

从观察到洞察：基于多智能体的人机协同课堂循证教研系统构建

李 玲 刘 静 王益卫 彭 源 姜元昊 周爱民

(华东师范大学 上海智能教育研究院, 上海 20062)

【摘要】 智能课堂观察可自动识别课堂教学的多维特征,为教研提供实时、精准、丰富的数据补给,但由于其分析结果缺乏情境化解读,反馈形式僵化,与教研话语体系脱节,未能与教研实践深度融通。本研究依托大模型与多智能体技术,构建人机协同的课堂观察与循证研究系统,赋能教师开展基于教学情境的个性化数据探究。系统在自动提取课堂多维特征的基础上,实现了三大突破:1)整合“课后反思”“同课异构”和“行动研究”三类典型教研场景的核心数据挖掘任务,并结合检索增强生成技术构建动态知识库,实现数据分析与教研语境的适配;2)利用大语言模型解析教研意图,将教师的非结构化提问拆解为可操作的教研问题 and 数据查询指令,推动自然语言交互与数据分析的无缝衔接;3)多智能体与分析工具自动化协同执行数据调用、分析及可视化等重复性任务,使教师能专注于高阶任务,即对课堂数据的解读、洞察与决策。初步研究表明,该系统能够解析教师基于情境的提问,引导教师将描述性的课堂观察数据挖掘解读为对教学规律、行为模式和教学策略的洞察。这一人机协同的知识共创机制实现了课堂观察从静态数据反馈到动态知识生产的范式跃迁,为人工智能支持下的循证教研开辟了新路径。

【关键词】 智能课堂观察; 循证教研; 大语言模型; 多智能体; 知识共创

【中图分类号】 G434

【文献标识码】 A

【文章编号】 1007-2179(2025)05-0101-10

一、引言

课堂观察作为教学改进的重要依据,其智能化转型已成为教育领域的研究热点。当前,基于计算机视觉、语音、自然语言处理等技术的智能课堂观

察系统能够自动识别课堂的多维教学特征,一定程度上解决了人工观察耗时费力、依赖专家、时效性低且观测视角单一的问题,可为教师日常教研提供实时、精准、丰富的数据补给(Karsli et al., 2024; Hoyos et al., 2020; Waheed et al., 2018)。然而,已有

【收稿日期】2025-07-20

【修回日期】2025-08-18

【DOI编码】10.13966/j.cnki.kfjyyj.2025.05.011

【基金项目】 华东师范大学人工智能赋能人文社科研究范式转型项目“多智能体驱动的循证教研范式转型与系统构建研究”(2025ECNU-AI006)。

【作者简介】 李玲,博士,副研究员,华东师范大学上海智能教育研究院,研究方向:智能课堂观察、循证教研、教师发展;刘静、彭源,博士研究生,华东师范大学上海智能教育研究院,华东师范大学计算机科学与技术学院,研究方向:智能课堂观察;王益卫,硕士研究生,华东师范大学上海智能教育研究院,华东师范大学计算机科学与技术学院,研究方向:智能课堂观察;姜元昊,博士研究生,华东师范大学上海智能教育研究院,华东师范大学计算机科学与技术学院,研究方向:智能教育、大语言模型及其教育应用、教育智能体;周爱民,博士,教授,博士生导师,华东师范大学上海智能教育研究院,研究方向:人工智能、智能教育(amzhou@cs.ecnu.edu.cn)。

【引用信息】 李玲,刘静,王益卫,彭源,姜元昊,周爱民(2025). 从观察到洞察:基于多智能体的人机协同课堂循证教研系统构建[J]. 开放教育研究,31(5): 101-110.

研究发现, 智能课堂观察系统与教研在实践层面并未深度融通, 体现在分析结果缺乏课堂语境、反馈呈现形式僵化、与教研实践话语体系脱节, 难以成为日常教研的有效支撑(郁晓华等, 2024; 张治等, 2024; Saar et al., 2022)。这一困境本质上源于智能系统与教研实践在价值取向和运作逻辑上的差异: 前者侧重输出标准化、通用、静态的数据报表, 后者侧重课堂情境的个性化、动态化聚焦。值得注意的是, 当前以数据为中心的建设思路弱化了教师在教研中的主体地位, 使其成为被观察、被量化、被评价、被建议的客体。

为弥合智能课堂观察与教研实践的鸿沟, 本研究依托大模型与多智能体技术, 构建人机协同的课堂观察与循证研究系统, 赋能教师自主开展课堂数据探究, 将静态数据反馈转变为知识与智慧。该系统以多源异构的课堂数据为基础, 以典型教研场景为框架, 以大语言模型为中枢, 以教研问题为线索, 以智能体为执行载体, 实现从数据采集到知识生产的全链条支持。这一技术架构旨在促进教师与课堂数据间的深度交互, 破解传统工具数据生产与知识转化之间的断层问题, 并重塑教师在智能化教研进程中的主体地位。

二、智能课堂观察: 局限与突破

(一) 现状回顾

20 世纪初, 研究者开始对课堂教学进行量化、系统化、结构化的观察与分析(O' Leary, 2020)。教师的行走移动、站位、话语、互动模式等都被研究证实会对学生及其学习产生影响(Ahuja et al., 2019)。20 世纪 50 年代, 大量课堂观察方法与工具涌现, 如弗兰德斯言语互动分析法、课堂评估评分系统(classroom assessment scoring system, CLASS)、本科 STEM 课堂观察量表(classroom observation protocol for undergraduate STEM, COPUS)从多个维度提炼课堂特征。研究表明, 课堂观察为教师提供了结构化的分析反馈, 对课堂把控、提升教学质量和学生成绩产生了成效(Ahuja et al., 2019; Ngoon et al., 2024)。然而, 人工课堂观察耗时费力且依赖专家, 难以支持常态化、规模化的教研工作。

