第30屆全球華人計算機教育應用大會
The 30th Global Chinese Conference on Computers in Education

Education in the Age of AI

AI時代的教育

何以可用、何以為界、何以可能
Opportunities, Boundaries, and Capabilities



博士生論壇論文集
Doctoral Student Forum Proceedings

30th GCCCE

30/5/2026 – 3/6/2026

第三十屆全球華人計算機教育應用大會（GCCCE 2026）

博士生論壇論文集

主編	楊宗凱、尚俊杰、莊紹勇、陳高偉、許庭嘉、林日坤、吳砥、尹弘飚、蔡敬新、趙建豐、曾嘉靈、李卓、徐禎、于玻、張一帆
發行人	楊宗凱
出版者	全球華人計算機教育應用學會
地址	320 台灣桃園市中壢區中大路 300 號
電話	+886-3-4227251 #35456
網址	http://www.gcsce.org/
電子郵件	secretary@gcsce.org
出版格式	PDF 檔
出版年月	2026 年 6 月初版
ISBN	ISBN 9786269323814

目錄

从工具到伙伴：以教师为主体的教师表现性评价系统研究.....3
多智能体人机协同如何支持批判性思维？——基于不同协同配置的实证分析.....7
多模态数据驱动的 STEM 项目式学习过程评估系统构建与应用.....11
基于设计思维的智能教学代理对中学生 AI 学习成就和批判性思维的影响.....15
数字化转型下代际教育期望铸就西藏边境青少年幸福的理路.....23
从静态测评到三维表征：大学生批判性思维的进阶测评框架构建及其实践应用.....31
Review of Multi-Agent Systems (MAS) in Education.....39
科学教育情境中好奇心向探索性学习行为转化的神经机制研究.....43
学习科学视角下拔尖创新人才早期 MAPS 素养模型构建——基于深度访谈的实证研究.....47
基于多智能体辩论的多视角反思支架对职前教师教学设计能力的影响研究.....51
人机协同视域下教师跨学科教学设计胜任力提升研究：模型构建与 GAI 平台干预.....55
高等教育环境下生成式人工智能何以影响学习结果？—基于 27 项实验与准实验的元分析..60
教育智能体从模型驱动走向意图驱动的编排范式.....68
适应性支架教学的运作机制：小学数学学习中支持策略与学生学习表现的交互作用.....73

从工具到伙伴：以教师为主体的教师表现性评价系统研究

From Tool to Partner: Research on a Teacher-Centered Human-Computer Collaborative Teacher Performance Assessment System

牛梦晓[1]，刘欢[2]，孙众[*]
[1 2 *]首都师范大学
[*] sunzhong@cnu.edu.cn

【摘要】 为破解传统教师表现性评价高度依赖专家、成本高昂且难以规模化的实践困境，本研究构建了一个以教师为反思主体的人机协同评价系统。系统遵循分析—协商—改进的人机协同范式，通过整合课前教学设计、课中教学实录视频及 AI 课堂分析报告等多源数据，构建结构化教学案例库。研究采用 BGE-M3 模型进行文本向量化处理以支撑案例检索，并训练了专用的教师表现性评价模型。该系统形成了从教师提问到人机协商，最终生成改进策略的闭环，旨在规模化降低评价成本的同时保障教师的主体性，为人工智能支持教师专业成长提供了可操作的实践路径。

【关键词】 教师表现性评价；人机协同；教师专业成长

Abstract: *To address the practical challenges of traditional teacher performance assessment,which heavily relies on experts,incurs high costs,and is difficult to scale,this study constructed a human-computer collaborative assessment system with teachers as the reflective subject. The system follows the analysis–negotiation–improvement paradigm of human-computer collaboration. It integrates multi-source data including pre-class instructional designs,in-class teaching videos,and AI classroom analysis reports to build a structured teaching case library. The study employed the BGE-M3 model for text vectorization to support case retrieval,and trained a dedicated teacher performance assessment model. The system forms a closed-loop empowerment path from teacher questioning to human-computer negotiation, ultimately generating improvement strategies. It aims to reduce assessment costs at scale while safeguarding teacher agency,providing an actionable practical path for AI-supported teacher professional growth.*

Keywords: Teacher Performance Assessment,Human-Computer Collaboration,Teacher Professional Growth

1. 研究背景

近年来，我国基础教育改革持续深化，教师专业成长与评价机制的完善已成为推动教育高质量发展的核心任务。中共中央国务院《关于深化教育教学改革全面提高义务教育质量的意见》提出要构建以过程性、发展性为导向的教师评价体系，促进教师在评价中实现专业成长。尽管政策导向已明确指向过程性、发展性评价，但教师在实际评价过程中仍面临实践困境。传统评价体系主要依赖于专家听评课，因资源局限，无法常态化，难以满足教师评价的实时性需求（Borko H et al.，2008）。教师表现性评价为困境解决提供理论支持，其强调将评价融入日常教学中，可以满足教师对于评价常态化与实时性的需求。同时可以为教师提供系统化的评价支持，促进教师专业成长。然而，教师表现性评价依赖于专家评价，导致成本高、反

项目基金： 本文系国家自然科学基金人机协同的教师表现性评价关键智能技术研究（项目编号：62477033）阶段性研究成果。通讯作者：孙众 sunzhong@cnu.edu.cn

馈慢、难规模化等问题（周文叶，2022），同时教师的主体性在评价过程中难以保证。人工智能技术的发展与应用，为破解这一难题提供思路。通过人工智能赋能课堂行为智能分析、学情数据精准研判、人机多轮对话与意义理解等，为教师专业发展提供了全新路径（王雯等人，2024）。人工智能不再是单纯的工具载体，而是深度融入教学全流程的协同伙伴。因此，本研究设计了人机协同的教师表现性评价系统，助推高素质专业化教师队伍建设迈向高质量新阶段。

2．理论基础

技术应用须回归评价的核心目标，正如教育评价学者斯塔弗尔比姆所言（Daniel L. Stufflebeam et al.，1966）,评价最重要的意图不是为了证明,而是为了改进。唯有教师成为评价主体之一,将作为评估与管理工具的课堂教学智能分析,转变为助力个人与群体教学改进的技术和方法,才能切实改进课堂教学质量，促进教师专业成长。正是基于这一主体性认识，教师表现性评价成为契合上述理念的应然选择。教师表现性评价是一种旨在评价真实教学实践的专业评估，它强调在真实教学情境中捕获教师的专业实践，通过反思性对话充分表达自身观点，促进教师主体性生成。然而，传统教师表现性评价面临三重困境。其一，评价过程高度依赖专家视域，易受主观性干扰且成本高昂，难以实现规模化覆盖；其二，碎片化观察记录难以形成系统性证据链，无法支撑教师持续的自我监控与循证改进；其三，评价反馈往往滞后且缺乏情境化解读，未能将理论框架有效转化为教师可理解、可操作的实践智慧。

随着人工智能技术的飞速发展，技术赋能为破解上述困境提供了新的可能。生成式人工智能可以自动识别教学行为并生成分析报告，结合大语言模型模拟专家提供即时反馈，理论上可大幅降低实施成本、提升反馈时效。孙众等人基于第四代评价理论，提出的以分析-改进-协商为特征的人机协同教师表现性评价范式，突出了教师主体地位，促进教师反思，提升教师教学能力（孙众等人，2025）。如何将 AI 技术优势与教师表现性评价深度融合，进而构建即时、有效、可操作的评价系统，成为实践性探索的迫切需求。为此，本研究旨在构建基于人机协同的教师表现性评价系统，在实现技术赋能的同时确保教师主体地位，进而促进教师专业成长，推动高素质专业化教师队伍建设。

为进一步为技术实现提供理论指导，本文对以下概念做出界定。

（1）教师表现性评价

教师表现性评价在教师评价领域的运用，是一种旨在评价真实教学实践的专业评估，它关注教师在促进学生学习中的实际表现。教师表现性评价有着明确的目标指向性——于教师评价而言即教师真实有效地促进学生学习（周文叶等人，2020）。具体而言，它基于真实课堂教学情境，通过教学计划、教学实施、评价学习三大任务收集教师的实践证据，评估教师促进学生学习的专业实践能力（周文叶等人，2021）。同时，它突出教师的主体地位，鼓励教师高度主动参与。

（2）人机协同

在本研究中，人机协同指教育评价场景下人类教师与人工智能系统构成的协作关系（武松等人，2025），其核心在于构建分析—协商—改进的智能支持体系。旨在通过技术赋能实现 AI 分析结果向个性化、可操作反馈的转化，同时恪守教师专业判断与反思主体性的教育本位。该界定并非简单的技术辅助逻辑，而是强调教师从被动评价对象向基于 AI 证据的实践性反思者的范式转型，这一转型直接回应了第四代教学评价理论所倡导的沟通与协商价值取向（孙众等人，2025）。实现从信息接收者→证据提供者→反思主导者的角色跃迁。

3. 技术实现

本研究旨在训练教师表现性评价模型，设计教师与模型的交互系统，最终构建一个以教师为反思主体、人机协同的教师表现性评价系统。通过本研究开发的人机协同的教师评价系统，帮助教师提升教学能力，促进教师专业成长。系统技术流程遵循多源资源库输入-教师表现性评价模型分析-人机协商、教学改进的闭环，如图 1 所示。

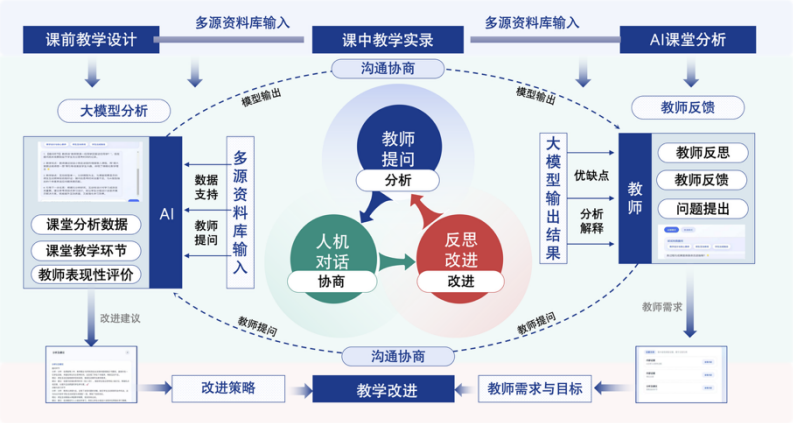


图 1 人机协同的教师评价系统

3.1 构建多源数据支持的教学案例库

系统首先要求教师上传涵盖教学实施前、中、后完整过程的资料，包括课前教学设计、课中教学实录视频及课后的 AI 课堂分析报告。在具体实现中，系统首先将课中教学实录视频转写为带时间戳的师生真实语录，根据时间特点将课堂实录内容与教学评价中的教学活动设计以及 AI 课堂分析报告中的分析数据一一对应。构建包含分析数据与课堂真实流程的映射关系。将其进一步进行文本向量化处理构建资料库，本研究选取三大文本向量化模型进行对比实验，最终选用 BGE-M3 模型作为文本向量化模型。通过实验验证 BGE-M3 模型在测试集上实现 96.7%的召回率，确保模型反馈精准。通过对三大维度映射内容进行向量化处理，形成一个结构化、体系化的课堂教学案例库，为后续的模型处理提供数据支持。

3.2 教师表现性评价模型

该模型的建构以教师表现性评价理论与团队自主研发的《教师表现性评价量表》为训练数据，量表已经过实验验证其有效性。通过采集并结构化处理教师表现性评价理论与相关文献构建专项数据集。并采用微调技术对基座大模型进行训练，训练模型的通用信息处理功能转化为教师表现性评价功能。在应用过程中，教师提出具体问题，模型对接教学案例库，融合教师问题与资源库依据作为模型输入，使用提示词调用模型进行处理。

RAG 层对接系统构建的多源数据支持的教学案例库。主要负责从海量教学资源中精准定位与当前评价情境高度相关的证据片段，为后续生成环节提供可靠的上下文支撑。该层以教师上传的课程资料及权威文献库为检索对象，构建行为数据—教学环节—教师表现性评价的三维语义索引。提示词工程是本系统实现从 AI 课堂分析数据到教师表现性评价转化的关键。通过结构化的提示词设计，将检索增强生成（RAG）所获取的多源证据转化为符合教师表现性评价标准的、可操作的反馈内容。

综合以上模型实现，系统接收教师查询意图后，首先通过 BGE-M3 嵌入模型将查询语义向量化，继而激活混合检索机制。检索结果经重排序算法筛选后，提取 Top-K 相关片段，涵盖：与查询契合的 AI 分析数据段落、对应教学环节的真实课堂实录、教师表现性评价规则及 edTPA 参考案例。

3.3 以教师为主体的人机协商交互系统

人机协商交互系统承接前述教师表现性评价模型，将其从静态分析框架转化为动态专业发展系统。该架构以教师表现性评价结果为协商基准，构建问题提出—证据协商—策略生成的协商路径。

在交互机制层面，本研究构建了以课程实例为单位的情境化对话系统。当教师选定特定课次发起提问时，后端自动聚合该课次的完整教学证据链。通过多模态数据与《教师表现性评价量表》的八条规则及教育理论文献共同构成大语言模型的上下文窗口。同时，系统支持教师对评价结果进行存疑、反驳与证据补充的多轮对话。当教师上传教案修订稿或学生作业样本时，触发二次检索与证据重评，生成协商轨迹的可视化输出。本研究已成功开发并部署了人机协同的教师表现性评价系统，如图 2 所示。



图 2 人机协同的教师表现性评价系统

系统完整实现了分析—协商—改进的机制。通过构建轻量化 Web 交互平台，支持教师用户上传教学设计文档、课堂实录视频及 AI 分析报告等多模态数据。技术验证表明，系统在测试环境中稳定支持并发请求，初步验证了架构设计的可行性与技术路径的有效性，目前该系统开展小规模试点应用。系统实现不仅将理论原理转化为可运行的技术实体，更通过真实教学场景的测试，暴露了当前协商深度不足、改进策略可利用性较差等实践困境。这为后续研究指明了优化方向——需进一步探索人机协同中教师的改进机制，以及基于持续交互数据的模型建议生成路径，从而切实弥合技术支持与教师专业成长之间的转化鸿沟。

参考文献

中共中央，国务院. (2019).关于深化教育教学改革全面提高义务教育质量的意见.中华人民共和国中央人民政府. https://www.gov.cn/zhengce/2019-07/08/content_5407361.htm

王雯和李永智. (2024). 国际生成式人工智能教育应用与省思.开放教育研究，30(3)，37–44.

孙众和韩硕. (2025).从人工智能课堂教学分析到人机协同的教师专业成长.教师教育论坛，38(04)，4–12.

周文叶.(2020).教师评价:评什么和怎么评——访斯坦福大学李·舒尔曼教授.全球教育展望，49(12)，3–12.

周文叶和董泽华. (2021).教师表现性评价系统的研发与实施逻辑：以 edTPA 为例.教育发展研究，41(12)，20–27.

周文叶. (2022).指向立德树人的教师表现性评价.全球教育展望，51(04)，129.

武松、安富海和张仁杰.(2025).功能论视角下人工智能赋能教育的边界与协同机制研究.教育理论与实践，45(34)，25–30.

Borko,H.,Jacobs,J.,Eiteljorg,E.,& Pittman,M.E.(2008).Video as a tool for fostering productive discussions in mathematics professional development. Teaching and Teacher Education,24(2),417–436.

