子或例子与概念无关	必须补 1 个“贴题例子”(优先)或“生活例子”; 例子需体现概念边界/公式条件; 例子不合格继续退回	对话中出现“例子段落”; 教师端“例子合格/不合格”

续表

深层解释治理	测评节点评分后(未达标)	只复述条件, 缺少原理/推导/几何意义	要求补“原理/推导依据/几何意义”; 至少写出一个关键步骤为何成立; 必要时加“自检点”	反馈标签: 依据不足/推导缺失; 再提交记录
数学语言精准度治理	测评节点评分后(未达标)	术语不严谨(如概念表述混乱)	三段式改写: 原句 - 问题 - 修正句; 同类错误累计 ≥ 2 次 \rightarrow 进入重点纠偏名单; 下次测评节点优先抽检语言	语言错误清单(可附录); 重点纠偏名单
公式背后原理治理	测评节点评分后(未达标)	只给结论, 不说明来源/适用条件/误用	要求补: 公式来源/适用条件/典型误用; 必须给出“能不能用”的判断; 判断缺失退回	输出结构中单列“适用条件/判据”; 评分维度记录
阶段性达标判定	每日结束/节点结束	“起初存在缺陷 \rightarrow 补充修正 \rightarrow 完成目标”	将“完成今日学习目标”作为过程性结局变量; 未达标者次日补交; 达标者进入变式任务/更高阶任务	每日达标名单、补交名单; 教师端干预记录

晚期(巩固深化期): 维持低风险 + 强化依据链 + 促进迁移

治理模块	触发节点	可观察证据/识别指标	处置规则	留痕方式
深度比较与联系治理	测评节点评分后(未达标)	能说概念但缺少联系/区别与迁移解释	补“联系 - 区别 - 何时用 A/何时用 B”; 必须出现至少 1 条迁移解释; 迁移缺失不达标	反馈标签: 联系不足/迁移不足; 二次提交记录
概念 - 题目映射治理	测评节点评分后(未达标)	概念解释与题干条件脱节	强制输出“概念要素 \rightarrow 题干条件 \rightarrow 求解目标”的映射链; 映射不完整退回重写	输出模板新增“与本题的联系”; 评分记录
例子补强治理	测评节点评分后(未达标)	例子缺失或只给单例	例子必须与判别条件一一对应; 同时给“满足/不满足”两类; 无反例不达标	例子对照表(可作为作业附件)
依据链完整性治理	推导/代入步骤出现跳步	步骤存在但缺依据说明	强制三段式: 做什么 - 依据是什么 - 服务哪个目标; 缺依据退回; 反复缺依据者重点抽检	评分维度增加“依据完整性”; 抽检记录
计算准确性治理	结果错误/计算失误	计算错误导致结论偏差	要求补“自检点”(代入验证/特殊值检验等); 修正后说明错因; 同类错误高发者纳入纠错清单	错因标签(可用于后续分析); 纠错清单
课程退出与长期治理	课程末/干预收束	学生能基于预习笔记提问题	从“强制节点抽检”过渡为“预习笔记问题-AI 问答 - 教师低频抽检”; 反作弊红线保持不变	周度抽检记录; 长期低频治理日志

3.4. 数据收集

本研究同时收集量化与质性两类数据。量化数据为实验班与对照班学生的一模和二模数学成绩, 用于检验干预效果。质性数据包括实验班学生与教育智能体的对话日志、教师在课堂中的过程干预记录, 以及学生在“反向输出”节点提交的文本材料。

在干预期间, 实验班共形成学生与智能体对话数据 744 条。为进一步分析学生在教育智能体支持下的自学行为演进特征, 研究从实验班中选取 10 名学生的代表性日志作为重点样本, 结合其提问内容、修改痕迹与教师干预记录进行追踪分析。

3.5. 数据分析方法

量化数据分析采用 SPSS 25.0 进行处理。首先, 对实验班和对照班的一模、二模成绩进行描述统计,

计算均值与标准差;其次,采用独立样本 t 检验对两组前测成绩进行基线差异检验;最后,采用 2 (时间:一模、二模) $\times 2$ (组别:实验班、对照班)的重复测量方差分析检验干预效果。

质性数据分析采用主题分析法。研究先对实验班对话日志进行匿名化整理,再结合教师干预记录和学生反向输出文本进行开放编码,重点识别与“目标表达”“提问深度”“概念理解”“形式化完成”“策略性规避”“教师纠偏”“迁移解释”等有关的行为特征;在此基础上,对相近编码进行聚合,并结合时间顺序考察其演进轨迹,最终形成“新接触-适应-策略性规避-引导修正-良好习惯”的阶段性发展模型。

4. 研究结果

4.1. 一模与二模成绩的描述统计及基线差异检验

实验班与对照班一模、二模成绩的描述统计结果见表 3。可以见到,实验班成绩均值由一模的 57.34 分上升至二模的 61.15 分,而对照班成绩均值由一模的 60.09 分下降至二模的 55.38 分。为检验两组学生在干预前是否具有可比性,对一模成绩进行独立样本 t 检验。结果显示,实验班与对照班在一模成绩上的差异不显著, $t(84) = -0.780$, $p = 0.437$,表明两组在干预开始前具有较好的基线一致性。

Table 3. Descriptive statistics of the first and second mock exam scores for the experimental class and the control class
表 3. 实验班与对照班一模、二模成绩的描述统计

组别	n	一模 $M \pm SD$	二模 $M \pm SD$
实验班	41	57.34 \pm 16.06	61.15 \pm 14.54
对照班	45	60.09 \pm 16.54	55.38 \pm 18.21