随着视觉、语音和文本智能分析技术的高速

发展, 智能课堂观察应运而生。智能技术能自动采集课堂教学的多模态数据, 利用不同模态数据之间的信息互补机制精准识别课堂特征, 还原教学全貌(王一岩等, 2022)。近年, 国内外智能课堂观察研究涵盖课堂活动类型(Ngoon et al., 2024; 李玲等, 2023; 杨晓哲, 2021)、提问策略(Pérez-Sanagustín et al., 2012)、课堂氛围(Alfredo et al., 2025)、行动轨迹(Alfredo et al., 2025; 李玲等, 2023)、学生参与度等(Aslan et al., 2019), 极大拓展了课堂观察的范围, 为教师审视课堂教学提供了多元、穿透性的视角。自此, 基于智能课堂观察的“智能教研”“大数据驱动精准教研”“智能循证教学”等概念相继被提出。其核心观点在于智能课堂观察输出的数据反馈可作为“证据性”资源, 支持教师教研的反思与决策, 推动教研向数据化、科学化、精准化方向发展。

然而, 有研究指出, 在实践层面, 智能课堂观察与教研之间并未深度融通。首先, 智能分析提供的反馈观测指标往往独立于教学语境, 如提问次数、学生参与度、师生行为占比等, 难以反映教学方法的优劣(郁晓华等, 2024; 张治等, 2024)。当反馈缺乏可操作性和可解释性时, 教师采纳意愿往往较低(Alfredo et al., 2025; Ngoon et al., 2024)。其次, 教师教研提出的问题是场景化、个性化、动态螺旋聚焦的(Ma et al., 2023; Saar et al., 2022)。例如, 教师常会从宏观的数据分布切入(全班答题准确率呈何种分布特征), 进而锁定关键群体(哪些学生持续作答错误), 最后向教学场景纵深追问(学困生课堂行为有哪些特征)。然而, 当前课堂观察报告分析结果大多呈标准化、通用、静态的, 难以满足动态分析要求。最后, 智能课堂观察与教研之间话语体系不一。前者输出的分析反馈往往是描述师生的行为, 如“教师行动轨迹”“学生发言时长”, 后者关注教学过程、效果、规律的探究, 如“学生学习遇到哪些困难?”“新的干预手段能否提升学生参与度?”“成绩优秀学生的课堂行为有哪些特征?”话语体系的差异使课堂分析输出结果无法与教研需求直接对齐(Alfredo et al., 2025; Ngoon et al., 2024)。上述问题导致智能课堂观察难以作为有效的协同工具支撑教研开展。

(二)智能系统与教研实践的融合路径

智能课堂观察与教研实践之间的鸿沟,本质源于二者运作逻辑与价值取向的差异。前者通过第三方视角客观分析课堂数据,生成标准化的观察报告;后者是教师立足实践开展反思性探究,最终形成情境化的教学洞察。

洞察(insight)是超越表层观察的认知跃迁,能实现对问题本质、内在模式和关联现象的深刻把握,进而促进个体认知框架的革新与重构(Yi et al., 2008)。从生成机制看,基于数据的洞察具有三个典型特征:1)产生于个体与数据的深度交互,通过动态调整数据维度(如粒度调节、信息筛选)和主动识别潜在模式(如差异比较、趋势分析)实现认知积累(Card et al., 1999);2)依赖心智模型匹配(match mental model),即将可视化表征与领域知识相结合,从而产生情景化理解(Agutter et al., 2001);3)高质量的洞察往往具有意外性和创造性(North, 2006),表现为对既有认知框架的超越或重构。

在教育研究领域,基于课堂的教研本质上是系统的洞察生成过程。这一过程遵循“问题驱动—迭代探究—知识建构”的循证研究范式(张治等, 2024; 吴雨宸等, 2023),致力于实现从实践经验到教育智慧的价值跃迁。这一过程始于教师对教学现象的观察,继而收集数据、对比差异、关联分析和识别规律等验证既有假设或补充原有认知(Mandinach et al., 2016)。这种知识生产遵循“数据→信息→知识→智慧”的递进逻辑,其核心在于对教学数据的深度解读与意义建构。然而,当前智能课堂观察系统多止步于呈现基础数据,未能构建支持教师个性化探究的认知支架。

上述问题随智能技术的发展愈发凸显:一方面,技术侧重通过算法优化不断扩充可观测维度和观测指标,但指标体系的扩容并未同步提升教研支持效能,反而导致教师面临“数据丰裕而洞见匮乏”的信息过载困境。近年涌现的教研实践研究表明,教师群体正通过二次加工、跨周期整合、情境关联等深度分析方法,主动延展智能课堂观察系统的数据价值链条,如对比课例研磨前后课堂数据(江雪贞, 2024)、识别长周期课堂数据中的教学趋势(高德模, 2024)、师生对话深度和课堂教学环节之间

的关联分析(魏宁, 2024)。这些实践不仅印证了一线教研对结构化课堂数据的再挖掘需求,更折射出现有智能分析系统支持教学循证教研存在“数据生产—知识转化”的断层问题。

要突破这一难题,需要智能课堂观察系统构建教师与数据的深度对话机制,开发支持教师自主开展数据挖掘、分析与解读的探究环境。为此,本研究引入大语言模型与智能体技术,通过嵌入教研知识框架,提升系统在教研垂直领域的智能交互能力,实现系统从“数据看板”到“教研伙伴”的转型,实现从课堂“观察”到“洞察”的认知跃迁。