多智能体人机协同如何支持批判性思维？——基于不同协同配置的实证分析

How Do Multi-Agent Human–AI Collaborations Support Critical Thinking? An Empirical Study Across Collaborative Configurations

楚肖燕[1*]，季爽[1*]，朱凌芸[1]，钱春雨[1]，翟雪松[1]
[1]浙江大学教育学院
[*] xiaoyanchu@zju.edu.cn

【摘要】 随着生成式人工智能在教育领域的应用逐渐从工具辅助走向认知协作，单一智能体在促进批判性思维方面仍存在局限。多智能体系统通过多角色分工与反馈，为支持高阶认知提供可能，但不同人机协同配置的影响仍有待检验。本研究以 87 名本科生为研究对象，采用准实验设计考察不同协同配置（单人-多智能体和双人-多智能体协同）对学业表现与批判性思维的影响。结果表明，人机协同模式在知识测验成绩上无显著差异，但双人-多智能体协同的项目产出更优，而单人组在批判性思维意识上表现更佳。滞后序列分析进一步揭示了批判性思维行为的差异特征。

【关键词】 生成式人工智能；多智能体系统；人机协同学习；批判性思维

Abstract: *With the increasing adoption of Generative Artificial Intelligence (GenAI) in education, human–AI collaboration has shifted from tool support toward cognitive partnership. Single-agent interactions may suffer from limited perspectives and echo-chamber effects. Multi-Agent Systems (MAS) offer a potential solution, yet the effects of different human–GenAI collaborative configurations remain underexplored. This study examined how different MAS-supported human–GenAI collaboration configurations influence learning performance and critical thinking. A quasi-experimental study involving 87 undergraduates was conducted. Results showed that collaborative learning with MAS led to higher project performance, whereas individual MAS-supported learning enhanced critical thinking awareness. Lag Sequential Analysis further revealed distinct interaction patterns across configurations.*

Keywords: Generative artificial intelligence, Multi-agent systems, Human–AI collaborative learning, Critical thinking

1. 引言与研究问题

随着生成式人工智能（Generative Artificial Intelligence, GenAI）的快速发展，人机协同学习逐渐从“工具辅助”阶段迈向“认知协作”阶段。尤其是在高阶认知目标导向的学习情境中，GenAI 进一步被视为能够参与分析、评价与反思的认知伙伴。然而，现有研究多聚焦于单智能体支持的人机互动，其在支持批判性思维方面的有效性仍存在争议。多智能体系统（Multi-Agent Systems, MAS）提供了新的路径（Hao et al.，2026）。通过引入多个具有不同角色与功能分工的智能体，MAS 能够模拟多视角对话与分布式认知过程，创造更为复杂的认知冲突与意义协商情境。此外，MAS 的效果受到人机协同配置的影响，即以个体还是协作方式与 MAS 互动，以及同伴与 AI 在学习过程中的角色分工与协作关系（Chu & Li，2025；Tai & Chen，2024）。现有研究虽然开始关注多智能体支持的人机协同学习，但较少比较不同人机协同配置的差异化影响，且对批判性思维的加工过程关注不足。本研究在 MAS 支持下，系统考察不同协同配置(单人-多智能体和双人-多智能体)对学业表现与批判性思维过程的影响：

RQ1：在多智能体支持下，不同人机协同配置对学习者的学习表现有何影响？

RQ2：在多智能体支持下，不同人机协同配置如何影响学习者的批判性思维意识？

RQ3：在不同协同配置下，批判性思维行为序列模式呈现出哪些特征？

2. 研究背景

2.1. 多智能体系统与批判性思维

批判性思维指的是对“该做什么”和“该相信什么”做出合理判断的能力（Ennis，1985），涉及问题识别、信息理解、论证分析、证据评价以及解决方案生成等一系列认知过程（Facione，1990；Murphy，2004）。通过互动，学习者得以表达自身推理、检视潜在假设并不断修正判断，从而支持批判性思维的发展。GenAI 对学习者的分析性与评价性思维过程起到支架作用，但单智能体交互也面临一些挑战，如“回音室效应”，即确认偏差与视角多样性的不足会削弱学习者进行真实认知投入的机会。

为应对上述局限，多智能体系统被认为是一种架构性解决方案。与单一智能体不同，MAS 通过协调多个具有专门功能的智能体，形成一种分布式探究形态（Hao et al.，2026）。通过模拟多视角对话与复杂问题解决情境，MAS 支持的学习环境能够为学习者创造接触竞争性观点、评估替代性主张并论证自身推理的机会。

2.2. 人机协同配置

除 GenAI 本身的能力，人机协同的效果还受到协同配置方式的显著影响，包括个体或协作式与 AI 进行交互（Chu & Li，2025；Tai & Chen，2024）。协同配置规定了学习者、同伴与 AI 的互动组织形式，决定了认知资源分配方式以及信息流动路径。

多智能体系统的引入，进一步拓展并复杂化了人机协同配置。MAS 可以作为共享的认知支架，帮助建立协作团队的整体感并促进对复杂问题的共同理解；此外，也可能作为自主的第三方参与者，通过引入分歧性观点来挑战群体共识，进而激发对既有观点的再评价。目前，系统性比较不同人和多智能体协同配置如何影响学业表现与批判性思维的实证研究仍然稀缺，且对于协作过程中的动态交互机制，亦缺乏深入洞察。

3. 研究方法

3.1. 研究对象与环境

本研究以《现代教育技术》课程作为实验实施情境。该课程具有显著的任务导向与项目驱动特征，为嵌入多智能体支持的人机协同学习活动提供了适切的环境。研究对象为中国一所综合性高校教育学院的本科生，共有 87 名学生参与实验，男性 32 人、女性 55 人。

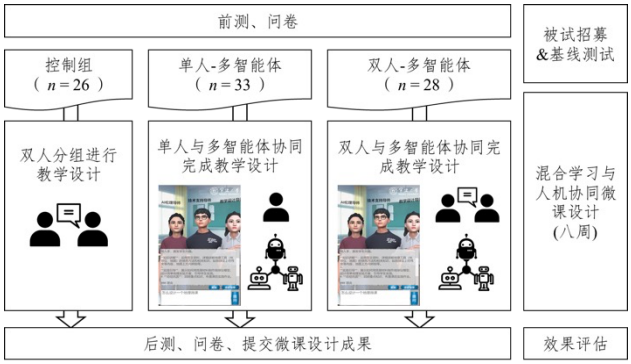
本研究环境基于 Unity3D 构建前端界面，后端集成 CozeAI 多智能体框架与 DeepSeek 模型，包含技术支持导师、教学设计导师与 AI 伦理导师，承担技术操作指导、教学设计咨询与伦理规范引导等功能。系统依托长时记忆管理与知识库集成等功能，通过角色提示词预设智能体的职责边界与语言风格，形成差异化功能分工与情境化认知支持，按需触发分别提供反馈。基于 RAG 机制接入垂直领域知识库，以保障反馈的专业性与语境相关性。三位虚拟智能体均支持语音、表情与动作等多模态交互。

3.2. 研究流程

本研究采用为期 8 周的准实验设计（如图 1）。实验在微课设计课程中实施，学习者需要与同伴或多智能体协同完成一个微课视频的制作。设置三种学习条件：控制组以两人一组完成任务，不使用 AI 支持；单人-多智能体组由学生个体与多智能体系统协；双人-多智能体组则由两名学生在线下协作，共同与多智能体系统交互完成设计活动。三种条件均由同一位教师授课，课程进度、学习目标、教学内容保持一致。干预前后分别实施知识测验与批判性思维意识问卷，结束时提交最终微课设计作品。

图 1

研究实验流程



3.3. 测量工具与数据分析

学业表现由知识测验与期末微课设计项目两部分构成。知识测验包含 10 道选择题，内容涵盖微课设计关键原理与基础软件操作，选自标准化题库。微课设计项目采用包含教学目标（10 分）、教学设计（30 分）、技术实现（20 分）、内容从呈现（20 分）与创意设计（20 分）五个维度的量规进行评价，两位专家独立评分，Cohen's kappa 值为 0.806，一致性良好。

批判性思维意识测量改编自 Hwang 等（2018）的五点李克特量表，Cronbach's α 系数在前后测中分别为 0.774 和 0.924，表明信度良好。

为分析学生在与多智能体系统交互过程中的批判性思维加工模式，基于 Murphy（2004）的框架对对话数据进行编码（见表 1）。两名受训专家独立编码并达到良好一致性（0.848）。

表 1

批判性思维编码框架

阶段	描述	编码
识别	识别或发现已存在的问题、困境或情境	R
理解	探索相关证据、知识、研究、信息与观点	U
分析	澄清问题，组织已知信息，识别未知信息，并将问题分解为基本组成部分	A
评价	对信息、知识或观点进行批判性判断	E
创造	生成新的知识、观点或策略，并加以实施或应用	C

由于数据未完全满足正态分布，使用 Kruskal-Wallis 检验比较学习表现与批判性思维意识方面的差异，采用 Dunn 事后检验并进行 Bonferroni 校正。三组在前测知识成绩和批判性思维意识无显著差异，基线水平相当。采用 GSEQ 软件对编码后的对话进行滞后序列分析。

4. 结果与讨论

在学业表现层面，三组在后测知识测验上未发现显著差异（$p = 0.161$），但项目得分组间差异达到显著。事后检验表明，双人-多智能体协同组显著高于对照组（$p = 0.039$）和单人-多智能体协同组（$p = 0.044$）。这一结果表明，不同学习配置显著影响综合性学习成果，而非知识性测验。双人组的显著优势，显示了同伴协作与多智能体支持的结合在复杂任务情境中具有明显的协同效应。

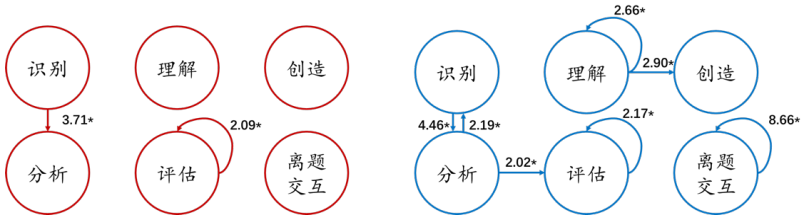
在批判性思维意识层面，三组后测得分存在显著差异（$p = 0.024$）。事后比较结果表明，单人-多智能体协同组的得分显著高于双人-多智能体协同组（$p = 0.019$）。该结果表明，与协作情境相比，单人-多智能体协同学习为学习者提供了更为集中的认知空间，使其需要独立承担问题理解、方案评估与决策判断的责任。

滞后序列分析结果（图 2）显示，在单人组中，显著的行为转移链主要表现为***识别→分析***以及***再评估***，即学习者形成了从问题识别到深入分析的路径，在评估阶段进行检验与反思，反映以个体为中心、强调逻辑推进与评估的批判性思维加工模式。相比之下，双人组的显著行为转移链更加多样化且分散，表现为***识别***与***分析***之间的双向转化、***分析→评估***的递进关系，

以及***理解***、***评估***与***离题交互***的循环共现，并最终从***理解***延伸至***创造***。这表明，协作情境中学习者的认知加工并非线性推进，而是通过反复协商、澄清与整合逐步达成共同理解，并生成新的设计方案。结合自我调节学习及共享调节学习理论，双人组的社会协商与任务导向利于整合产出，但可能减少个体元认知反思深度；而单人直面多视角反馈更易激活思维自我监控，但产出整合效率可能较低。

图 2

滞后序列分析结果



（a）单人-多智能体组行为序列　（b）双人-多智能体组行为序列

5. 研究不足与未来展望

本研究仍存在一定局限。由于样本来源于单一高校课程且采用准实验设计，研究结论的普适性仍需谨慎看待；同时，批判性思维过程分析主要依赖对话数据，尚未整合更多多模态证据。后续研究将从以下几个方面进一步拓展。首先，在研究设计层面引入更具层次性的协同配置与任务复杂度操控，探究不同配置在高、低复杂度任务中的适配性差异。其次，在过程分析层面进一步整合多方法、多模态的数据分析框架，对人机协同路径进行更为精细的刻画。最后，在应用与推广层面可尝试将多智能体人机协同学习模式拓展至不同学科与教育阶段，检验其在不同教学目标与学习群体中的可迁移性。

参考文献

Chu, X., & Li, Y. (2025). Interacting with a humanized dual-LLM agent: An exploratory study of individual and collaborative modes in language proficiency and critical thinking awareness. *Interactive Learning Environments*, 1–21.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership*, *43*(2), 44-48.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction : research findings and recommendations*. P.A. Facione.

Hao, Z., Cao, J., Li, R., Yu, J., Liu, Z., & Zhang, Y. (2026). Mapping student-AI interaction dynamics in multi-agent learning environments: Supporting personalized learning and reducing performance gaps. *Computers & Education*, *241*, 105472.

Hwang, G.-J., Lai, C.-L., Liang, J.-C., Chu, H.-C., & Tsai, C.-C. (2018). A long-term experiment to investigate the relationships between high school students' perceptions of mobile learning and peer interaction and higher-order thinking tendencies. *Educational Technology Research and Development*, *66*(1), 75–93.

Murphy, E. (2004). An instrument to support thinking critically about critical thinking in online asynchronous discussions. *Australasian Journal of Educational Technology*, *20*(3).

Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, *220*, 105112.

多模态数据驱动的 STEM 项目式学习过程评估系统构建与应用

Construction and Application of a Multimodal Data-Driven STEM Project-Based Learning Process Assessment System

谢颖[1]，乔爱玲[1*]
[1] 首都师范大学教育学院
* qiaoal@126.com

【摘要】 针对 STEM-PBL 评价实践困境，本研究计划开发并验证一套学习过程评估系统。采用基于设计的研究范式，沿“理论构建、原型开发、实证检验”路径展开。首先，构建涵盖认知、行为、社会与元认知维度的评估指标框架。继而开发集成多模态数据采集、融合分析与可视化反馈的原型系统，运用认知网络分析等方法，实现对学习协作与设计过程的自动化解析。最后通过准实验与案例研究，在真实课堂中检验系统有效性，并探究其对学习投入、协作质量与元认知的影响机制。本研究计划为 STEM-PBL 过程性评价与反馈提供整合性方案，也为技术赋能的动态评估提供实践参照。

【关键词】 STEM-PBL；学习评估；多模态数据；学习分析

Abstract: *This study plans to address practical challenges in STEM-PBL assessment by developing and validating a learning process assessment system through design-based research. Following a "theoretical construction—prototype development—empirical testing" sequence, it first establishes a multi-dimensional assessment framework. around cognitive, behavioral, social, and metacognitive dimensions. A prototype system is then built to integrate multimodal data collection, fusion analysis, and visual feedback, applying methods like epistemic network analysis to automatically parse collaborative and design processes. Finally, classroom quasi-experiments and case studies evaluate the system's effectiveness in enhancing learning engagement, collaboration, and metacognition.. This research plan to offer an integrated solution for formative assessment and feedback in STEM-PBL, along with actionable insights for technology-supported dynamic evaluation.*

Keywords: STEM-PBL, Learning assessment, Multimodal Data, Learning analytics

1. 研究目的

针对 STEM 项目式学习（STEM-PBL）中长期存在的评价理论与实践脱节问题，即传统评价方式难以有效衡量复杂问题解决过程中的高阶思维、协作与实践能力（Gao et al., 2020），本研究计划构建、实施并验证一个内嵌于 STEM-PBL 学习过程的多模态动态评估系统。在理论层面，整合学习科学、表现性评价与多模态学习分析的相关成果，构建能够刻画认知、社交、行为与元认知等多维能力发展的过程评估模型，以弥合宏观教育目标与微观学习证据之间的鸿沟。在技术与方法层面，研发多模态数据融合分析引擎与可视化反馈系统，通过对非结构化过程数据的自动化解析，将“协作质量”等抽象概念转化为可量化的计算指标。在实践层面，于真实课堂环境中检验该系统的教学有效性及用户体验，探究其如何通过实时、可视化的过程性反馈影响师生的元认知与教学决策，最终形成可直接支撑 STEM-PBL 教学的评估工具与设计原则。

2. 研究背景

作为开展科学、技术、工程、数学多学科于一体的跨学科综合教育，STEM 教育倡导通过基于科学和数学的概念和程序解决问题，结合工程策略和应用适当技术，对于发展学生的高阶思维能力、人际交往能力以及创造能力等核心能力具有重要作用（Sánchez Carracedo et al., 2018）。其中 PBL 与 STEM 教育的结合（STEM-PBL）可以提供“情境化的真实经验，使学生能够构建有意义的、强大的科学、技术、工程和数学概念”（Capraro & Slough, 2013），成为全球备受欢迎的教学模式。然而，其评价体系的发展滞后于 STEM 教学实践的创新，陷入三重困境。

其一，评价内容与学习目标错位。标准化测试或作品评分难以评估项目进程中动态发展的核心能力（Gao et al., 2020），如迭代设计中的决策推理、跨学科知识整合及从失败中学习的学习韧性。其二，评价方法在信度、效度与可操作性上存在局限。教师观察或学生自评主观性强且易产生回顾性偏误（Hsiao et al., 2022）；而学习管理系统日志分析多聚焦于点击流、学习时长等表层行为（Xia & Qi, 2024），难以触及认知与社会性互动等深层内涵。其三，技术赋能评价的潜力未被充分挖掘。当前技术应用多为零散的、事后分析式的工具（Karakaya & Yılmaz, 2022），缺乏对多源异构数据同步采集、融合分析与即时反馈的完整架构。

基于上述分析，本研究以多模态学习分析为核心理论视角与方法论框架。学习本质上是具身化与情境化的，其过程证据广泛蕴含于语言、行为、生理等多模态数据流中。通过系统性采集并计算融合这些数据，能够形成对复杂学习过程更为全面、客观且精细的动态表征。因此，本研究致力于解决的核心问题是**如何设计并实现一个基于多模态学习分析的技术系统，以自动化、可视化的方式对学生的 STEM-PBL 能力表现进行过程评估，从而有效提升教学与学习的质量与效率？**本研究拟定提出一种创新的评价方案，为技术赋能的动态学习评估提供实践参照。