4.2. 重复测量方差分析结果

在基线具有可比性的前提下,采用 2 (时间:一模、二模) $\times 2$ (组别:实验班、对照班)的重复测量方差分析考察干预效果。结果显示,时间主效应不显著, $F(1, 84) = 0.070$, $p = 0.793$, $\eta^2 = 0.001$,说明整体样本在一模与二模之间未表现出一致方向的显著变化;组别主效应不显著, $F(1, 84) = 0.237$, $p = 0.628$, $\eta^2 = 0.003$,说明从整体平均水平来看,两组成绩无显著差异;但时间 \times 组别交互作用显著, $F(1, 84) = 6.145$, $p = 0.015$, $\eta^2 = 0.068$,表明两组学生在成绩变化趋势上存在显著差异。结合均值变化可见,实验班成绩呈上升趋势,而对照班成绩呈下降趋势,提示教育智能体支持下的自学干预可能与实验班成绩变化趋势的改善有关。

4.3. 过程性证据对阶段模型的支撑

除成绩变化外,过程性资料还显示,学生在与教育智能体交互过程中提出问题的质量存在明显差异,并呈现出阶段性变化特征。部分学生的问题停留在“如何直接得出答案”的层面,而部分学生则逐步转向追问数学依据、方法边界与知识联系。例如,有学生追问“求异面直线夹角时,为什么结果范围要控制在 $0^\circ \sim 90^\circ$ ”“方向向量与法向量的分工边界是什么”“为什么点到平面的距离可以转化为法向量与数量积的计算”“圆的标准方程和一般方程分别适用于哪些场景”等。这类问题已不再局限于求解步骤,而是进一步触及概念原理、方法比较与结构化理解。

从 744 条对话数据及 10 名学生重点日志样本来看,学生在教师规则引导与智能体支架支持下,提问方式逐步由答案导向转向依据导向、比较导向和迁移导向。该过程性证据为“五阶段自学发展模型”中“引导修正”与“良好习惯”阶段提供了支持。

5. 讨论

5.1. 干预效果的可能机制

本研究结果显示, 实验班与对照班在一模与二模成绩变化趋势上存在显著差异。结合过程性资料来看, 这种差异可能不仅与智能体提供的即时反馈、概念澄清和结构化支架有关, 也与教师在课堂中的规则设定、节点抽检和过程纠偏有关。换言之, 本研究所观察到的效果, 更可能体现为“教育智能体支持+教师过程治理”的综合作用, 而不宜简单归因于智能体本身。

5.2. 潜在混淆因素与研究限制

需要注意的是, 本研究结果仍可能受到若干潜在混淆因素的影响。首先, 实验班学生处于新型教学情境中, 可能因“被关注”而在短期内提高学习投入, 这种霍桑效应可能对成绩变化产生一定影响。其次, 实验班教师在干预过程中承担了目标规范、日志抽检、过程反馈和个别纠偏等工作, 教师额外关注本身也可能促进学生学习行为改善。因此, 本研究更适合说明一种综合性课堂干预机制的潜在价值, 而不是单独证明某一技术工具的独立效果。

5.3. 实施中的现实挑战

此外, 该方案在推广中仍面临若干现实挑战。其一, 教师需要投入额外时间进行目标规范、日志抽检与个别纠偏, 这对教师的时间分配和专业判断提出了更高要求。其二, 智能体系统的稳定运行依赖持续的技术维护, 包括模型调用、知识库更新、账号管理和设备保障等。其三, 不同学生在数字工具使用能力、规则意识与自我监控水平上存在差异, 这可能导致干预效果呈现较大的个体差异。总体而言, 该方案更适合被理解作为一种需要教师参与、技术支持与课堂制度共同保障的复合型实施模式。

6. 结论

本研究围绕县域普通高中高一数学自习场景, 构建了“目标设定-交互学习-反向输出评价”的教育智能体支持机制, 并结合教师过程治理与分阶段规则设计, 探索了大班条件下教育智能体辅助自学的实施路径。基于可匹配的一模、二模成绩数据进行分析后发现, 实验班与对照班在成绩变化趋势上存在显著差异, 提示该干预方案在县域普通高中数学自习场景中具有一定应用潜力。

同时, 结合 744 条对话数据、10 名学生重点日志样本以及教师干预记录的分析, 研究初步提炼出“新接触-适应-策略性规避-引导修正-良好习惯”的阶段性演进模型, 为识别学生使用风险与实施分阶段干预提供了操作框架。

需要指出的是, 本研究结论仍可能受到霍桑效应、教师额外关注、样本范围有限及技术实施条件等因素影响。因此, 相关发现仍需在更大范围、更长周期和更多学科场景中进一步验证。

参考文献

- [1] http://www.moe.gov.cn/jyb_xwfb/s271/202509/t20250926_1415032.html, 2026-05-09.
- [2] http://www.moe.gov.cn/srcsite/A06/s7053/202112/t20211216_587718.html, 2023-01-03.
- [3] 胡娟, 陈嘉雨. 怎样理解高等教育普及化进程中的“县中困境”——基于高等教育类型结构变迁的探讨[J]. 教育发展研究, 2023, 43(2): 1-10.
- [4] 郑富, 李宛濛, 孙婉宁. 初中数理智能体的构建与应用[J]. 科学咨询, 2025(3): 163-167.
- [5] 梁昭. AI“幻觉”: 认知困境、术语反思与范式嬗变[J]. 民族学刊, 2025, 16(8): 82-87+161.
- [6] 杜修平, 王崑羽. 检索增强生成赋能智能导学系统构建研究——基于本地大模型与私有知识库[J]. 中国电化教育, 2025(5): 117-127.