三、构建人机协同的课堂观察与循证教研系统

(一)理论框架

循证实践(evidence-based practice)形成于20世纪70年代的医学领域,强调临床决策应基于严谨的科研证据,而非个人经验或传统做法(Sackett et al., 1996)。随着该理论在管理学、心理学等领域的应用与完善,逐渐形成普适性的循证决策方法论。在教育领域,哈格里夫斯(Hargreaves, 1996)首次论证了循证理念对教育变革的价值,主张教学决策应建立在科学证据而非直觉或惯例的基础上。这一理念不仅重构了教师的角色定位——从被动的研究对象变为主动的证据应用者与生成者,更形成教学方法的创新范式。通过实践证据与教学智慧的动态交互,循证教研能够不断催生本土化、情境化的创新教学策略。教学知识因此得以动态更新,理论与实践相互滋养,最终形成良性循环的教育创新生态(吴雨宸等, 2023)。

国内外学者根据不同的实践情境和政策背景,提出了循证教研划分阶段框架(Davies, 1999; 陈峥等, 2022)。这些框架包含五个环节:1)问题界定,即识别教学情境中具有研究价值的实践问题;2)收集数据,即获取课堂观察、学生表现等多元数据;3)分析数据,即完成数据清洗、分析与可视化;4)解读数据,即将分析结果与教学知识融合,提炼有指导价值的信息,形成改进方案;5)干预验证,即实施改进方案并评估效果,继而生成新的研究问题。

然而,循证教研的有效开展对教师专业素养提

出较高要求。国内外研究表明,教师在循证教研实施的每个阶段都面临挑战:教师提出的教研问题往往模糊宽泛,难以将实践问题精准转化为清晰聚焦且可操作的教研问题;数据采集所需的专业设备与技术操作对教师提出严峻挑战;复杂的数据分析工具超出大多数教师的技术能力范围(Mandinach et al., 2016);教师只有将分析结果与教育知识深度融合,才能实现从证据到策略的有效转化(Gummer et al., 2015)。这一能力缺口在传统教研模式下难以得到有效弥补。

“基于大语言模型的多智能体系统”(large language model-based multi-Agent systems)为攻克这一难题提供了新思路。系统可将复杂任务拆解为多个子任务,通过多个自治智能体互动和协同完成复杂任务(Janbi et al., 2023; 翟雪松等, 2024)。其自主决策、动态交互与资源共享特性,有望精准解决教师课堂循证教研存在的能力缺口,实现教研各环节的高效协同与深度推进。

本研究将多智能体的协同机制嵌入课堂循证教研的知识生成过程,以多智能体系统的分布式架构与交互逻辑为核心脉络,将教师的专业洞察与传统智能课堂观察的特征提取能力、大语言模型的语义理解能力、教研知识库的认知增强能力、数据分析工具的量化分析能力等技术模块进行动态耦合,形成“有理、有据、有思”的交互式知识生成链条。这一创新探索将人机协同机制引入智能课堂观察领域,推动智能课堂观察工具实现从“静态数据反馈”到“动态知识生产”的范式跃迁。

(二)建设目标

本研究基于人机协同理念,构建智能课堂观察与循证教研系统——ClassInsight,旨在解决教师循证教研实践面临的挑战。该系统融合多模态数据分析、大模型与多智能体等前沿技术,同时整合循证教研的理论范式与实践场景,构建自然语言交互环境的智能化教研支持平台。具体来说,该系统集成了两个相互关联的模块:1)智能课堂观察模块,它通过多模态数据分析与融合,实现对多维课堂特征的实时解析与可视化呈现;2)智能循证教研模块,它基于自然语言交互界面,实现从教研意图解析、数据挖掘到智能解读的一站式循证教研支持。下

文将系统阐述各模块的技术框架与实现路径。

(三)智能课堂观察模块的设计与实现

智能课堂观察模块的核心在于构建支持教学实践反思的数据基础。本智能课堂观察系统采用三层级技术架构(见图1)。

在文本分析方面,系统基于大语言模型,结合自然语言处理技术,构建课堂话语文本的多维度特征提取框架,通过 few-shot learning 策略设计动态提示模板,实现知识点自动抽取、疑问句识别、关键词提取、情感词汇分析等功能。在日志数据处理方面,系统通过描述性统计、聚类分析、时序分析等方法,挖掘学生与数智教材平台的交互特征。

标准报表层统计分析模型层输出的结构化数据,参考课堂观察领域的前沿成果,构建4个维度的30余项指标(见图1)。教学内容分析维度包括课堂语音转写、智能大纲生成、主题脉络生成和核心概念提取等,旨在辅助教师审视课堂教学内容的系统性与完整性。教学策略分析维度包括教师行动轨迹识别(Alfredo et al., 2025)、教师头部朝向识别、积极/消极词汇识别(Ramakrishnan et al., 2021)、课堂提问次数及类型(Jensen et al., 2020)等。研究显示,上述教学策略对课堂氛围、学生心理、专注度、参与度有深刻影响。教学方法分析维度涵盖教学活动类型、课堂交互层面(李玲等, 2023)、S-T曲线与 Rt-Ch图、师生活动占比等,旨在还原课堂教学形态。教学效果分析维度包括学生抬头率动态曲线、举手频次、学习资源交互(Ahuja et al., 2019; Aslan et al., 2019)等体现参与度的指标,还包括作答准确率等体现知识掌握度的指标。此维度旨在辅助教师联动思考教与学的内在关联。需强调的是,本指标体系旨在客观呈现课堂教学特征,不评价教学优劣,只是辅助教师自主评估教学有效性,为后续数据挖掘提供基础数据。

(四)智能循证教研模块的设计与实现

智能循证教研模块整合基于大语言模型的多智能体系统、教研知识库、检索增强生成技术、多模态数据库、数据分析与可视化工具等核心技术组件,构建面向课堂教研场景的交互式智能数据探究链路(见图2)。该模块采用中央协调式架构,由 workflow 调度智能体,统一负责与教师的交互沟通及