3. 研究方法

本研究拟采取“理论构建-原型开发-实证检验”的路径。第一阶段为理论框架构建（图 1），通过系统的文献综述确立理论基础，随后对一线 STEM 教师及学科专家进行深度访谈与焦点小组讨论，获取实践需求。在此基础上，采用德尔菲法组织专家咨询，对初步构建的评估维度与指标进行评议，确立具有共识的多模态 STEM-PBL 过程性评估框架。

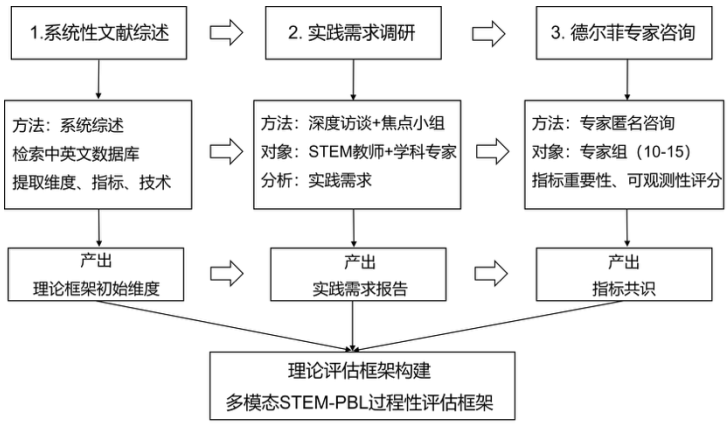


图 1 理论框架构建流程图

第二阶段为系统原型开发（图 2），核心任务是将理论框架转化为可运行的技术原型。为此，本研究设计了包含“数据采集、处理分析、可视化反馈”三层结构的系统架构。在数据采集层，系统整合定向麦克风、全景摄像机与学习管理平台的 API 接口，在 PBL 任务执行过程

中同步捕获音频、视频及数字制品日志，形成多模态数据流。处理分析层则对上述数据进行综合处理。运用自动语音识别与自然语言处理技术，对小组对话进行转写与语义编码，再通过认知网络分析刻画协作对话中的概念关系结构，从而量化评估 STEM-PBL 学习过程的关键维度；其次，为捕捉协作中的非言语互动，系统引入基于 MediaPipe 库的计算机视觉方法，通过实时姿态估计提取成员的身体朝向、手势与间距等特征，进而计算“小组空间凝聚度”与“任务协同焦点时间比”，以客观反映协作参与模式与注意力分布。此外，采用滞后序列分析识别设计迭代中的时序依赖模式，对应元认知调控行为。三类数据经毫秒级时间对齐后融合为统一事件日志，通过跨模态事件关联实现因果推断。数据处理采用近实时批处理，教师仪表盘每 10 分钟刷新一次，以平衡反馈及时性与课堂流畅性。在可视化反馈层则基于 Web 技术开发交互式教师仪表盘，将上述分析结果以时间线、网络图、热力图等形式动态呈现，为教学干预提供直观依据。

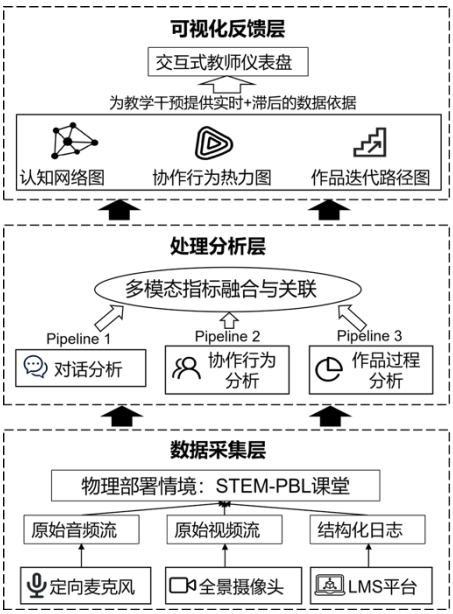


图 2 系统原型开发流程图

第三阶段为实证检验，在中学真实情境中开展两轮迭代研究。参考基于设计的研究范式，首轮迭代旨在对系统原型进行形成性评估与优化，核心是确保工具的可用性与可靠性。首轮迭代在一个初中 STEM 项目班中展开，采用单组案例研究设计，全面部署并测试系统原型（V1.0）。数据收集聚焦于技术可行性与初始用户体验，基于初步测试结果，重点识别出原型系统在算法稳定性、用户交互逻辑及教学场景适配性等方面存在的主要不足，改进并形成 V2.0 系统版本，为后续开展严格的效能实证研究奠定了可靠基础。次轮迭代旨在对优化后的系统进行效果验证与影响机制探索，核心是检验其教育有效性并深化理论理解。采用准实验设计，同步采集多模态过程数据，两组教师保持相同的干预频率与时长，唯一差异在于实验组教师基于仪表盘数据进行诊断，对照组教师基于常规观察进行诊断。收集前后测量表、过程性多模态数据、作品成绩及访谈资料。基于此轮分析，研究最终形成完善的评价模型、定型的工具系统，并总结出可迁移至同类情境的 STEM 学习过程评价设计原则与实践指南。

4. 研究阶段与初步成果

目前，本研究处于正式实施的启动与奠基阶段。前期构想与规划已完成，并已转化为一个具体、可操作的研究方案，为核心工作的开展做好了全面准备。在研究整体规划与设计方面，已经完成了核心研究问题的凝练并确立了以 “基于设计的研究”为核心，融合混合方法的总体研究范式。研究的技术路线已经明确，即通过构建多模态数据采集与分析系统，实现对

STEM-PBL 学习过程的动态评估。包含“理论构建、原型开发、实证检验”的研究路线图已绘制完成，并完成了对初步合作研究场域的联络与沟通。

其次，在理论基础与技术准备方面，已经完成了初步的探索性工作。研究已围绕“多模态学习分析““STEM 过程性评价“等主题进行了广泛的文献调研，确定了以学习科学、学习过程评价和社会文化理论作为本研究的核心理论视角。同时，已对关键技术的可行性进行了初步探索与测试，例如利用现有开源工具（如 MediaPipe）尝试了课堂视频中人体姿态的初步提取，并使用在线语音识别服务测试了对话转写的效果等。

5. 未来研究工作

接下来将有序推进第一阶段至第二阶段的工作，具体计划如下。第一步，完成理论评估框架的构建与验证。通过系统文献综述及对 15-20 名 STEM 教师与学科专家的访谈调查，明确评价需求。整合文献与实证数据，形成多模态评估指标框架。随后邀请教育学、学习分析、STEM 教学等领域专家（约 10-15 人）开展两至三轮德尔菲咨询，通过统计分析各指标的重要性与可观测性评分，确立具有良好内容效度的正式评估框架。预计此步骤持续 3-4 个月。第二步，启动多模态分析系统的原型设计与开发。在框架基本确定后，进行系统架构设计，明确数据流与技术选型，并依据指标进行算法模型选择与适配。同步开展前端仪表盘原型设计，与教师进行可用性测试迭代，目标是完成核心功能可演示的集成原型系统 V1.0，为第一轮实证循环做好准备。第三步，开展第一轮实证循环。在合作学校选取一典型 STEM 项目，在班级中部署系统 V1.0。本轮关注形成性评估，收集系统技术性能数据、师生用户体验反馈及多模态数据初步分析结果，检验指标计算的有效性。通过分析识别技术瓶颈、交互设计问题及教学融合障碍，为开发更稳定易用的 V2.0 系统提供改进依据，并同步完善下一轮效果验证的研究设计。

参考文献

Capraro, R. M., & Slough, S. W. (2013). Why PBL? Why STEM? Why now? An introduction to project-based learning: An integrated science, technology, engineering, and mathematics (STEM) approach. In R. M. Capraro & S. W. Slough (Eds.), *STEM Project-based learning: An integrated science technology engineering and mathematics (STEM) approach* (pp. 1–6). Rotterdam, The Netherlands: Sense.

Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *International Journal of STEM Education*, *7*(1), 24.

Hsiao, J. C., Chen, S. K., Chen, W., & Lin, S. S. (2022). Developing a plugged-in class observation protocol in high-school blended STEM classes: Student engagement, teacher behaviors and student-teacher interaction patterns. *Computers & Education*, *178*, 104403.

Karakaya, F., & Yılmaz, M. (2022). Teachers' views on assessment and evaluation methods in STEM education: A science course example. *Journal of Pedagogical Research*, *6*(2), 61-71.

Sánchez Carracedo, F., Soler, A., Martín, C., López, D., Ageno, A., Cabré, J., ... & Gibert, K. (2018). Competency maps: an effective model to integrate professional competencies across a STEM curriculum. *Journal of Science Education and Technology*, *27*, 448-468.

Xia, X., & Qi, W. (2024). Driving STEM learning effectiveness: dropout prediction and intervention in MOOCs based on one novel behavioral data analysis approach. *Humanities and Social Sciences Communications*, *11*(1), 1-19.

基于设计思维的智能教学代理对中学生 AI 学习成就和批判性思维的影响

The Impact of Design Thinking -Based Intelligent Pedagogical Agents on Middle School Students' AI Learning Achievement and Critical Thinking

郭晓戈[1]，张屹[1]，涂芸芳[2*]，陈邓康[1]，卢逸杰[1]
[1]华中师范大学人工智能与教育新形态实验室
华中师范大学人工智能教育学部
[2]東吳大學 資料科學系
[*]sandy0692@gmail.com

【摘要】 AI 项目学习活动具有跨学科整合的特点，已成为学习人工智能的有效途径。然而，学生在 AI 项目学习活动中往往关注作品制作，限制了批判性思维的发展。基于此，本研究开发了具有学习指导与反思鼓励功能的智能教学代理(IPA)，探讨其作为学习辅助工具对学生 AI 学习成效和批判性思维的影响。参与者为 45 名高一年级学生。实验组使用基于设计思维的智能教学代理 (IPA-DT)，控制组采用传统的设计思维 (C-DT) 方法。结果显示，与 C-DT 相比，IPA-DT 能有效促进学生的批判性思维，但 AI 学习成效没有提升。此外，实验组活动后报告了对 IPA-DT 的积极看法。该研究为教学设计者提供了参考。
【关键词】 智能教学代理；设计思维；批判性思维；AI 教育

Abstract: *AI project-based learning activities feature interdisciplinary integration. They have become an effective means of learning about artificial intelligence. However, students often focus on completing work, limiting their critical thinking development. Based on this, this study developed an Intelligent Pedagogical Agent (IPA) with guidance and reflection functions. It explores the IPA's impact as a support tool on students' AI learning achievement and critical thinking. Participants included 45 first-year high school students. The experimental group used the IPA-based design thinking (IPA-DT) approach. The control group used the traditional design thinking (C-DT) approach. Results show the IPA-DT approach promotes students' critical thinking more than the C-DT approach. However, it does not significantly enhance AI learning achievement. Additionally, the experimental group reported positive perceptions of the IPA-DT approach following the activities. The findings provide valuable insights for instructional designers.*
Keywords: Intelligent pedagogical agent, design thinking, critical thinking, AI education

1. 研究目的

本研究旨在探讨智能教学代理 (Intelligent Pedagogical Agent, IPA)如何优化基于设计思维的 AI 项目学习，以促进学生批判性思维的发展。 当前，学生在 AI 项目活动中常过度关注技术工具与作品产出，而缺乏对问题本质、解决方案及学习过程的深度反思，限制了高阶思维能力的培养。同时，传统课堂中教师难以提供持续、个性化的指导以支持这一深度过程。为应对这些挑战，本研究开发基于计思维的智能教学代理，旨在通过 IPA 提供实时、支架化的引导与反思支持。通过一项准实验研究，本研究具体比较了 “IPA-DT 方法” 与 “C-DT 方法” 对学生 AI 学习成效与批判性思维的影响，并调查了学生对 IPA 干预的感知。预期成果将为理论上阐释 IPA 在促进元认知与批判性思维中的作用机制提供实证依据，并为实践上设计有效的、支持思维发展的 AI 项目学习环境提供具体参考。

2.研究背景

人工智能 (AI) 已融入人们的生活，在教育、医疗和金融等领域发挥关键作用(Kalmatov et al., 2025)。因此，培养适应智能时代的人才已成为教育领域的重要议题 (Fang et al., 2025)。一些研究者指出，融入设计思维的 AI 学习，引导学生参与原型开发、测试、解决方案优化和作品实现的系统过程，帮助他们有效地学习和应用 AI 知识和技能 (Aranda et al., 2020; Azizan & Abu Shamsi, 2022; Wu et al., 2022)。这种方法强调持续反思和改进，提升了学生的学习表现和批判性思维 (Tang, 2020)。例如，Ng et al. (2024)招募了 35 名中学生制作人工智能驱动的回收箱，研究结果显示该项目对学生的学习动机和学习成效具有积极影响。Kuo et al. (2019) 的研究发现基于设计的项目式学习，不仅深化了学生对知识的理解和应用，还提升了学业成绩和批判性思维。

然而，一些研究表明，在基于设计思维的 AI 项目学习活动中，学生往往更注重应用工具制作作品，缺乏对问题本质的深度探究、对解决方案的批判性评估以及对自身学习过程的反思，导致批判性思维未能得到有效培养(Choi et al., 2026)。此外，在课堂活动中，教师难以实时关注所有学生的进展并提供及时、个性化的指导。部分学生因在关键环节遭遇困难时未能获得有效支持，导致学习兴趣衰减、参与度降低，甚至产生挫败感 (Lesseig et al., 2016)。这种反馈的缺失阻碍了学习的深度和持续性。

随着大语言模型(Large language models, LLMs)的出现，智能教学代理能够为学习者提供信息和指导，在学习者遭遇困难时提供及时的个性化帮助。它不仅帮助学习者克服学习障碍，还支持学习者进行深入反思，从而促进批判性思维的发展(Baha et al., 2023; Daradoumis & Arguedas, 2020; Kim & Baylor, 2016)。例如，Alharbi (2017) 探索使用会话代理 Botocrates 让用户参与反思和论证，通过结构化对话增强了他们的批判性思维能力。Chin et al. (2014) 指出智能教学代理能作为虚拟同伴或虚拟导师，能提供个性化即时的反馈和支持学生的多元观点与认知发展，以提升学习兴趣和发展高阶思维。Fang 等人（2025）发现，在 STEAM 课程中使用人工智能教学代理的学生，其学习行为序列表现出更频繁的寻求协助—自主探索—问题解决模式，而非仅仅聚焦于任务监控与成果评估。这表明 IPA 能有效将学生的注意力从表层执行引导至深度探究与反思，从而在行为层面上为批判性思维的发展创造有利条件。

智能教学代理因其智能和具身的特点，在促进学生学习成效和批判性思维方面显现出巨大的潜力(Leavy et al., 2023; Hostetter et al., 2023)。因此，本研究以人机交互(Human Computer Interaction) 理论为依据，开发了一种能够进行学习指导和鼓励反思的智能教学代理。这个智能教学代理旨在支持设计思维核心环节的实施，具体包括需求定义（define）、创意发想（ideate）、原型制作（prototype）和实际测试（test）。本研究将该智能教学代理(IPA)应用于高中人工智能课程，探讨基于设计思维的智能教学代理 (IPA-DT) 方法和传统的设计思维 (C-DT) 方法对高中生的 AI 学习成效和批判性思维的影响。以下是本研究的研究问题：

（1）相较于 C-DT 方法，IPA-DT 方法能提高高中生的 AI 学习成效吗?

（2）相较于 C-DT 方法，IPA-DT 方法能增强高中生的批判性思维吗?

（3）IPA-DT 组和 C-DT 组学生的学习经验和感受是什么？

3.基于智能教学代理的设计思维方法

本研究借鉴了 Li and Zhan (2022) 的建议，提出了一个基于设计思维的学习框架，并根据具体教学内容进行了适当调整与优化。每个设计项目都遵循需求定义(define)、创意动脑(ideate)、制作原型(prototype)以及实际测试(test)四个关键步骤，本研究据此提出了 IPA-DT 方法(见图 1)，旨在系统性地引导学生学习和应用人工智能技术解决现实中的问题。老师向与 IPA

合作的实验组学生介绍了它的用法。这包括如何使用 IPA，该工具的局限性，以及它在学习中的应用。为了确保实验的完整性，同一位老师会指导控制组的学生，讨论有效协作的策略、资料查询的方法。

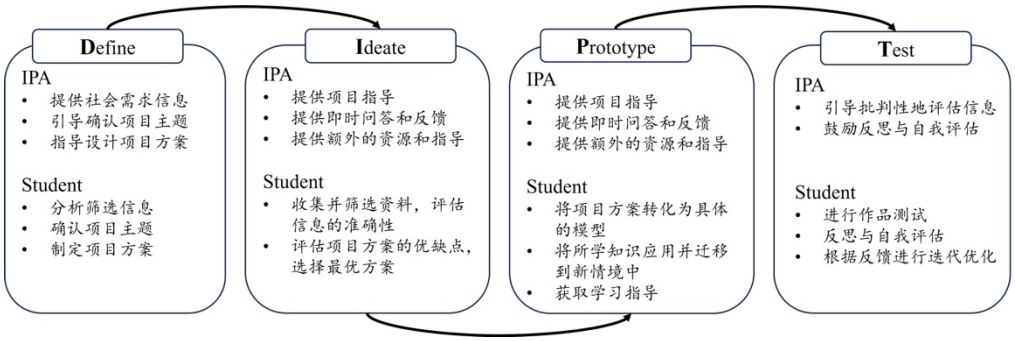


图 1 IPA-DT 方法

在需求定义 (Define)阶段，活动重点在于激发学生的设计兴趣，帮助他们理解用户需求并明确设计主题。教师展示故事、视频和图片引导学生选择项目主题，并确保设计聚焦于实际问题。实验组的学生在 IPA 的辅助下，探索并分析社会议题的核心内容，并批判性地评估议题的重要性(如图 2)。

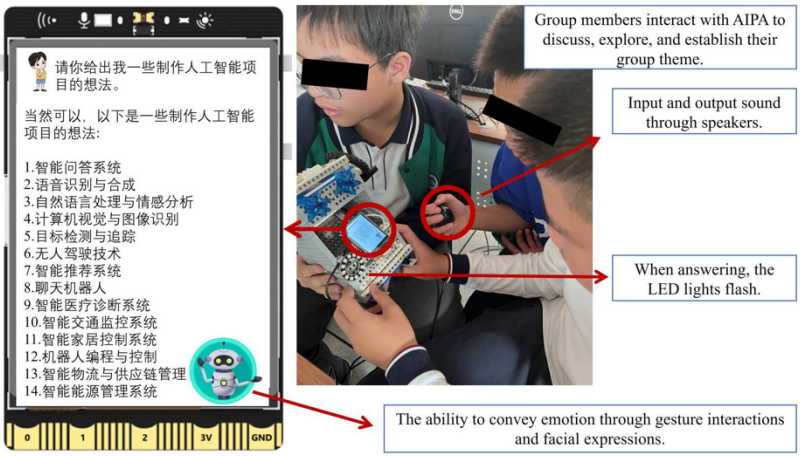


图 2 IPA 交互界面

在创意动脑 (Ideate)阶段，小组进行头脑风暴，制定 AI 舌诊器项目方案。实验组学生在此阶段可利用 IPA 提供的项目方案参考，IPA 不仅帮助学生反思制作项目方案中遇到的问题，还鼓励他们思考解决策略。

在制作原型 (Prototype)阶段，学生将设计方案具体化为 AI 舌诊器作品，并根据设计要求实现各项功能。学生收集相关素材构建数据集，并在 IPA 的指导下完成模型的训练、推理、转换与部署。在此过程中，IPA 不仅向学生提供知识解释，还通过任务引导帮助他们完成项目的各个步骤。

在实际测试 (Test)阶段，学生测试 AI 舌诊器的性能和用户体验。根据测试结果，学生识别出改进空间（如功能优化、界面改进等），并基于反馈进行迭代优化，IPAs 根据存储在记忆库中学生互动记录对作品进行评价与反馈。

4. 研究阶段

如图 3 所示为本研究的具体流程。干预开始前，两组学生均接受前测，以评估其 AI 学习成效与批判性思维水平。干预阶段共持续 8 周，每周进行 2 次、每次 80 分钟的教学。在前 4

周，实验组与对照组学生在同一位教师的指导下，共同参与“情绪识别助手”AI 项目活动，两组的学习内容与活动安排完全一致。期间，我们为学生提供了与智能对话代理（IPA/大语言模型）进行有效沟通的询问技巧与示范，以降低认知负担并提供初步支持。在掌握基本方法后，实验组采用 IPA-DT 教学法，对照组则采用 C-DT 方法，两组学生分别开展为期 4 周的自主 AI 项目设计与制作。干预结束后，两组学生均完成 AI 学习成效与批判性思维的后测。为深入了解学生对两种教学法的看法，本研究随后从实验组与对照组中各随机抽取 9 名学生进行访谈，每次访谈约为 20 分钟。

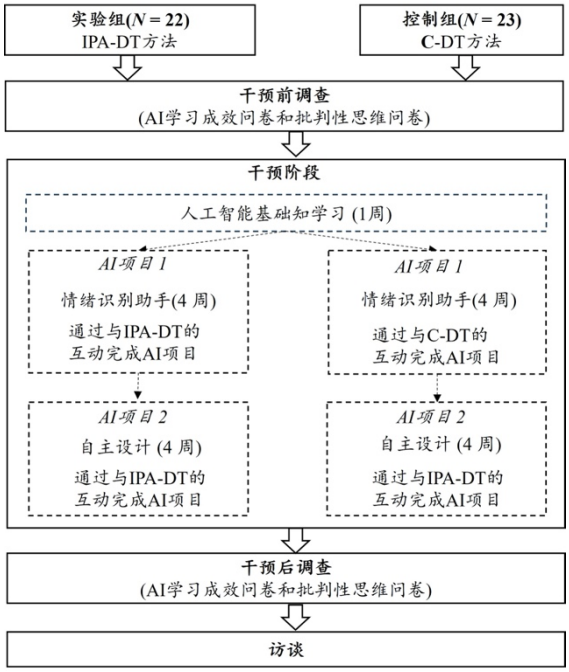


图 3 实验流程

本研究采用的测量工具包括人工智能作品评价量规，批判性思维问卷和半结构化访谈大纲。

人工智能作品评估量规是根据 Fang et al. (2025)的评估量规修改而来的，如表 1 所示。该评估量规作为评价学生人工智能作品成绩的评分依据，共五个评价维度(功能、外观、品质、创新、元素)，各维度打分范围为 0-20 分。两位教学经验丰富的专业教师进行了评分，两位评分者之间的一致程度非常高，达到 0.89 (Cohen, 1960)。

表 1. 人工智能作品评价量规

Item	4	3	2	1
功能	连续识别 5 次，全部正确识别	连续识别 5 次，正确识别 4 次	连续识别 5 次，正确识别 3 次	连续识别 5 次，仅正确识别 1-2 次
外观	在 3 个或以上的位置有图案或装饰	有 2 个图案或装饰	有 1 个装饰或图案	没有装饰
品质	没有结构不牢固的地方	有 1 处结构不牢固	有 2 处结构不牢固	有 3 处或更多地方结构不牢固

Item	4	3	2	1
创新	具备 3 处或以上的创新 (项目概念新颖，与老师教授的项目不同，且能解决实际问题)	有 2 处创新 (项目概念新颖，与老师教授的项目不同)	有 1 处创新 (例如，与老师教授的项目相似，但数据集不同等)	无创新 (类似于老师教授的项目)
元素	添加了 2 个或更多新元素 (如摄像机、传感器等)	添加了 1 个新元素 (如摄像机、传感器等)	基本元素完成 (如仅完成基础框架)	几乎没有完成的元素

批判性思维量表基于 Chai et al. (2015)开发的批判性思维问卷修改而来。它由 6 道问题组成，采用李克特 5 点量表(1 代表非常不同意，5 代表非常同意)。其 Cronbach’ s alpha 值是 0.90。

访谈大纲是根据 Hwang et al. (2009)的访谈提纲改编而来的。根据研究的具体情况，这些问题略有改动。

5. 研究结果

5.1.AI 学习成效

本研究验证了 AI 学习成效的正态性，前测数据(Shapiro-Wilk = 0.79, $p < .001$)和后测数据(Shapiro-Wilk = 0.82, $p < .001$)的数据不具正态性。因此，本研究采用了 Mann-Whitney U Test 方法。如表 2 所示，实验组的 Mean = 17.64 , SD = 2.36，控制组的 Mean = 17.22, SD = 2.24。Pairwise Mann-Whitney U Test (U = 227.5 , Z = -0.66 , $p = 0.51 > 0.05$)的结果显示，两组学生 AI 学习成效前测不存在显着差异。

表 2. AI 学习成效前测之 Mann-Whitney U 检定结果

变量	组别	N	均值	标准差	中位数	U	*p*	r
AI 学习成效前测	实验组	22	17.64	2.36	16	227.5	0.51	0.10
	控制组	23	17.22	2.24	16			

如表 3 所示，实验组的 Mean = 18 , SD = 2.39，控制组的 Mean = 17.39, SD = 2.86。Pairwise Mann-Whitney U Tests (U = 227 , Z = -0.66 , $p = 0.51 > 0.05$)的结果显示，实验组和控制组在 AI 学习成效后测不存在显着差异。

表 3. AI 学习成效后测之 Mann-Whitney U 检定结果

	组别	N	均值	标准差	中位数	U	*p*	r
AI 学习成效后测	实验组	22	18	2.39	20	227	0.51	0.10
	控制组	23	17.39	2.86	16			

5.2.批判性思维

本研究验证了批判性思维的 ANCOVA 分析的前提条件。结果显示，实验组数据 (Shapiro-Wilk = 0.961, $p = 0.134 > 0.05$) 和控制组数据 (Shapiro-Wilk = 0.950, $p = 0.051 > 0.05$) 满足正态和独立性；Leven’s test 的结果($F(1,43) = 0.388$, $p = 0.537 > 0.05$) 显示实验组和控制组

符合方差同质性假设；线性结果(F(1,41)=0.27，p = 0.606> 0.05)显示实验组和控制组符合组内回归系数同质性假设。如表 4 所示。实验组的批判性思维平均分数为 4.29，调整后平均为 4.29，控制组的平均分数为 3.85，调整后平均为 3.84。排除前测的影响之后，实验组显着优于控制组 (F(1,42) = 9.184，p = 0.004 < 0.01)。

表 4. 批判性思维之 ANCOVA 分析结果

变量	组别	N	均值	SD	调整后均值	SE	F	η2
批判性思维	实验组	22	4.29	0.54	4.29	0.11	9.184**	0.179
	控制组	23	3.85	0.43	3.84	0.10		

**p< .01

5.3. 访谈结果

本研究参考 Chen and Lee（2021）的框架对访谈内容进行编码，共获得 112 条有效回复（实验组 71 条，控制组 41 条）。表 5 展示了编码主题及频次。

在实践与反思方面，学生普遍展现出质疑、验证与改进的反思意识。实验组学生更倾向于主动判断 IPA 提供的信息是否合理，并交叉比对多方资源，例如 E01 提到：“我会先判断是否真的符合我的需要。”控制组则更多通过小组讨论和事后复盘进行分析与优化。关于技能提升的原因，实验组学生普遍认为 IPA 在即时答疑、原理讲解、资料提供与思路拓展方面给予重要支持。例如 E05 表示，IPA 帮助其“理解知识点”，E06 认为它“拓宽了我的思路”。相比之下，仅使用大语言模型的控制组学生虽能获取丰富信息，但也面临信息筛选与可信度判断的负担。在使用体验上，实验组学生肯定 IPA 操作简便、反馈及时，且具有一定的情感理解力（E08：“它似乎懂我的感受”），能有效减轻认知负荷。但部分学生指出其回答有时不够精准或深入。控制组学生则反映自主搜索信息效率较低，容易陷入信息过载。

表 5. 访谈结果的主题、编码项目及提及次数

主题	编码项目	次数	
		实验组	控制组
实践与反思	独立思考	8	5
	多重验证	7	6
	选择性接受	8	8
	批判性分析	8	5
	持续学习与改进	6	6
技能提升的原因	提供及时反馈	7	0
	提供个性化指导	5	0
	提供学习材料	4	0
	拓展思维	4	2
使用 IPA 工具的经验	易用性	7	5
	可用性	4	4
	人性化程度	3	0

6.讨论与结论

针对研究问题 1，IPA-DT 与 C-DT 在 AI 学习成效上无显著差异，与 Fang et al. (2025) 一致。实验组更多参与问题分析、反思等高阶认知活动，控制组侧重任务执行，符合 IPA 引导

深度思考的功能。但 IPA 的多轮对话在初期可能增加认知负荷，影响后续效率；同时有用性与易用性问题(Sikström et al., 2024)及智能水平不也可能形成阻碍，访谈中亦有反映。

针对研究问题 2，IPA-DT 较 C-DT 更能强化批判性思维，与既往研究一致(Alharbi, 2017)。IPA 通过建议引导深度思考，避免确认偏差；作为具身伙伴，其提供的解释与元认知线索可促使学习者反思思维过程。依据人机交互理论，教学代理的社会线索能激发深层认知加工(Sikström et al., 2024)。在 AI 项目制作中，IPA 的资源与指导促进了反思与自我评估。

研究问题 3，IPA-DT 较 C-DT 更能促进积极认知反应，与既往研究一致(Hostetter et al., 2023)。实验组在理解、分析、评价维度上显著更优，得益于 IPA 的即时反馈、批判性引导与资源支持，促进了深度加工与反思；控制组在创造维度上略高，可能源于其自主性与对外部资源的依赖。

参考文献

Alharbi, A. H. F. (2017). *Modelling a conversational agent (Botocrates) for promoting critical thinking and argumentation skills* (Doctoral dissertation, University of Leeds).

Aranda, M. L., Lie, R., & Selcen Guzey, S. (2020). Productive thinking in middle school science students' design conversations in a design-based engineering challenge. *International Journal of Technology and Design Education*, *30*(1), 67-81. https://doi.org/10.1007/s10798-019-09498-5

Azizan, S. A., & Abu Shamsi, N. (2022, May). Design-based learning as a pedagogical approach in an online learning environment for science undergraduate students. In *Frontiers in Education* (Vol. 7, p. 860097). Frontiers Media SA. https://doi.org/10.3389/feduc.2022.860097

Chai, C. S., Deng, F., Tsai, P. S., Koh, J. H. L., & Tsai, C. C. (2015). Assessing multidimensional students' perceptions of twenty-first-century learning practices. *Asia Pacific Education Review*, *16*, 389-398. https://doi.org/10.1007/s12564-015-9379-4

Chin, K. Y., Hong, Z. W., & Chen, Y. L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Transactions on learning technologies*, *7*(4), 333-345. https://doi.org/10.1109/TLT.2014.2346756

Choi, S., Jeong, S., Park, S., Kim, Y., & Han, I. (2026). Analyzing teacher–AI interaction patterns across teacher experience and AI proficiency in student-centered lesson design. *Teaching and Teacher Education*, *169*, 105266. https://doi.org/10.1016/j.tate.2025.105266

Daradoumis, T., & Arguedas, M. (2020). Cultivating students' reflective learning in metacognitive activities through an affective pedagogical agent. *Educational Technology & Society*, *23*(2), 19-31. https://www.jstor.org/stable/26921131

Fang, J. W., Guo, X. G., Meng, X. P., Hwang, G. J., & Tu, Y. F. (2025). Realization of parasocial interaction theory in STEAM education with an AI pedagogical agent: insights from learning performance, collaboration tendency, and problem-solving tendency. *Interactive Learning Environments*, 1-24. https://doi.org/10.1080/10494820.2025.2508326

Fang, J. W., He, L. Y., Hwang, G. J., Zhu, X. W., Bian, C. N., & Fu, Q. K. (2022). A concept mapping-based self-regulated learning approach to promoting students' learning achievement and self-regulation in STEM activities. *Interactive Learning Environments*, *31*(10), 7159-7181. https://doi.org/10.1080/10494820.2022.2061013

Hostetter, J. W., Conati, C., Yang, X., Abdelshiheed, M., Barnes, T., & Chi, M. (2023, September). XAI to increase the effectiveness of an Intelligent Pedagogical Agent. In *Proceedings of the*

23rd ACM International Conference on Intelligent Virtual Agents (pp. 1-9). https://doi.org/10.1145/3570945.3607301

Hwang, G. J., Yang, T. C., Tsai, C. C., & Yang, S. J. (2009). A context-aware ubiquitous learning environment for conducting complex science experiments. *Computers & Education*, *53*(2), 402-413. https://doi.org/10.1016/j.compedu.2009.02.016

Kalmatov, R., Karimovna, K. G., Zhamin, S., & Rahim, F. (2025). Artificial intelligence. In *The Palgrave Encyclopedia of Disability* (pp. 1-15). Cham: Springer Nature Switzerland.

Kim, Y., & Baylor, A. L. (2016). based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education*, *26*, 160-169. https://doi.org/10.1007/s40593-015-0055-y

Leavy, A., Dick, L., Meletiou-Mavrotheris, M., Paparistodemou, E., & Stylianou, E. (2023). The prevalence and use of emerging technologies in STEAM education: A systematic review of the literature. *Journal of Computer Assisted Learning*, *39*(4), 1061-1082. https://doi.org/10.1111/jcal.12806

Lesseig, K., Nelson, T. H., Slavit, D., & Seidel, R. A. (2016). Supporting middle school teachers' implementation of STEM design challenges. *School Science and Mathematics*, *116*(4), 177–188.