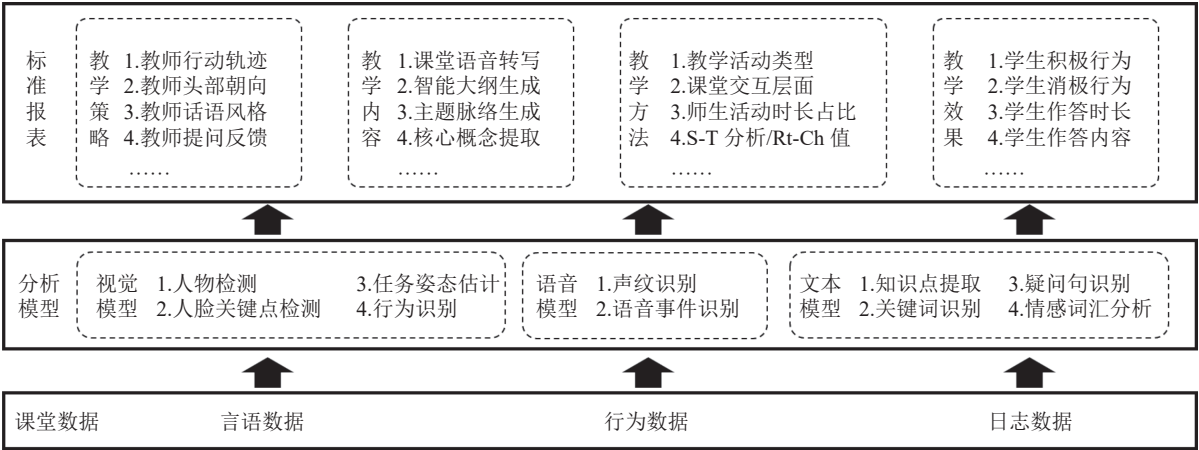


图 1 智能课堂观察模块的技术架构

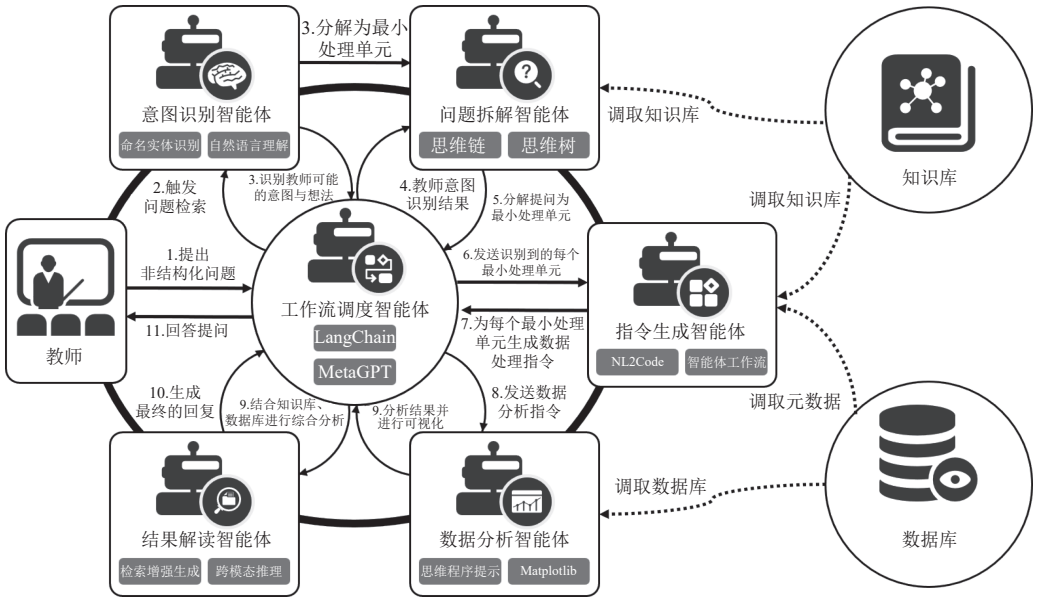


图 2 智能循证教研模块的技术架构

向五个功能智能体分发任务,具体流程如下:

首先,教师提出非结构化教研查询(例如,最近三个学期学生的课堂参与度呈何种趋势?),随后, workflowscheduling智能体调用 LangChain(Asyrofi, 2023)、MetaGPT(Hong, 2023)等智能体协同框架处理这一查询指令,并将该问题分发至意图识别智能体。该智能体先通过检索增强生成技术从教研知识库调取相关语义模式,再依托命名实体识别与自然语言理解技术解析教师查询内容——通过识别“学期”“学生”“课堂参与度”等关键实体,结合教育领域语义规则梳理逻辑关系,最终将模糊的原始查询转

化为结构化问题集合,并通过 workflowscheduling智能体反馈给教师(Li et al., 2020; McShane, 2017)。教师选定目标问题后,任务将被分配至问题拆解智能体。该智能体同样结合教研知识库的支撑,采用思维链与思维树技术,将教师的核心提问拆解为可独立处理的最小单元(Wei et al., 2022; Yao et al., 2023)。针对每个拆解后的最小单元,指令生成智能体在教研知识库的增强支持下,通过自然语言到代码(NL2Code)转换技术与智能体 workflow 协同机制,自动生成标准化的数据处理指令(Bai et al., 2025; Dai et al., 2024)。接下来,数据分析智能体依据生成的

指令开展数据调用与分析工作。该智能体集成数据库接口、专业数据分析工具(如 Python 数据分析库)及可视化工具(如 Matplotlib), 通过思维程序提示(thought process prompting)技术实现数据的调用、分析及可视化(Chen et al., 2022)。最后, 结果解读智能体结合课堂观察数据与循证教研理论框架, 阐释分析结果, 并向教师返回包含数据证据、趋势洞察及教学改进建议的综合性反馈。