Li, T., & Zhan, Z. (2022). A systematic review on design thinking Integrated Learning in K-12 education. *Applied Sciences*, *12*(16), 8077. https://doi.org/10.3390/app12168077

Ng, D. T. K., Su, J., & Chu, S. K. W. (2024). Fostering secondary school students' AI literacy through making AI-driven recycling bins. *Education and Information Technologies*, *29*(8), 9715-9746.

Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity*, *37*,

Wu, Q., Lu, J., Yu, M., Lin, Z., & Zhan, Z. (2022, January). Teaching design thinking in a C-STEAM project: A case study of developing the wooden arch bridges' intelligent monitoring system. In Proceedings of the 2022 13th International Conference on E-Education, E-Business, *E-Management, and E-Learning* (pp. 280-285). https://doi.org/10.1145/3514262.3514313

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2024). Pedagogical agents communicating and scaffolding students' learning: High school teachers' and students' perspectives. *Computers & Education*, *222*, 105140.

数字化转型下代际教育期望铸就西藏边境青少年幸福的理路

How Intergenerational Educational Expectation Shapes Adolescents' Well-being in Tibet's Border Areas under Digital Transformation

龙彦文，吴瑞林*，黄可盈
北京航空航天大学人文社会科学学院
wuruilin@buaa.edu.cn

【摘要】代际教育期望一致性与青少年幸福紧密相关，数字化学习是青少年自我实现的重要支点。基于2025年西藏10250位中小学生和169位校长的调查，构建有调节的中介效应模型发现，代际教育期望一致性正向影响学生主观幸福感，在边境地区正向影响被削弱。数字化学习投入是代际教育期望一致性影响主观幸福感的中介因素，其中，在边境地区发挥完全中介作用，在非边境地区发挥部分中介作用。随教师支持水平提升，数字化学习投入的中介作用持续增强，且在边境地区增幅更大。落实家校社协同育人，推动区域协调发展，发挥边境在地优势；弥合家庭禀赋差距，辐射学校公共价值；把握数字赋能机遇，实现个体整全成长。

【关键词】 代际教育期望；主观幸福感；学习投入；教育数字化转型；家校社协同

Abstract: *The consistency of intergenerational educational expectations is closely related to teenagers' happiness. Digital learning can support teenagers to realize life values. Drawing on ecosystem theory, this study analyzes data collected from 10,250 elementary and secondary school students and 169 school principals in Tibet in 2025. The results indicate that the consistency of intergenerational educational expectations positively impacts on students' subjective well-being, although this positive impact is attenuated in border counties. Digital learning engagement mediates the relationship between the consistency of intergenerational educational expectations and subjective well-being, exhibiting a complete mediating effect in border counties and a partial mediating effect in non-border counties. As the improvement of teacher support levels, the mediating effect of digital learning engagement is strengthened, with a more pronounced increase in border counties. Implementing home-school-community collaborative education. Promoting regional coordinated development and leveraging local advantages in border areas, narrowing disparities in family endowments and enhancing the public value of schools, seizing the opportunity of digital empowerment and achieving holistic adolescent growth.*

Keywords: intergenerational educational expectation, subjective well-being, learning engagement, educational digital transformation, home-school-community collaboration

1.引言

党的二十大报告提出"支持革命老区、民族地区加快发展，加强边疆地区建设，推进兴边富民、稳边固边"。处于中国西南边陲，西藏边境教育对内承担着文化传承、稳定边疆的国家

责任，对外担负着“人文交互、文化辐射、精神守土”的特殊使命。由于地理位置偏远、自然环境恶劣、民族构成复杂、经济发展滞后，西藏边境基础教育面临人口结构失衡加速、县级教育保障能力薄弱、师资质量偏低、内涵发展动力不足等挑战（李建民，2024）。教育数字化促进了教育资源共享，催生了学习方式变革，为西藏边境基础教育打破发展桎梏、实现质量跃升注入了新动能（万力勇，2022）。适应数字化转型、提升学习投入、力争学业成就是西藏边境青少年积累人力资本、打破贫困代际传递、通向美好生活的关键途径。教育公平始于期望公平。代际教育期望一致性能够激发子代学习动机，促使其在学习实践中获致充实与满足。然而，在西藏边境，受家庭社会经济地位和“贫困文化”影响，部分家长认为接受教育挤占了从事劳动创收的现实机会，对子代教育投入的热情降低，采取“自然放任型”教养方式，代际教育期望缺乏共识。从积极发展观来看，过于倚重教育不平等再生产的结构性力量将抹杀行动者的主体性。伴随教育数字化转型，处于家庭先赋性资源弱势的边境青少年能否通过自致性学习投入，从学校系统和教师获取资源补足，将代际教育期望落实为幸福？对边境青少年能动性实践的真实写照，将为家校社协同助推青少年健康快乐成长提供有益借鉴。

2.理论基础与研究假设

幸福的生产是个体与环境的复合函数。美国心理学家布朗芬·布伦纳提出的生态系统理论认为，个体嵌套于交互影响的一系列环境系统之中，这些系统与个体相互作用并影响着个体发展。查尔斯·扎斯特罗进一步建构了“微观——中观——宏观”系统模型，阐释了个体行为和成长与社会环境中多系统之间的交互关系。根据爱普斯坦的交叠影响域理论，家庭、学校、社会在个体成长中会产生交互叠加的影响，应以“关怀”为核心进行高质量互动（张俊，2019）。综上，本文提出图 1 所示理论框架，以个体主观幸福感为旨归，将个体数字化学习投入作为转化支点。家庭是与个体发展联系最密切的微观系统，代际教育期望一致性是激发个体数字化学习投入的动因。学校是个体社会化的中观系统，教师承担育人的主体责任，教师支持能够激发个体数字化学习投入进而影响其幸福感。社会作为宏观系统，形塑了个体与家庭、学校互动的制度规范和文化土壤。

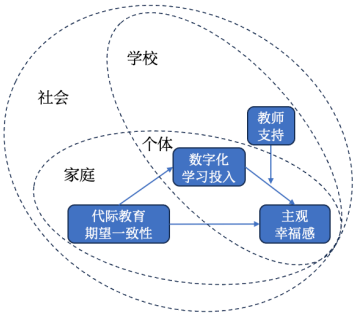


图 1“个体-家庭-学校-社会”系统框架

2.1 代际教育期望与主观幸福感

代际教育期望包括父代对子代、子代对自我未来最高学业成就的希望和期待。父母教育期望与自我教育期望越接近，表示代际教育期望一致性越高。主观幸福感指青少年以家庭生活和学校生活为基础，对自我存在状态的积极主观心理体验（Diener, 1984）。代际教育期望一致性较高时，亲子联结和互动往往更加紧密。而亲子沟通能够满足农村青少年的心理需求，进而提升其幸福感（郭海英，2014）。而代际教育期望一致性较低时，容易产生代际矛盾和冲突，造成对孩子的心理压迫（王建武，2025）。在西藏边境，受艰苦恶劣的自然环境影响，家庭教育内容偏重生产技能、风俗习惯等方面。家长多从事边贸、农林牧渔等职业（徐媛媛，

2025），由于文化水平不高，对孩子学业的辅导有限。且教育投入大、回报周期长、优质教育机会稀缺，部分家长认为“读书无望”，对孩子学业失败的宽容度不高，甚至早早认为孩子并非“读书的料”。加之家校距离遥远，孩子长期在校寄宿，产生了“脱嵌化”的家庭关系，亲子间缺乏深度沟通和相互理解，对于学业目标缺乏共识。而家庭支持的缺位会使孩子出现心理偏差和行为失范的风险增加（赵家红，2021）。综上，提出如下假设：

H1a：代际教育期望一致性对主观幸福感产生正向影响。

H1b：在边境，代际教育期望一致性对主观幸福感的正向影响被削弱。

2.2 数字化学习投入

教育期望的达成是具身化过程，需要个体持续、自主的学习投入。数字技术再造了学习生态。置身于多模态数字化学习场景，学习者借助图像、视频等视觉媒介促进认知加工，学习兴趣和想象力被激发，即时互动体验得到深化。代际教育期望一致性较高时，子代能够在相对宽松的外部期待中自主设定目标，自愿投入更多时间和认知资源参与学习（吴贤华，2021）。代际学业目标的合致使孩子信赖父母，敢于探索，遭遇学业挫折时能够得到家长的理解、接纳和鼓励，孩子的情绪更加稳定，抗逆思维能力更强（李玲，2025）。课堂是学习的主战场。根据资源补充理论（Ross, 2006），教育数字化释放的“学习红利”能够对家庭和区域资源“先天不足”的边境学生产生更大的边际效应，先赋性资源匮乏的边境学生珍视“变道超车”的数字化机遇，充分发挥主观能动性，独立思考、解决问题、迁移创新，在追寻学习目标的过程中产生意义确认和价值认同，获得不断超越自我的充实体验。由此，提出如下假设：

H2a：数字化学习投入在代际教育期望一致性与主观幸福感之间发挥中介作用。

H2b：在边境，数字化学习投入在代际教育期望一致性与主观幸福感之间的中介作用更大。

2.3 教师支持

教师的关注、鼓励和信任能够提升学生的学业成绩、心理健康、自我认同和学校适应（邵卓越，2024），降低学生的焦虑、抑郁情绪。教师支持通过自尊、心理弹性的中介作用，促进学生的学习投入（许怀雪，2023）。在西藏边境，受自然条件和人文环境限制，家庭发展能力较弱，教师是学生弥补家庭文化资本不足、获取外部支持的重要他人。教师的引导能够提升学生及其家长对接受教育的重视程度。当学生遭遇学业挫折、自我放弃时，教师的指点能够帮助学生重塑信心，提升自我效能感，缓解学业焦虑，推动其对自我的肯定与悦纳。在教师的关怀下，学生充满安全感和归属感。对教师的尊重使学生自觉理解教育目标并激发自我潜能，形成积极的学习态度和良好的学习习惯；对教师的感激使学生自愿付出更多努力，在投入学习实践中感到幸运与满足。由此，提出如下假设：

H3a：随教师支持水平的提升，数字化学习投入在代际教育期望一致性与主观幸福感之间的中介作用增强。

H3b：在边境，随教师支持水平的提升，数字化学习投入在代际教育期望一致性与主观幸福感之间的中介作用增幅更大。

3.研究设计

3.1 数据来源

2025 年 4 月，课题组赴西藏昌都市、日喀则市中小学开展预调研，通过与教师、校长座谈的方式了解学校教育数字化转型、学生心理健康等方面状况。2025 年 7 月对西藏 7 个地市的五年级、七年级学生展开问卷调查，采用分层抽样、整班抽取的方式，根据各地市和地区内各区县的人均 GDP 对区县进行分层，根据各地市和地区的学校总数量确定该地抽取的学校数。调查内容涉及代际教育期望、数字化学习投入、教师支持、主观幸福感等内容。同时，向所调查学校的校长了解教研培训、家校合作等情况。删除缺失和无效问卷后，纳入分析了

来自 42 所初中和 127 所小学的 10250 份学生问卷、169 份校长问卷。其中，包括来自 18 所边境初中和 58 所边境小学的 3234 名学生和 76 名校长。全部学生样本中，男生占 48.6%、女生占 51.4%；藏族学生占 96.4%；城市学校学生占 26.3%，乡村学校学生占 73.7%；五年级学生占 65.2%，七年级学生占 34.8%。学生的平均年龄为 13.87（SD=0.77）。

3.2 变量说明

代际教育期望包括父母教育期望和自我教育期望，测量希望将来学业达到的水平，设置没有要求、初中、高中、大学、研究生 5 个层级，从 0-4 依次编码。父母教育期望与自我教育期望差值的绝对值越小，表示代际教育期望一致性越高。主观幸福感参考世界卫生组织幸福感指数量表，测量学生感到快乐，心情舒畅、充满活力，精力充沛等 4 道题的均值。采用里克特四点计分，“0”表示从不，“3”表示总是，科隆巴赫α系数为 0.804。数字化学习投入包括学生在数字化课堂环境中的行为、认知和情感投入。测量老师在使用课件、视频等数字媒介开展教学时，学生是否感到有趣、能够理解、解除学习困惑等，共 5 道题的均值。采用里克特五点计分，“0”表示非常不同意，“4”表示非常同意，科隆巴赫α系数为 0.837。教师支持测量班主任关注与鼓励（贾娟，2012）。包括学生是否从班主任的神情或言语中感受到认可、成绩落后时受到班主任鼓励等 6 道题的均值。采用里克特五点计分，“0”表示非常不符合，“4”表示非常符合，科隆巴赫α系数为 0.844。

4.研究结果

4.1 描述性统计与差异性分析

分析全样本、边境县学生与非边境县学生各个特征的水平、分布以及边境县与非边境县之间的差异。边境县学生的主观幸福感略高于非边境县学生。边境县父母教育期望的均值为 2.73，显著低于非边境县的 2.80。边境县父母教育期望为研究生的比例为 17.81%，显著低于非边境县的 24.10%。边境县自我教育期望的均值为 2.96，显著低于非边境县的 3.08。边境县自我教育期望为研究生的比例为 27.71%，显著低于非边境县的 36.05%。边境县代际教育期望一致的比例为 64.19%，显著低于非边境县的 66.29%。边境县父母教育期望和自我教育期望均为研究生的占 20.10%，显著高于非边境县的 13.32%。边境县学生数字化学习投入与感知教师支持水平高于非边境县。

4.2 代际教育期望一致性对主观幸福感的回归结果

如表 1 所示。代际教育期望一致性越高，主观幸福感越高(M1)，在纳入学校层面的控制变量后正向影响仍然显著(M2)，H1a 成立。为了探讨边境县与非边境县之间的差异，构建代际教育期望一致性与边境县分布的交互项，进行回归发现(M3)，在边境县，代际教育期望一致性对主观幸福感的正向影响被削弱，H1b 成立。

表 1 代际教育期望一致性对主观幸福感的回归分析

变量	M1		M2		M3	
截距	0.723***	(0.123)	0.768***	(0.129)	0.728***	(0.130)
代际教育期望一致性	0.016**	(0.006)	0.016**	(0.006)	0.027***	(0.008)
数字化学习投入	0.275***	(0.009)	0.276***	(0.009)	0.276***	(0.009)
教师支持	0.212***	(0.009)	0.210***	(0.009)	0.210***	(0.009)
代际教育期望一致性*边境县					-0.034**	(0.014)

R^2	0.264	0.265	0.265
样本量	10250	10250	10250

注：控制变量：性别、年龄、民族、独生子女、家庭社会经济地位、年级、寄宿、父母同住、学校城乡分布、边境县分布、数字化教研、教师培训、教研员指导、家校合作。汇报主要影响变量的回归系数、标准误；*p<0.1;**p<0.05;***p<0.01。

4.3 数字化学习投入的中介作用分析

在全样本中，代际教育期望一致性正向影响数字化学习投入，纳入数字化学习投入对主观幸福感进行回归，代际教育期望一致性对主观幸福感的影响效应下降，说明数字化学习投入在代际教育期望一致性与主观幸福感之间发挥中介作用，H2a 成立。在边境县样本中，纳入数字化学习投入进行回归，代际教育期望一致性对主观幸福感的影响效应不再显著，说明数字化学习投入起到完全中介作用。而在非边境县，纳入数字化学习投入后，代际教育期望一致性对主观幸福感的影响效应下降，说明数字化学习投入起到部分中介作用，H2b 成立。采用 Bootstrap 方法进行中介效应检验，重复抽样 5000 次，纳入与回归分析一致的控制变量，以确保估计结果的稳健性。进一步分析数字化学习投入作为中介的间接效应、直接效应与总效应。在全样本中，间接效应为 0.033，占 58.38%，直接效应为 0.023，占 41.62%。在边境县，数字化学习投入完全中介了代际教育期望一致性对主观幸福感的影响。在非边境县，数字化学习投入部分中介了代际教育期望一致性对主观幸福感的影响，间接效应为 0.028，占 44.47%，直接效应为 0.035，占 55.53%。

4.4 教师支持对数字化学习投入中介作用的调节效应分析

数字化学习投入在代际教育期望一致性与主观幸福感之间的中介作用随着教师支持水平提升得到增强，H3a 成立。如图 2 所示，边境县回归方程直线的斜率为 0.007，大于非边境县回归方程直线的斜率 0.005，即在边境县，学生感知到更强的教师支持时，能够更大程度激发自身数字化学习投入，进而提升幸福体验，H3b 成立。

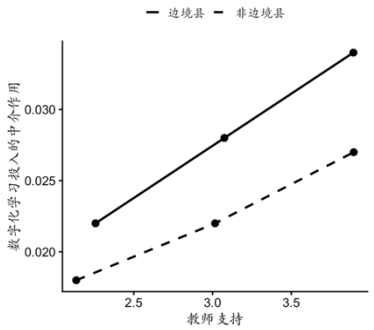


图 2 不同教师支持水平下数字化学习投入的中介作用

5.结论与启示

结构性因素经由行动者对自身生存环境的理解和参与产生影响，个体在数字化学习实践中创生幸福密钥。本文发现，代际教育期望一致性对主观幸福感产生正向影响。代际教育期望一致时，亲子凝心聚力，家庭内部资源传递和积累的效率更高。边境地区家庭面临收入结构单一、文化资本不足、社会网络局限等困境，家庭教育投入的经济负担重，子女接受教育的机会成本更高。低社会经济地位的家庭多采取“成就自然成长”的养育方式（Lareau，2011），将教育责任托付给学校，缺少对孩子学业目标的引导和规划，造成代际教育期望一致性对主观幸福感的正向影响被削弱。数字化学习投入是代际教育期望一致性转化为幸福的重要中介，其在边境地区起到完全中介作用，说明努力对社会底层家庭子女的意义更大（阳义南，2024）。本文侧重探讨课堂场域的数字化学习投入，弱化家庭社会经济地位可能导致的课外数字化学

习不公。边境个体充分利用学校公共性资源弥补家庭原始禀赋的弱势，在自致性学习投入中获致幸福体验。说明处境不利学生可以依靠自然心性和人伦传统的精神力量，塑造担当、乐观、尊重、感恩等“底层惯习”，形成具有差别优势的“亲学校文化”。教师支持强化了上述中介作用。且在边境地区，随教师支持水平提升，数字化学习投入中介作用的增幅更大，表明学生更需要和依赖教师支持，从教师的帮助中受益更明显。董永贵（2022）等发现对教育改变命运的“信”和对教师的“从”是底层文化资本发挥作用的密钥。教师通过提供文化资源、赋予正向情感和建构社会关系，能够满足学生自主性、胜任感和归属感的内在心理需求，为学生改变不利处境、跨越底层藩篱、攀登幸福阶梯积累智识与动能。党的二十大报告提出“健全学校家庭社会育人机制”。应确保政府统筹、部门协作、学校主导、家庭尽责、社会参与，扎实推进家校社协同育人“教联体”落地生根，为青少年打造健康快乐成长的“大本营”。

5.1 推动区域协调发展 发挥边境在地优势

习近平总书记指出：“改变藏区面貌，根本要靠教育。”处在地理区位、资源禀赋与经济基础的相对劣势，青少年的幸福感仍是可调节、可改变的主观社会建构。罗尔斯强调教育以“差别原则”保障弱势群体的教育权利。要立足实际，构建服务边境地区可持续发展的优质均衡基础教育体系，践行“关注每一个、发展每一个、成就每一个”。其一，遵循教育数字化环境与人文环境“和合共生”的逻辑，加强边境数字基础设施建设，实现“5G+智慧课堂”全覆盖。保障数字化资源的精准供给，依托“中心校+教学点”的办学模式，促进校际间优质教育资源共享。借助数字化平台进行“云端教研”，开展跨区域集体备课、同课异构等教研活动，推动课堂教学改革与创新。其二，打造边境地区高素质专业化创新型教师队伍，建立“招得来、留得住、教得好”的教师发展长效机制。创新教师编制管理制度，确保边境县教师，特别是抵边乡镇学校、村小教学点教师享有编制。完善教师培训政策，推动教育人才“组团式”援边，提升边境教师专业素养。健全教师福利保障与荣誉制度，增强边境教师从教积极性。其三，弘扬教育家精神，活化边境文化资源。增强教师投身边境教育事业发展与固边守边的责任感与使命感，厚植团结互助、吃苦耐劳、忠诚奉献的优良品质，涵养教师家国情怀。运用数字技术挖掘和呈现西藏传统文化中的非遗技艺、历史典籍等教育元素，促进民族传统文化在教育过程中的创造性转化与创新性发展。

5.2 弥合家庭禀赋差距 辐射学校公共价值

家庭是社会的基本细胞，是个体社会化的第一场域，知识、思想、习惯在代际之间无形传承，为个体身心发展提供了重要保障。应明确边境家庭养育的主体责任地位，引导家长重视教育、崇尚教育，提升家长育人胜任力。对于家长而言，首先，应理性认识教育的本真价值，助推孩子“生命成人”而非仅仅是“个体成才”。避免过度追求教育的功利性价值、以过高标准要求孩子，对孩子身心造成压抑。其次，营造温馨融洽的家庭氛围。家庭情感体验产生的“爱的倾向”是家庭代际动力的核心。家长的情感参与对孩子的学业成就有更强的预测力（邢淑芬，2023）。最后，加强代际融通对话，整合代际教育观念。充分激活家庭资本，与学校进行高质量互动。当代际教育期望达成共识，孩子能够理解家长的用意，并自觉投入更多时间和精力学习，回馈家长的支持。

学校是教育发挥影响力的制度化、系统化机构。要发挥义务教育阶段学校育人主阵地的功能，逐步改变因区位因素、家庭因素等造成的起点的不平等，实现个体发展机会和发展结果的相对公平。一方面，学校教育应与边境独特的自然环境、社会环境之间相互适应、互促共生，融入民族文化和乡土文化，在学校的现代性文化与边境地方性文化之间建立有意义联结。依托优质数字化资源开齐开足国家规定课程。并积极利用校外活动场所开展实践育人活动，促进学生健康全面发展。另一方面，教师是学生的学业引导者和心灵守护者。首先，教师应根据学生的知识结构、兴趣经验，善用数字化资源破解教学重、难点，将多媒体素材与教学

内容有机整合，营造具身化学习环境，鼓励学生自我探索，并为学生提供个性化反馈。其次，尊重每个学生的节奏，不让一个学生掉队。如第斯多惠所言，教育的本质不在于传授本领，而在于鼓励、唤醒和鼓舞。对待学习吃力、课堂沉默、学业厌倦的学生，教师要帮助其解决学习困惑，克服胆怯与自卑心理，避免学生形成“差生”“考不上大学”的自我意象，产生消极抵抗的“反学校文化”。最后，学生是有生命、情感、思想的完整的人。教师要用尊重、信任、亲近和同理心建立与学生之间的桥梁，关注学生的情绪状态，肯定学生的独特价值，通过关爱、倾听和欣赏帮助学生建立良好的自我概念，锤炼学生的意志品格。

5.3 把握数字赋能机遇 实现个体整全成长

个体本质上是由内在能动性推动的、具有无限多种发展可能性的动态生命体，而非完全依据外在结构性力量形塑的沉默者。当数字化已然成为内在性生命特征，个体需要清醒“洞察”自身的生存境遇，认识自身局限，具备内在的觉醒和改变的勇气，与家庭、学校和社会多主体进行“物质、能量和信息”的流通，滋养自身成长为乐学会学的终身学习者。首先，要在困难和逆境中意识到教育功能的重要性，改善学习态度和方法，砥砺自身投入数字化学习实践，在自主探索、解决问题、批判思考的“意义创造”过程中建立知识与现实世界的关联，将自身潜能转化为现实存在，满足对生命本体的感悟、体察和超越。其次，准确认知自我生态位，在客观评价自我的基础上，确立符合实际的教育期望，将期望细化为具体可行的目标，并合理计划、管理和监控自身行为，从近期目标过渡到长远目标，不断确证自我价值，拓展和优化发展生态空间。最后，养成正向的人生观、价值观，培养兴趣爱好，参与社团活动，排解不良情绪，提升情绪适应能力，避免过度依赖学业成就评价体系的单向度发展。

参考文献

万力勇，解攀科（2022）．教育信息化促进教育公平的现实路径——以民族地区基础教育为例．中南民族大学学报（人文社会科学版），42（2），166–173+188.

王建武，王思杨（2025）．促进还是抑制：父母教育期望与投入对子女心理健康的影响．*青年研究*，（3），175–192，198.

邢淑芬，刘伊宁（2023）.父母教育参与与学龄儿童学业成绩——多多益善还是过犹不及.*教育研究*，2023，44（08）：130-142.

阳义南，卫松（2024）．“努力”与“拼爹”影响社会地位的代际流动研究．*暨南学报（哲学社会科学版）*，46（1），112–127.

许怀雪（2023）．体认的教师支持对城乡结合区高中生学习投入的影响研究——基于学生自尊-心理弹性的链式中介效应分析．基础教育，20（3），94–102+112.

李建民（2024）．优化区域教育资源配置推动国门学校高质量发展．*中国民族教育*，（12），21–23.

李玲，姚梦繁，黄宸（2025）．早期留守经历何以影响中学生的核心技能抗逆力——来自西部农村的证据．*教育科学*，41（5），37–44.

吴贤华，刘永存，孙利（2021）．中小学生教育期望匹配、家庭教养方式与学习投入的关系．基础教育，18（2），42–52.

邵卓越，董玥彤，安桂清（2024）．数智时代教师支持对学生数学学习结果的影响——基于 PISA 2022 数据的发现及启示．全球教育展望，53（4），30–42.

张俊，吴重涵，王梅雾，等（2019）．面向实践的家校合作指导理论——交叠影响域理论综述．教育学术月刊，（5），3–12．

张越（2023）.现代化中的边境儿童及其文化生产——保罗·威利斯《在中国寻觅现代性踪迹》的另一面.终身教育研究，34（02）：13-20.

郭海英，朱婉灵，朱倩，等（2014）．亲子沟通与农村青少年幸福感的关系：基本心理需要满足的中介作用．心理发展与教育，30（2），129–136．

赵家红，任娟（2021）．西藏教育强边的现实意义与实现路径．*西藏民族大学学报（哲学社会科学版）*，42（6），142–146．

贾娟（2012）．教师情感支持及其对中学生自尊、学业自我效能感的影响研究．西南大学．

徐媛媛，刘玉杰（2025）．边境地区教育赋能强国建设：战略效能、制约因素与实施路径．*民族高等教育研究*，13（6），23–29．

董永贵，王静宜（2022）．“信”与“从”：底层文化资本发挥作用的密钥．*中国青年研究*，（1），104–110，119．

尤尔根·哈贝马斯（1994）．交往行动理论：第 2 卷，论功能主义理性批判（洪佩郁，蔺青，译）．重庆：重庆出版社．

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*(3), 542–575.

Epstein, J. L. (2010). School, family, community partnerships: Caring for the children we share. *Phi Delta Kappan*, 92(3), 81–96.

Lareau, A. (2011). *Unequal childhoods: Class, race, and family life* (2nd ed.). Berkeley, CA: University of California Press.

Ross, C. E., & Mirowsky, J. (2006). Sex differences in the effect of education on depression: Resource multiplication or resource substitution? *Social Science & Medicine, 63*(5), 1400–1413.

Rawls, J. (1988). *A theory of justice* (H. He, B. He, & S. Liao, Trans.). China Social Sciences Press. (Original work published 1971)

Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 well-being index: A systematic review of the literature. *Psychotherapy and Psychosomatics*, 84(3), 167–176.

从静态测评到三维表征：大学生批判性思维的进阶测评框架构建及其实践应用

From Static Assessment to Three Dimensional Representation: Advanced Assessment and Practical Application on Critical Thinking of College Students

程玉梅[1]， 王瑞宁[1] ，乔爱玲[1]
[1]首都师范大学 教育学院

【摘要】 批判性思维作为培养学生核心素养的基石，也是高等教育在数字化转型期间培养创新型人才的关键抓手，但关于批判性思维的进阶测评仍存在测评框架缺乏层级表征、对批判性思维发展过程数据挖掘不足，难以揭示思维动态发展的问题。基于文献研究与德尔菲专家咨询法，重构涵盖行为表征、认知调节与情感倾向的大学生批判性思维进阶测评框架。进而，通过三维的认知网络分析对实证数据进行深度挖掘揭示了不同阶段的学习者分别在群体、小组及个体层面的认知网络特征。研究发现学习者的批判性思维进阶并非线性增长，呈现出由离散扁平向立体耦合跃迁的结构化特征，且情感倾向在思维进阶中发挥着关键作用；并围绕小组协作学习活动设计、智能体提示词与工作流的设计逻辑提出如何针对测评结果反哺批判性思维的进阶培养。

【关键词】批判性思维；批判性思维进阶测评；德尔菲法；认知网络分析

Abstract: *Critical thinking, as the cornerstone of cultivating students' core competencies, is also a key lever for higher education to cultivate innovative talents during the digital transformation period. However, there are still problems with the advanced evaluation of critical thinking, such as the lack of hierarchical representation in the evaluation framework, insufficient data mining of the development process of critical thinking, and difficulty in revealing the dynamic development of thinking. Based on literature research and Delphi expert consultation method, reconstruct an advanced assessment framework for critical thinking of college students that covers behavioral representation, cognitive regulation, and emotional orientation. Furthermore, deep mining of empirical data through three-dimensional Epistemic Network Analysis revealed the cognitive network characteristics of learners at different stages at the group, group, and individual levels. Research has found that learners' critical thinking progression does not increase linearly, exhibiting a structured transition from discrete flatness to three-dimensional coupling, and emotional tendencies play a key role in thinking progression; And propose how to cultivate advanced critical thinking based on the design logic of group collaborative learning activities, intelligent agent prompt words, and workflow, in response to the evaluation results.*

Keywords: Critical Thinking; Advanced Critical Thinking Assessment; Delphi Method; Epistemic Network Analysis

1.引言

批判性思维作为 21 世纪核心素养的基石，不仅是 5C 能力之一(Ravitz et al., 2012)，更是高等教育在数字化转型期培养创新型人才的关键抓手。然而，随着生成式人工智能(Generative Artificial Intelligence, GenAI)的广泛应用，学习者面临认知外包的风险(余胜泉 & 汪凡淙, 2023)，若学习者缺乏独立的主体性与批判性审视能力，极易陷入对生成内容的无批判性接受，丧失独立思考与深度学习的能力。因此，在人机协同新生态下，如何通过精准的测评引导学生批判性思维进阶发展已成为教育技术领域亟待解决的命题。尽管学术界对批判性思维的重视程度日益提升，但现有的测评仍存在显著结构性滞后，表现为绝大多数测评工具仍沿用标准化测验或自我报告式问卷。这种结果导向的静态测评关注的是学习者在某一时间点上所具备的批判性思维的状态，虽然能通过量化分数划分能力等级，却割裂了思维要素间的复杂的耦合关系，难以揭示学习者在解决复杂问题过程中思维结构的动态演化轨迹与不同阶段的思

维表现差异和发展规律。同时，研究指出学生的思维发展并非突变式，而是呈现由低阶到高阶的层次性增长(Kaldaras et al., 2024)，学生的批判性思维也应当具有螺旋递进式的发展特征(Liu et al., 2014)。然而，目前关于批判性思维的进阶测评研究仍存在测评框架缺乏层级表征、对批判性思维发展过程数据挖掘不足以及测评结果与教学反馈闭环缺失三大短板。

为突破上述瓶颈，学习科学中的学习进阶理论为透视的思维发展提供了过程性视角，强调能力发展呈现由低阶到高阶的层次性跃迁(Wilson, 2009)。同时，随着量化民族志与认知网络分析(Epistemic Network Analysis, ENA)等技术的发展，得以从话语交互数据中提取认知结构，将隐性的思维过程转化为可视的网络拓扑图，通过分析节点间连接强度的时空变化，能够动态评估学习者专业思维发展的显著特征(Shaffer, 2017)。这为从关系与结构维度重新审视批判性思维进阶提供了新的方法论支撑。

鉴于此，本研究试图构建面向大学生的批判性思维进阶测评框架，以 3D ENA 为工具将隐性的思维结构转化为可视化的网络拓扑图，通过实证数据揭示学习者在不同进阶阶段的认知网络特征。本研究将重点回答以下问题：（1）如何通过专家共识形成科学的大学生批判性思维进阶表征框架？（2）基于多模态话语数据，探究不同进阶水平的学习者在群体、小组及个体层面呈现出怎样的认知网络拓扑特征？（3）批判性思维测评的结果如何反哺批判性思维的进阶培养？

2.大学生批判性思维进阶测评的表征框架构建

2.1 批判性思维的结构与进阶测评的表征框架设计

鉴于批判性思维是一种多维度、多路径并发的复杂认知过程，其发生过程往往伴随行为、认知与情感因素的交织协同。因此，研究以系统性、交互性与层级性为基本原则，以国内外具有代表性的四种批判性思维结构为参照，以“行为—认知—情感”三位一体的系统视角对批判性思维结构进行重新设计。研究采用**德尔菲专家咨询**的方式确定批判性思维结构与进阶测评的表征框架。批判性思维结构中，行为指认知技能外显的行为表现，聚焦批判性思维在具体学习任务中的言语表达、问题探究、观点建构、协作辩论等外显行为；认知指元认知主导的自我反思，体现个体在面对复杂、不确定、冲突信息时主动修正、反思与优化思维路径的能力；情感指思维意愿与价值导向的情感驱动。批判性思维进阶测评的表征框架部分，研究参考吴亚婕等人在研究中修订的批判性思维三层次评分量表(吴亚婕 et al., 2015), 在批判性思维结构的基础上，将二级维度划分为低阶、中阶与高阶三个水平层次。最后，研究初步设计了包括 3 个一级指标、11 个二级指标、33 个观测点的批判性思维进阶测评的表征框架。