这一人机协同的数据探究链路包含两项特有的技术组件: 一是团队自主研发的中国版教育大模型——启创•InnoSpark。它用于支持智能体运行。该模型基于大规模教育领域语料训练而成, 覆盖 400 余部教育经典教材、1.5 万册专业著作、100 多万篇高水平科研论文及 40 余门核心教育学课程资源, 形成教育领域专属的知识图谱与语义理解能力, 可赋能多智能体在循证教研场景下为教师提供专业化、个性化支持。二是作为核心组件的教研知识库, 它能够通过检索增强生成技术赋能多智能体, 使其能够准确理解教研场景的工作逻辑与话语体系。本研究选取三种最常见的教研场景——课后反思、同课异构和行动研究, 通过文献分析和教师访谈, 系统梳理各类场景的目标定位、典型教研问

题, 以及对应的可量化观测指标和数据查询指令, 并基于此构建教研知识库(见表 1), 进而提升智能体在教研意图理解、问题拆解及查询指令生成方面的能力。

课后反思是最常见的教研实践, 指教师课后根据实际教学情况评估课堂教学是否达到预期、教学难点是否突破、教学环节是否符合预设时长、教学容量是否适度等。此类教研活动的产生源于课堂教学的设计与实施之间的偏差。这种偏差既来源于教师自身教学信念与实际教学水平之间的偏移, 如教师课堂把控能力不足以支撑教学设计的有效实施, 也来源于理想与真实课堂环境之间的偏移, 如学生的参与度、学习能力和先验知识的欠缺导致预先设计的目标、活动、流程不能顺利达成(Li et al., 2022)。

同课异构是一种群体性、研讨式的教研活动。教师可通过观察其他教师的教学设计、实施, 识别自己的优势与不足, 达到优化教学策略的目的。这种教师间基于证据的探讨有助于推动教学共同体的成长(袁永民, 2011)。然而, 传统的同课异构教研活动受限于人力资源投入与操作成本, 仅能实现单节课的片段化观察。这种局限衍生出双重困境:

表 1 教研知识库

非结构化问题	数据探究流程	可量化观测指标	答案准确率	课堂活动类型	课堂交互层面	师生活动比例	教师行为占比	师生行为转换率	学生抬头率	提问次数	学生发言时长	积极 / 消极词汇	学生举手率	头部朝向	...	数据查询方法	数据合并	数据比较	关联挖掘	趋势分析	...
			×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×
课后反思	教师探究问题	教学目标是否达成?	×														×				
		教学难点是否突破?	×														×				
		教学环节是否符合预设时长?		×													×				
		教学容量是否适度?		×													×				
同课异构																					
		两位教师在课堂活动编排上有何异同?		×														×			
		两位老师的课堂教学形态有何异同?					×	×										×			
		两位教师在课堂交互层面上有何异同?			×													×			
行动研究		两位教师在师生交互活跃度上有何异同?						×										×			
		两位教师在课堂师生活动比例上有何异同?				×															
																					
		教师丰富了课堂行动轨迹后是否提升了学生的抬头率?							×	×									×		
		教师增加了提问次数后是否提升了学生的参与度?								×	×								×		
		教师提升了积极词汇后是否提升了学生的参与度?										×	×						×		
		教师增加了与学生的眼神交流后是否提升了学生的专注度?							×					×					×		
		教师优化了某个环节后是否提升了学生对知识的把握?	×																	×	
		教师优化了某个环节后是否提升了学生的参与度?								×			×							×	
		教师优化了某个环节后是否提升了学生的专注度?							×											×	
																					

首先,知识建构与素养培育有显著的时序性与递进性特征,需依托长周期教学设计框架。单一课堂的局部观察仅能捕捉孤立的教学片断,难以透视教师在大单元整体目标导向下对知识和能力培养阶梯式的规划。其次,学习本质上是隐形且非线性的认知建构过程(白倩等,2022)。单节课的片段化观察,仅能获取学生表层的行为特征,难以捕捉学生长周期知识与能力的发展轨迹,教师也难以获取全面的学习数据来佐证教学效果。

行动研究作为一种实践导向的研究方法,通常由教育工作者自主发起,旨在通过系统性探究生成可操作的解决方案以应对实际教学问题。例如,如何改善学生注意力分散现象?如何构建有效的小组竞争机制以提升学生参与度?研究者基于真实教学情境,遵循“设计—实施—评估—改进”的循环模式解决实际问题。该方法强调教师作为实践者与研究者的双重角色,体现“为行动而研究(research for action)、在行动中研究(research in action)、由行动者研究(research by actors)”的核心特征。基于课堂情境的行动研究往往需要探究多重教学要素之间的动态关联,或长周期追踪特定干

预措施对学生课堂行为的影响。智能课堂分析系统能同步采集多模态教学数据(如语音、表情、姿态、互动频率等),为教师全面审视课堂教学各要素间的内在关联、开展行动研究提供基础。

四、有效性验证

为验证智能课堂观察与循证教研系统的实践价值,本研究从6所学校选取40名一线教师作为研究对象:三所市级教育数字化转型示范学校18人(小学组N=5;中学组A校n=7、B校n=6),三所区级普通学校22人(小学N=10,中学N=12),覆盖多教龄段与多学科。参与教师均承担教研课题且持续开展基于课堂数据的循证研究。

研究分三个阶段:1)系统培训阶段,研究者介绍系统设计理念及核心功能模块;2)模拟操作阶段,参与教师在模拟数据环境中与系统交互(系统界面见图3);3)焦点小组访谈阶段,研究者采用半结构化访谈收集教师反馈并利用主题分析法编码与归类数据。