2.2 批判性思维进阶测评表征框架的修正

为验证批判性思维进阶测评表征框架的科学性与有效性，研究编制了批判性思维进阶测评表征框架的专家意见征询问卷。研究选取 17 名高校教育技术学专业的教师作为专家，采用微信与邮件发送的方式进行。经过统计分析，两轮专家意见征询后，各个指标项的平均值逐渐增大，变异系数逐渐缩小（如表 1 所示），表明专家意见逐渐趋同。

表 1 两轮专家意见征询数据分析结果

征询轮次	征询人数	均值范围	标准差	变异系数临界值	变异系数
第一轮	17	2.65～4.82	0.39～1	0.25	2 项（0.27、0.28）
第二轮	15	4.31～5	0～0.52	0.12	0 项>0.12

2.2.1 第一轮专家意见征询

量化分析结果得出，第一轮专家的权威系数为 0.812，高于 0.7，表明专家具有较高的权威性，咨询结果可靠(吴扬, 2018)。根据专家反馈意见，研究对表征框架的各维度设置、指标命名与各观测点的表述进行优化，修改内容包括：(1)从一级指标维度和二级指标维度来看，行为主要指认知的外显行为，认知主要指元认知，建议一级维度名称调整为行为表征、认知调节与情感倾向；(2)二级指标的解释与阐释有重复嫌疑，建议与解释合并，且推断与分析均属逻辑推理的核心环节，合并为分析可避免概念重叠。(3)按照元认知理论所强调的行为修正需依赖策略选择，体现元认知的主动性，建议修改为自我调节；(4)思维开放性表述侧重态度，建议修改名称，修改后的概念应涵盖对自身认知局限的觉察、对自身认知的谦逊，同时，“认知公正性”所指内容与思想开放性中部分重合，建议合并，且合并后的名称修改为“思维的开放性”；(5)求知动机和批判性思维之间似乎缺少直接的逻辑关系，按照上述结构框架，是否指面对复杂问题时需持久探究的意志力，建议修改为探究欲；(6)针对三阶观测点，专家认为部分二级指标的三阶层次递进不明确或表述主题不一，建议系统审查。

2.2.2 第二轮专家意见征询

第二轮专家意见征询主要主要的修改内容如下:(1)各个观测点表述中举例内容的关键取决于省略号后面的语义信息如何，建议表中谨慎举例。(2) 建议各个观测点的低阶、中阶与高阶的起点水平要一致。经过两轮专家意见征询后的批判性思维的结构如表 2 所示，批判性思维进阶测评表征框架如表 3 所示。

表 2 批判性思维的结构

一级维度	二级维度	能力表现
行为表征：认知技能外显的行为表现		
	解释	能够准确理解并表达各类信息的含义或用论证的形式清晰地呈现自己的认知过程及结果
	分析	能够对多种形式的信息进行分解，准确识别各类信息中蕴含的要素及其功能
	评价	能够恰当地评价信息来源的可信度以及信息内容的逻辑性
认知调节：元认知主导的自我反思		
	自我监控	对自己的认知活动及活动中使用的要素，推导结果等进行监控，反思个人态度、情感、动机、价值观是否公正
	自我评价	对自己的认知活动、推理过程、结果做出分析与判断
	自我调节	自我检查在意识到错误或不足后，及时补救或修正错误
情感倾向：思维意愿与价值导向的情感驱动		
	思维的开放性	尊重容纳不同观点，避免立场先行或预设偏见，以一贯标准对待各种观点、证据与立场，展现理性与客观
	探究欲	面对复杂问题时是否展现出持续探究的兴趣、愿意提出问题

表 3 批判性思维进阶测评表征框架

一级维度	二级维度	三阶观测点	描述
行为表征	解释	低阶水平	能简单复述语境中的观点、概念、问题、标准、规则等的含义，但存在表达模糊、结构松散或与原意偏离明显的情况； 对概念含义解释笼统、缺乏上下文依据
		中阶水平	能较准确地说明信息或解释语境中核心观点等内容，但仍无法结合上下文进一步阐释，仅能提供简单说明
		高阶水平	能准确理解信息含义，结合上下语境进行对比解释，常伴随举例或关联推理
	分析	低阶水平	能识别部分信息，但无法区分主次要点或缺乏对结构的认识，分析视角局限，列出观点

			但未具体说明
		中阶水平	能识别信息中的基本结构与主次内容，尝试比较观点间的差异或联系，对话中出现对比分析行为
		高阶水平	能揭示观点、证据与逻辑之间的隐含关系，或指出矛盾、偏差或逻辑漏洞。如忽略了……，没有解释……
	评价	低阶水平	能够在对话情境中不判断或无法判断信息的可信度，容易顺从他人的观点，缺乏主动质疑的意识和深度的逻辑分析
		中阶水平	能够从可信度、相关性、价值性等多维度评估信息，或识别信息中的逻辑漏洞或不合理之处，能够提出质疑，但更多是基于直觉或简单思考
		高阶水平	能够辩证地分析信息的可靠性，主动质疑信息中的假设、偏见和潜在错误，并提出观点支撑或修正方向
认知调节	自我监控	低阶水平	极少在发言中出现对自己理解或任务进展的意识描述，只陈述观点，缺少“我是否理解正确”“是不是可以这样想”等元话语。
		中阶水平	在对话中不确认自己理解或表达的正确与否，依赖他人反馈，或出现自我觉察语言
		高阶水平	能够在对话中清晰反思自己是否存在理解偏差、思路是否清晰、判断是否合逻辑，对话中出现“我觉得我之前理解偏了，重新来看……”“我们应该忽略了……”“我改一下之前的说法，因为……”等行为。
	自我评价	低阶水平	学生在对话过程中缺少自我评价
		中阶水平	仅从单一或片面或结果导向对个人观点进行简单判断，缺乏相应证据支撑或自我评价维度单一
		高阶水平	能够对自身观点进行客观评价，并进行自我反思与分析
	自我调节	低阶水平	对个人观点中存在的漏洞、缺陷或问题，无任何反馈行为
		中阶水平	能够意识到个人观点的缺陷，并尝试提出修正方案，但缺乏针对性与可行性
		高阶水平	能够反思自己的错误或不足，设计有效的补救方法
情感倾向	思维的开放性	低阶水平	倾向于维护自我立场，对他人观点不予理解或不考虑他人证据与理由
		中阶水平	能考虑或能认可他人观点，或被动地对异议进行客观评估，但尚未能主动地将不同观点联系起来进行深入剖析。
		高阶水平	主动整合多方观点，引入异见进行对比，体现公正性与对他人观点的包容性
	探究欲	低阶水平	在外部任务驱动下被动回应问题，缺乏主动提问或进一步探究意图
		中阶水平	能在任务情境中提出需要进一步探索的问题，展现基本的探究意愿
		高阶水平	能够在对话或任务情境中持续提出问题，或者在多个角度提出问题，体现出持续、主动的问题探究意识

3.大学生批判性思维进阶测评表征框架的实践应用

3.1 应用情境

3.1.1 研究对象

研究以北京市某高校教育技术学专业的 49 名本科二年级师范生为研究对象，平均年龄 20-21 岁，男生 28 人，女生 21 人。所有学生均具有网络学习经验，能够利用计算机进行信息检索以及与 GenAI 对话完成小组协作学习任务。

3.1.2 研究过程与研究情境

研究在本科生《信息技术教育》课程中开展，研究过程共 10 周。第一阶段（第 1-2 周）为批判性思维水平前测并进行学生分组，批判性思维水平前测为未进行任何干预下对学生在

师星学堂讨论区围绕给定议题进行的发帖与回帖内容进行编码与分析。第二阶段（3-10 周）为教学干预过程，主要开展由生成式智能体赋能的小组协作学习。完成小组任务过程中，所有学生均需要与智能体进行对话并通过微信群或录音的方式进行小组讨论以完成协作任务，学生所参与的学习活动内容均相同。教学干预结束后，对收集到的所有数据（学生个体与智能体的文本对话内容、小组微信群讨论、小组录音等）进行文本转录，并进行编码与分析。

3.2 数据收集与分析

研究收集了教学干预前后学生发帖与回帖以及学生与生成式多智能体的对话内容，用来分析学生的批判性思维水平发展与特征。在删除无效对话内容后，共收集 1273 条对话数据。随后，两位编码员首先对随机选取的 537 条（占全部对话数据的 42%）数据分别进行独立编码，编码一致性为 0.856，并对不一致地方进行充分讨论，随后分工完成剩余数据的编码。

3.3 研究结果与分析

研究采用 ENA 的方法对学生的批判性思维水平进行分析，探索学生干预前后的批判性思维发展路径。ENA 常用于分析学习者对特定主题的观点，以构建可视化的个人或群体的认知网络来揭示学习者概念之间的关系和结构(Shaffer et al., 2016)。然而，传统的 ENA 通常通过奇异值分解将前两个维度在二维平面上显示，其余维度通常不被检查，这往往忽视了其他维度的重要性。因此，与上述构建的批判性思维三维结构相对应，本研究以三维空间中的认知网络分析为思路(Yu et al., 2024)，通过构建 3D ENA 分析学生批判性思维在行为表征、认知调节与情感倾向的各个维度在低阶、中阶与高阶水平的发展路径，旨在识别批判性思维发展过程中最常出现的核心要素及其共现模式，从而揭示它们的内在关系。

研究对干预前（第一阶段, First stage）与干预后（第二阶段, Second stage）的学生交互文本，分别围绕**学习者群体、学习者小组、学习者个体进行了全样本的认知网络分析**。研究发现上述三个维度的学习者认知网络结构基本保持一致，表现出由离散扁平向立体耦合的发展变化趋势，也体现出情感倾向作为思维进阶的关键作用。因此，**研究随机选择一个学习者小组为例进行展示。**

3.3.1 学习者小组批判性思维的频次分析

学习者小组批判性思维进阶前后（First stage VS. Second stage）的交互文本数据如图 3 所示。

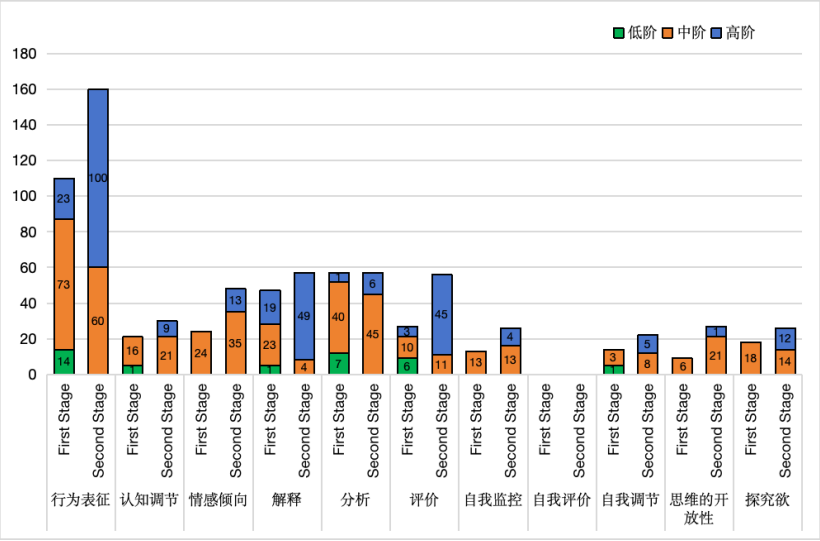


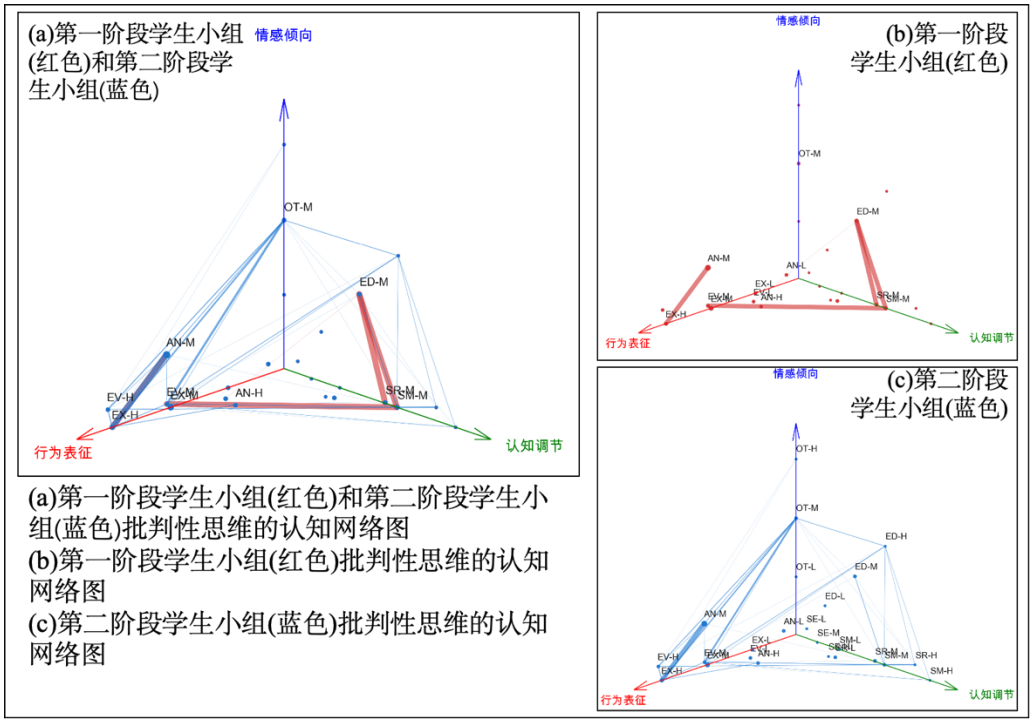
图 3 学习者小组批判性思维进阶前后的行为频次

学习者小组的批判性思维在三个一级维度的行为频次对比中，频次最高的是行为表征和情感倾向，分别占比 69.41%、18.51%；而认知调节行为较少，占 12.08%。行为表征维度中，出现行为频次最高的为分析，占比为 36.67%，解释与评价稍次之，分别占比为 35.56%、27.78%。

认知调节维度中，学生小组的自我评价在两个阶段的低中高阶的行为频次仍均为 0，且该维度的交互行为频次主要集中在自我监控与自我调节，分别占比为 63.83%、36.17%。情感倾向维度中，探究欲（61.11%）的行为频次要高于思维开放性（38.89%）。通过对以上每个维度 First stage 和 Second stage 在低中高三个层次出现的行为频次进行对比分析，经过教学干预后，学生小组第二阶段的批判性思维水平在各维度的中高阶层次中的交互行为频次明显上升。

3.3.2 学习者小组批判性思维进阶特征

研究采用 3D ENA 对两个阶段学生的交互文本进行分析，得到学习者小组批判性思维特征的可视化表征，如图 4 所示。



(a)第一阶段学生小组(红色)和第二阶段学生小组(蓝色)批判性思维的认知网络图
(b)第一阶段学生小组(红色)批判性思维的认知网络图
(c)第二阶段学生小组(蓝色)批判性思维的认知网络图

图 4 学习者小组批判性思维进阶前（第一阶段）后（第二阶段）认知网络分析图

图中 X 轴代表行为表征、Y 轴代表认知调节、Z 轴代表情感倾向，节点的大小反应批判性思维能力运用的多少，节点间连线的粗细和颜色饱和度反应批判性思维各能力水平之间的关联程度。图 4(a)展示的是学习者小组在第一阶段(红色)与第二阶段(蓝色)批判性思维的认知网络图，对比两个阶段的群组质心位置与网络拓扑结构，可以发现小组协作从浅层互助向深度协同发展，两个阶段最直观的差异体现在节点数量与 Z 轴(情感倾向)的延展度上。进一步分析发现，图 4(b)第一阶段中解释(EX-M)与分析(AN-M)与自我监控(SM-M)和探究欲(ED-M)之间的关联程度较强，也呈现出一定的思维开放性(OT-M)，但分析、自我监控、自我调节等维度在高阶水平层面不足。同时，批判性思维的节点与节点间的连线集中在由解释(EX-M)与自我监控(SM-M)、自我调节(SR-M)构成的二维层面，这表明初期的小组互动主要局限于任务层面的操作执行与基础的认知协调，缺乏高维度的情感投入。图 4(b)学生小组第二阶段批判性思维的认知网络图中，对比来看，第二阶段的学生小组展现出更多的批判性思维节点与节点之间更密集的联系。同时，解释(EV-H)、评价(EV-H)与思维的开放性(OT-M)、探究欲(ED-M)、自我监控(SM-M)、自我调节(SR-M)之间展现出更强的关联性，也呈现出一定的探究欲(ED-H)、思维的开放性(OT-H)、自我调节(SR-H)，这表明学生的批判性思维发展进阶尚可，呈现出较大程度的提升。此外，第二阶段的网络结构在 Z 轴方向有所提升，形成了以情感倾向(OT-M)为顶点的立体锥形结构。这一变化也揭示了学生小组中批判性思维进阶的关键机制，情感因素是支撑群体批判性思维进阶的核心，表明小组成员在面对复杂的问题解决时所构建起的积极的探究欲精神与思维开放精神为促进批判性思维进阶提供了契机。