数据显示,教师普遍认可系统的实践价值。首先,相较于传统智能课堂分析报告的静态呈现模式,

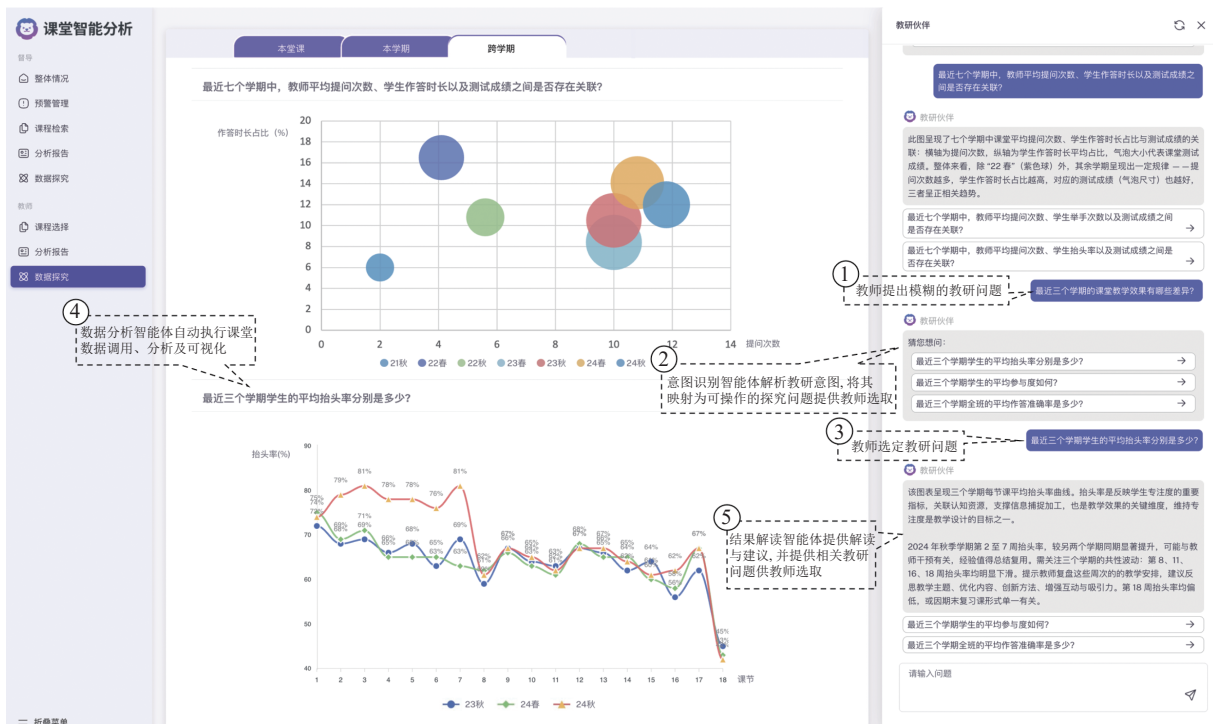


图3 交互式智能课堂数据探究系统界面

基于分析结果的个性化探究功能受到教师的高度评价。如教师所言“横看成岭侧成峰”,长周期的数据能够揭示教师的课堂形态,帮助教师评估干预策略的效果。这一反馈在数字化转型示范校尤为突出。教师此前在集体教研时常需人工整合课堂分析报告,横向对比数据(如平行班差异)与纵向追踪(如干预效果验证)。本系统对数据的智能化、自动化挖掘有效解决了手动操作的繁琐与数据偏差等问题。

其次,教师高度评价与智能体的交互体验,认为其扮演了“教研伙伴”角色。它能将教师情景化提问映射为教研语境的结构化问题,辅助解读数据分析结果的教育意义。同时,这种认知脚手架功能缩短了教师与新系统的磨合期。最后,多数教师提到能从分析结果中获得教研思路。这种“数据初探→问题生成”的逆向探究路径符合日常教研的自然演进规律(Wise et al., 2019)。

教师提出系统需改进的地方在于:1)系统基于教师提问自动生成的教研问题与教师预期存在一定偏差,原因是课堂可识别的基础数据维度有限。2)课堂并非孤立的教学单元,而是学习链条的关键环节。教师通过收集学生在预习阶段的行为特征数据,并结合课后作业的质量分析,将其与课堂教学数据整合,可为教研提供更完整、系统化的证据支持。3)系统应强调算法透明的重要性,清晰标注每个观测指标的推导逻辑,以及每个教研问题的数据分析方案,增强分析结果的可信度。4)教师特别关注数据归属权问题,并提出数据归属权和使用权应由教师主导,避免技术异化为监控工具,促进监控文化的形成。数据驱动的核心目标应是赋能教研和教学模式的创新。

五、反思

本研究突破传统静态反馈模式,构建了交互式课堂观察与循证教研系统,通过三大创新机制实现人机协同的知识生产范式跃迁:1)整合“课后反思”“同课异构”和“行动研究”三类典型教研场景的核心数据挖掘任务,并结合检索增强生成技术构建动态知识库,加强数据分析与教研语境的适配;2)利用大语言模型解析教研意图,将教师的非结构

化提问拆解为可操作的教研问题和数据查询指令,推动自然语言交互与数据分析无缝衔接;3)智能体与分析工具协同,自动化执行数据调用、分析及可视化等任务,使教师能专注于高阶教研任务。实证研究表明,系统能有效支持教师将描述性观察转化为对教学规律、行为模式和教学策略的洞察,推动智能课堂观察从“数据看板”到“教研伙伴”的转型,从静态数据反馈到动态知识生产的范式跃迁。本系统处于研发初期,尚未大规模应用,后续将推进真实课堂的数据应用,邀请教师基于教学实践数据进行个性化探究,获取更细致、深入的用户体验反馈,以优化系统交互设计;聚焦系统的应用验证,长周期追踪系统使用对教研行为和教学效果的影响,为构建新一代智能教研生态系统提供理论与实践依据。

[参考文献]

- [1] Agutter, J., Syroid, N., Drews, F., Westenskow, D., Bermudez, J., & Strayer, D. (2001). Graphic data display for cardiovascular system[C]//Proceedings of the IEEE Symposium on Information Visualization: 163-166.
- [2] Ahuja, K., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., & Agarwal, Y. (2019). EduSense: Practical classroom sensing at Scale[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3): 1-26.
- [3] Alfredo, R., Mejia-Domenzain, P., Echeverria, V., Rahayu, D., Zhao, L., Alajlan, H., Swiecki, Z., Käser, T., Gašević, D., & Martinez-Maldonado, R. (2025). TeamTeachingViz: Benefits, Challenges, and Ethical Considerations of Using a Multimodal Analytics Dashboard to Support Team Teaching Reflection[C]//Proceedings of the 15th International Learning Analytics and Knowledge Conference: 58-69.
- [4] Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., D'Mello, S. K., & Arslan Esme, A. (2019). Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms[C]//Proceedings of the 2019 chi conference on human factors in computing systems: 1-12.
- [5] Asyrofi, R., Dewi, M. R., Lutfi, M. I., & Wibowo, P. (2023). Systematic literature review langchain proposed[C]//International Electronics Symposium (IES). IEEE, 2023: 533-537.
- [6] 白倩,李艺(2022).皮亚杰发生建构思想下的学习发生[J].电化教育研究,43(12): 11-17.
- [7] Bai, X., Huang, S., Wei, C., & Wang, R. (2025). Collaboration between intelligent agents and large language models: A novel approach for enhancing code generation capability[J]. Expert Systems with Applications, 269: 126357.
- [8] Bayrak Karsli, M., Cilligol Karabey, S., Kaba, E., Guler, M., Aydemir Arslan, M., & Kursun, E (2024). Research Trends in Teaching Analytics: Bibliometric Mapping and Content Analysis[J]. Technology, Knowledge and Learning: 1-25.
- [9] Card, S. K., Mackinlay, J., & Shneiderman, B. (Eds.) (1999). Readings in information visualization: Using Vision to Think. [M]. Mor-

gan Kaufmann: 16.

[10] Chen, W., Ma, X., Wang, X., & Cohen, W. W. (2022). Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks[J]. arXiv preprint arXiv: 2211.12588.

[11] 陈峥, 禹双文, 卜凡燕(2022). 用 6A 循证教研模式突破教师用证与制证难题[J]. 教师教育论坛, 35(12): 19-22.

[12] Dai, L., Jiang, Y. H., Chen, Y., Guo, Z., Liu, T. Y., & Shao, X. (2024). Agent4EDU: Advancing AI for Education with agentic workflows[C]//Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education: 180-185.

[13] Davies P. (1999). What is evidence - based education?[J]. British Journal of Educational Studies, 47(2): 108-121.

[14] 高德模(2024). 智能精准教研对教师专业能力提升的影响研究[J]. 大学, (S1): 131-133.

[15] Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy[J]. Teachers College Record, 117(4): 1-22.

[16] Hargreaves, D. H.(1997). In defence of research for evidence-based teaching: A rejoinder to Martyn Hammersley[J]. British Educational Research Journal, 23(4): 405-419.

[17] Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Wang, J., & Schmidhuber, J., et al. (2023). MetaGPT: Meta programming for a multi-agent collaborative framework[C]//The Twelfth International Conference on Learning Representations.

[18] Hoyos, A. A. C., & Velasquez, J. D. (2020). Teaching analytics: Current challenges and future development[J]. IEEE Revista Iberoamericana de Tecnologías del Aprendizaje, 15(1): 1-9.

[19] Janbi, N., Katib, I., & Mehmood, R.(2023). Distributed artificial intelligence: Taxonomy, review, framework, and reference architecture[J]. Intelligent Systems with Applications, (18): 200231.

[20] Jensen, E., Dale, M., Donnelly, P. J., Stone, C., Kelly, S., Godley, A., & D' Mello, S.K. (2020). Toward automated feedback on teacher discourse to enhance teacher learning[C]//Proceedings of the 2020 chi conference on human factors in computing systems: 1-13.

[21] 江雪贞(2024). 基于智能研修平台下的小学校本教研路径探索与实践[J]. 福建基础教育研究, (12): 33-37.

[22] Li, J., Li, D., Xiong, C., & Hoi, S. (2022). Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation[C]//International conference on machine learning. PMLR: 12888-12900.

[23] Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 34(1): 50-70.

[24] Li, L., Farias Herrera, L., Liang, L., & Law, N. (2022). An outcome-oriented pattern-based model to support teaching as a design science[J]. Instructional Science, 50(1): 1-32.

[25] 李玲, 肖君, 顾小清(2023). 智能课堂分析的多维指标体系构建: 面向开放教育 OMO 智慧教学环境的研究[J]. 远程教育杂志, 41(1): 48-59.

[26] Ma, P., Ding, R., Wang, S., Han, S., & Zhang, D. (2023). Insight Pilot: An LLM-Empowered Automated Data Exploration System [C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations: 346-352.

[27] Mandinach E. B., Gummer E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions[J]. Teaching and Teacher Education, 60: 366-376.

[28] McShane, M. (2017). Natural language understanding (NLU,

not NLP) in cognitive systems[J]. AI Magazine, 38(4): 43-56.

[29] Ngoon, T. J., Sushil, S., Stewart, A. E., Lee, U. S., Venkatraman, S., Thawani, N., Mitra, P., Clarke, S., Zimmerman, J., & Ogan, A. (2024). ClassInSight: Designing Conversation Support Tools to Visualize Classroom Discussion for Personalized Teacher Professional Development[C]//Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: 1-15.

[30] North, C. (2006). Toward measuring visualization insight[J]. IEEE Computer Graphics and Applications, 26(3): 6-9.

[31] O' Leary, M.(2020). Classroom observation: A guide to the effective observation of teaching and learning[M]. Routledge: 3-5.

[32] Pérez-Sanagustín, M., Santos, P., Hernández-Leo, D., & Blat, J. (2012). 4SPPIces: A case study of factors in a scripted collaborative-learning blended course across spatial locations[J]. International Journal of Computer-Supported Collaborative Learning, 7: 443-465.

[33] Ramakrishnan, A., Zylich, B., Ottmar, E., LoCasale-Crouch, J., & Whitehill, J. (2021). Toward automated classroom observation: Multimodal machine learning to estimate class positive climate and negative climate[J]. IEEE transactions on affective computing, 2021, 14(1): 664-679.

[34] Saar, M., Prieto, L. P., & Rodríguez Triana, M. J. (2022). Classroom data collection for teachers' data-informed practice[J]. Technology, Pedagogy and Education, 31(1): 123-140.

[35] Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S.(1996). Evidence based medicine: What it is and what it isn' t[J]. British Medical Journal, 312(7023): 71-72.

[36] Waheed, H., Hassan, S. U., Aljohani, N. R., & Wasif, M. (2018). A bibliometric perspective of learning analytics research landscape[J]. Behaviour & Information Technology, 37(10-11): 941-957.

[37] 王一岩, 郑永和(2022). 多模态数据融合: 破解智能教育关键问题的核心驱动力[J]. 现代远程教育研究, 34(2): 93-102.

[38] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q., & Zhou, D.(2022). Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in neural information processing systems, 35: 24824-24837.

[39] 魏宁(2024). 课堂智能分析赋能教研的路径与思考[J]. 中小学信息技术教育, (6): 8-10.

[40] Wise, A. F., & Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making[J]. Journal of Learning Analytics, 6(2): 53-69.

[41] 吴雨宸, 宋萑, 徐兴子(2023). 循证教研: 指向教师实践性知识生成的教研转型[J]. 教师教育研究, 35(1): 13-19.

[42] 杨晓哲(2021). 基于人工智能的课堂分析架构: 一种智能的课堂教学研究[J]. 全球教育展望, 50(12): 55-65.

[43] Yao, S., Yu, D., Zhao, J., Shafraan, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models[J]. Advances in Neural Information Processing Systems, 36: 11809-11822.

[44] Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2008). Understanding and characterizing insights: How do people gain insights using information visualization?[C]//Proceedings of the 2008 Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization: 1-6.

[45] 郁晓华, 彭源, 胡婷玉(2024). 数字化课堂观察与反馈: 现状、评述与实现路径[J]. 开放教育研究, 30(3): 98-108.

[46] 袁永民, 王秀华(2011). “同课异构”应遵循的原则[J]. 当代教育科学, (16): 44-45.

- [47] 翟雪松, 季爽, 焦丽珍, 等(2024). 基于多智能体的人机协同课堂分析框架和技术实现 [J]. 开放教育研究, 30(5): 87-100.
解决复杂学习问题实证研究 [J]. 开放教育研究, 30(3): 63-73.
[48] 张治, 程抒一, 王天蓉(2024). 基于计算教育学的人工智能课 (编辑: 赵晓丽)

From Observation to Insight: A Study on the Construction of a Human-Machine Collaborative System for Evidence-Based Classroom Teaching Research Based on Large Language Models

LI Ling, LIU Jing, WANG Yiwei, PENG Yuan, JIANG Yuanhao & ZHOU Aimin

(Shanghai Institute of Artificial Intelligence for Education, East China Normal University, Shanghai 200062, China)

Abstract: Intelligent classroom observation provides real-time, accurate, and rich data for teaching research by automatically identifying multi-dimensional characteristics of classroom teaching. However, it has not yet deeply integrated into practices due to the lack of contextual interpretation of analysis results, rigid feedback forms, and disconnection from the teaching research. This study constructs a human-machine collaborative classroom observation and evidence-based research system relying on large language models and agent technology to empower teachers to carry out personalized data exploration based on teaching contexts. Specifically, on the basis of automatically extracting multi-dimensional classroom characteristics, the system has achieved three major breakthroughs: 1) It integrates core data mining tasks in three typical teaching research scenarios, namely "post-class reflection", "heterogeneous lesson designs with the same topic", and "action research", and constructs a dynamic knowledge base combined with retrieval-augmented generation technology to realize the adaptation of data analysis to teaching research contexts; 2) It uses large language models to parse teaching research intentions, decomposing teachers' unstructured questions into operable teaching research issues and data query instructions, thus achieving seamless connection between natural language interaction and data analysis; 3) It realizes the collaborative and automatic execution of repetitive tasks such as data calling, analysis, and visualization through the collaboration of multi-agent and analysis tools, enabling teachers to focus on high-value teaching research tasks, such as the interpretation of classroom data for decision-making. Preliminary verification shows that the system can parse teachers' context-based questions and guide teachers to further mine and interpret descriptive classroom observation data to explore insights into teaching rules, behavior patterns, and teaching strategies within the framework of the teaching research discourse system. This human-machine collaborative knowledge co-creation mechanism realizes the paradigm shift of classroom observation from static data feedback to dynamic knowledge production, opening up a new path for artificial intelligence-supported evidence-based teaching research practices.

Key words: intelligent classroom observation; evidence-based teaching research; large language model; multi-agent; human-machine collaborative knowledge co-creation