研究发现，学生小组在第一阶段的认知网络较为稀疏且低矮，连线多集中在左侧的行为表征区域，且节点间的共现强度较弱。这反映出学习者初期虽然具备一定的批判性思维技能，但这些技能是碎片化的，缺乏元认知层面的调节与统筹。且在解释(EX)、分析(AN)、评价(EV)、自我监控(SM)、自我调节(SR)、探究欲(ED)、思维的开放性(OT)各个节点中，均为低阶水平或中阶水平。在学生小组在第二阶段的批判性思维状态中，认知网络呈现出显著扩容与增强态势，整个网络在体积上覆盖了更广的空间，表明个体能够调用更多元的思维策略。同时，网络连线的粗细程度(权重)普遍增加，特别是在分析(AN-H)、评估(EV-H)与自我调节(SR-M)之间形成了闭环，这意味着批判性思维已从单一的技能应用转化为一种结构化的批判性思维。

4.讨论

实证数据验证了大学生批判性思维进阶测评表征框架的有效性，基于 3D ENA 实现了对大学生批判性思维进阶过程的精准测评与可视化表征，也证实了生成式多智能体赋能的小组协作学习能有效推动大学生批判性思维的进阶发展。基于此，需要进一步探讨如何利用测评结果来优化协作学习设计和生成式智能体的干预策略，以实现批判性思维的进阶培养。

一方面，与以往研究多将批判性思维视为单一维度的能力提升或静态的技能掌握不同，本研究的 3D-ENA 结果显示，批判性思维的进阶并非简单的线性增长，而是一种网络拓扑结构的质变，这一发现有助于修正以往过于强调孤立技能强化的教学倾向，也回应并深化了以往批判性思维的层级结构(文秋芳 et al., 2009)，揭示了高阶思维的形成需要超越单一技能训练，转向对思维结构系统性的培养。因此，教学需要优化小组协作学习活动的设计，如指定小组成员进行角色扮演，设定质疑者与评估者等角色，或设置任务的阶段性门槛，要求小组在进入下一阶段前必须先提交前一阶段的反思性报告，帮助小组在行为表征与认知调节之间产生结构化的互动。此外，对于教学教学实践来说，精准有效的层级评价框架也有助于教师了解学生群体内部的批判性思维发展状态，并对学生及时进行干预和指导。如基于评测结果，教师可为学生提供深度参与小组协作的问题讨论的支架。

另一方面，本研究引入生成式多智能体作为协作主体，区别于传统仅依靠教师或同伴的协作模式。研究发现，智能体在其中扮演了脚手架的角色。当学习者陷入认知僵局时，智能体通过苏格拉底式提问而非直接提供答案，有效地激活了学习者的反思。此外，研究表明基于 GenAI 技术的对话机器人能够促进批判性思维的发展(郑兰琴 et al., 2024)；这与研究者提到的 GenAI 的协作参与可以提升学习者的批判性思维水平也保持一致(杨宁 et al., 2025)。因此，后续研究可通过提供提示词交互支架与重新设置智能体工作流的逻辑，实现基于学生交互内容的持续监测与系统化反馈。如当智能体检测到学生批判性思维的认知网络中“分析”与“自我调节” 的连接权重过低，则智能体的工作流应立即从信息反馈模式切换为元认知反思模式，促使学生审视自己的思维过程而非仅关注任务结果。同时，赋予学生特定的提示词“请你用三句话阐释对方的核心论点，以确保你对其理解是开放且客观的，然后再进行评估”，这种基于结构缺陷的诊断式支架，能够精准地靶向缺失的连接，有效促进思维的结构化发展。

5.结语

本研究构建了面向大学生的批判性思维进阶测评表征框架，实证结果表明，该框架结合 3D ENA 技术，实现了批判性思维在行为表征、认知调节与情感倾向维度的多维可视化，也从学习者群体、学习者小组与学习者个体三个维度探索了学习者认知网络结构由离散扁平向立体耦合演进的动态轨迹，验证了情感倾向作为思维进阶的关键作用，并围绕小组协作学习活动设计、智能体提示词与工作流的设计逻辑三个方向提出如何针对测评结果反哺批判性思维的进阶培养。然而，受限于数据模态与实验周期，本研究当前主要依赖文本话语分析，难

以全面覆盖面部表情、生理信号等非语言维度的潜隐性信息，且对思维进阶的长效保持机制考察尚显不足。未来研究将致力于拓展多模态数据融合视角，提升测评的颗粒度与解释力；同时，构建嵌入生成式智能体的自适应干预闭环，实现从诊断向即时反馈的测评范式转型，并开展跨学科、长周期的纵向追踪。

参考文献

文秋芳, 王建卿, 赵彩然, 刘艳萍, & 王海妹. (2009). 构建我国外语类大学生思辨能力量具的理论框架. 外语界(01), 37-43.

吴亚婕, 赵宏, & 陈丽. (2015). 网络环境下大学生批判性思维培养教学模式的实践. 现代远程教育研究(02), 76-86.

吴扬. (2018). 融合幼儿班集体教学活动评价指标体系构建——基于德尔菲法的调查研究. 中国特殊教育(10), 18-25.

杨宁, 杨锦, 马子璘, & 林岚. (2025). 生成式人工智能支持的协作问题解决对小学生批判性思维的影响研究. In 电化教育研究 (pp. 104-112).

余胜泉, & 汪凡淙. (2023). 人工智能教育应用的认知外包陷阱及其跨越. 电化教育研究, 44(12), 5-13.

郑兰琴, 高蕾, & 黄梓宸. (2024). 基于生成式人工智能技术的对话机器人能促进在线协作学习绩效吗？. 电化教育研究, 45(03), 70-76+84.

Kaldaras, L., Akaeze, H. O., & Krajcik, J. (2024). Developing and validating an Next Generation Science Standards-aligned construct map for chemical bonding from the energy and force perspective. Journal of Research in Science Teaching, 61(7), 1689-1726.

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. ETS Research Report Series, 2014(1), 1-23.

Ravitz, J., Hixson, N., English, M., & Mergendoller, J. (2012). Using project based learning to teach 21st century skills: Findings from a statewide initiative. American educational research association conference, Vancouver, Canada,

Shaffer, D. W. (2017). Quantitative ethnography. Lulu. com.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis:Analyzing the structure of connections in cognitive,social,and interaction data. Journal of Learning Analytics, 3(3), 9-45.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 46(6), 716-730.

Yu, J., Hu, D., & Wang, C.-H. (2024). Development of ENA 3D: A Tool for Epistemic Network Analysis in Three-Dimensional Space. Advances in Quantitative Ethnography, 152-165.

Review of Multi-Agent Systems (MAS) in Education

Yujia Yang[1], Jianwei Zhang[1], Yifan Zhang[1*]
[1] Smart Learning Institute of Beijing Normal University, China
cqswuyu@163.com; 2361653573@qq.com; *zhangyifan@bnu.edu.cn

Abstract: *This systematic review synthesizes the distribution of Multi-Agent Systems (MAS) studies in education from 2015 to 2025, analyzing 523 peer-reviewed studies preliminarily. The results reveal a significant surge in publications after 2024 (n = 112 in 2025), possibly driven by the integration of Large Language Models (LLMs) and accessible agentic frameworks. Higher education dominates the research landscape (61.57%), partly due to its infrastructure and flexible ethical constraints, while the K-12 and vocational sectors remain underexplored. Pedagogically, MAS applications centralize subject-specific Knowledge Mastery (n = 128) and Learning Process Optimization (n = 91), while placing less emphasis on broader 21st-century competencies such as collaboration and ethical development. This study highlights a critical disconnect between technical advancements and holistic educational needs, aiming to provide a landscape for bridging the gap between MAS innovation and pedagogical effectiveness in the generative AI era.*
Keywords: Systematic Literature Review; LLM-based Agents; Multi-Agent Systems; Educational Technology

1. Introduction and Related Work

Multi-Agent Systems (MAS), rooted in distributed artificial intelligence (AI), integrate Agents, Environments, Interactions, and Organizations (AEIO model) and operate through decentralized cooperation (Hamal et al., 2021). While early agents relied on symbolic AI or conventional machine learning approaches (Jiang et al., 2024), the integration of Large Language Models (LLMs) has revolutionized the field. By incorporating LLMs, MAS now benefits from enhanced collaboration among specialized agents, more sophisticated reasoning capabilities, and dynamic adaptation to complex environments (Li et al., 2024). These LLM-based MAS surpass the limitations of single-agent systems, demonstrating remarkable success in addressing complex tasks.

In educational contexts, pedagogical agents (PAs) have been a subject of research since the late 1990s (Martha & Santoso, 2019; Sikström et al., 2022). Earlier review studies mainly focused on specific characteristics of these agents. Some explored their visual appearance and role design within virtual learning environments (Kim & Baylor, 2016; Martha & Santoso, 2019), while others examined their communication effectiveness and pedagogical strategies, particularly in K-12 education settings (Sikström et al., 2022; Zhang et al., 2024). These prior works primarily addressed single-agent systems developed before the emergence of LLMs and thus did not reflect the capabilities or dynamics introduced by LLM-enhanced multi-agents.

Researchers and educators need to discern emerging trends and develop corresponding strategies to align MAS innovations with pedagogical priorities by supporting the development of high-order competencies such as critical thinking, creative problem-solving, collaboration, and metacognitive regulation (Miao & Shiohira, 2024). Emerging reviews on MAS focused on diversified application scenarios in education (Jiang, 2024) and the framework for LLM-integrated intelligent multi-agent collaboration in personalized education (Zhao, 2025). However, an overview of MAS distribution in educational technology remains underexplored. To bridge the gap, this study synthesizes the distribution of MAS studies in education over the past decade. It aims to identify whether there is a potential shift in MAS publications in the context of LLM applications in education, examine relevant research across various educational levels, and assess their alignment with contemporary pedagogical objectives. This study seeks to answer the following research questions (RQs):

RQ1. How has the volume of publications on MAS in education evolved over the past decade?

RQ2. How are MAS applications distributed across different educational levels?

RQ3. What are the primary educational goals addressed in MAS research?

2. Methodology

In the searching and screening phase, this study employed a systematic literature review method following PRISMA guidelines (Page et al., 2021). A comprehensive search was conducted on December 3, 2025, across four major academic databases (i.e., Web of Science, Scopus, IEEE Xplore, and ACM Digital Library) restricted to publications from 2015 to 2025. Rounds of iteration resulted the search string: *("multi-agent*" OR "multi agent*" OR multiagent* OR "multi-agent system*" OR "multi agent system*" OR "multiagent system*" OR "multi-agent systems" OR "multi agent systems" OR "multiagent systems") AND (education OR student* OR teacher* OR teaching OR pedagogy OR pedagogical OR instruction).* The initial search yielded 7,626 records. After removing 1,423 duplicates, 79 non-English, 459 non-peer-reviewed (e.g., book chapters), 45 non-retrievable, 134 short papers (<= 5 pages), and 4,963 irrelevant abstracts, the remaining 523 abstract texts. This abstract was conducted to serve as the first screening round and to inform the subsequent full-text analysis.

In the coding and rating phase, to ensure data validity, the selected article abstracts were evenly assigned to two independent researchers for coding using the initial analytical frameworks of educational levels and research objectives. An inductive approach was subsequently adopted to capture publications on MAS in education that lacked a specific target learner group or learning goal. The initial inter-rater agreement was calculated at 73%. All discrepancies were subsequently resolved through a consensus discussion involving a third senior researcher, ultimately achieving agreement on the current classification.

3. Key Findings and Discussion

3.1. Potential LLMs Surge in MAS Research

Research temporal trends were analyzed by publishing year (see Figure 1), revealing a sharp acceleration in educational MAS research. Publications increased from a stable baseline (2015–2023) to 50 in 2024 and reached 112 in 2025. Related research indicates that authors' motivations were influenced by the integration of LLMs (Malik et al., 2024). The increase in publications observed in 2025, rather than immediately after the release of foundational LLM models in 2023, may be partly explained by a standard academic publication lag in social science disciplines (Björk & Solomon, 2013). This surge is further interpreted through emerging agentic frameworks that increased functionality (Durante et al., 2024), accessible agentic applications that lowered technical barriers (e.g., Coze, Dify), and supportive policy and post-pandemic education digitalization that provided fertile ground (Holmes & Miao, 2023).

Figure 1

Annual distribution of publications from 2015 to 2025

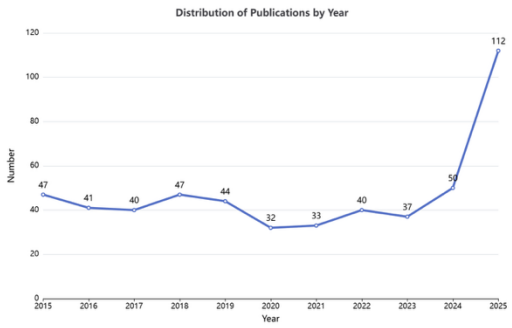
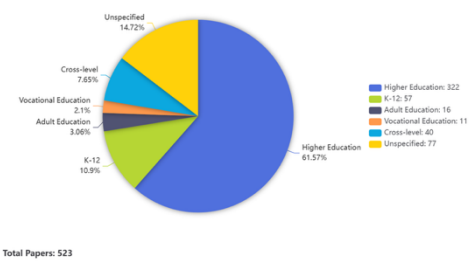


Figure 2

Distribution of MAS research by educational level



3.2. The Predominance of Higher Education in MAS Research

Among the 523 reviewed publications, higher education dominates (61.57%), followed by K-12 (10.9%), cross-level (7.65%), adult (3.06%), and vocational education (2.1%), with 14.72% unspecified (see Figure 2). The predominance is likely driven by readiness-optimized infrastructure and institutional capacity suitable for research in higher education (Pedró et al., 2019). Compared to higher education, the smaller K-12 portion might be due to the restrictive, protectionist safeguards typically required in K-12 settings (Akgun & Greenhow, 2022). Adoption in the adult and vocational sectors remains limited, underscoring the adaptable MAS's span across the educational spectrum. A notable portion of studies with unspecified levels is largely simulation-based or system-design research lacking direct human involvement, and these will be further analyzed to determine whether they should be excluded from the review.

3.3. Educational Objectives of MAS Research

The educational objectives emerged in seven categories (see Table 1): *Knowledge Mastery* for subject acquisition, such as STEM and language learning; *Cognitive and Metacognitive Skills* for critical thinking and self-regulation; *Ethical and Affective Development* for emotional growth; and *Social and Collaboration* for teamwork. The remaining categories include *Practical Skill Training*, *Learning Process Optimization* for personalization, and *Others* for technical studies that lack explicit human evaluation or pedagogical goals. The analysis reveals that Knowledge Mastery (n = 128) and Learning Process Optimization (n = 91) are the predominant educational goals, indicating that MAS is utilized primarily for subject-specific content delivery and adaptive tutoring, instead of fostering broader 21st-century competencies like critical thinking (n = 25) or collaboration (n = 18) (Miao & Shiohira, 2024). The marginal attention to Ethical and Affective Development (n = 6) highlights a critical gap in addressing contemporary priorities such as social-emotional learning (OECD, 2021). The high volume of technical studies (n = 231) in the current dataset needs to be further analyzed or removed to focus on the educational impact of MAS, rather than its technical architecture. This distribution suggests an imbalance in pedagogical objectives across MAS applications, underscoring more immersive, human-centered pedagogical goals (Zawacki-Richter et al., 2019).

Table 1

Pedagogical landscapes of MAS

Pedagogical Goals	Number	Pedagogical Goals	Number
Knowledge Mastery	128	Practical Skill Training	24
Cognitive and Metacognitive Skills	25	Learning Process Optimization	91
Ethical and Affective Development	6	Others	231
Social and Collaboration	18		

4. Limitation and Future Plan

The limitation in the current study lies mainly in the methodology. First, a preliminary, abstract-level coding may overlook complex pedagogical objectives or the specific technical nuances of MAS architectures, and may therefore not be precise. Second, a Cohen's Kappa coefficient or a percentage agreement of 80% should be established as the threshold for acceptable consistency, requiring a refined coding framework to ensure inter-rater reliability. Third, regarding RQ3, some studies exhibited potential pedagogical overlap and necessitated a multi-label coding approach or theoretical framework for MAS implementation that serves varied instructional goals. Future phases of this review will include an updated criterion and a full-text synthesis to elucidate MAS innovations and pedagogical functions in education, focusing on the evolution enabled by LLMs across system architecture, pedagogical configuration, and empirical impacts on learning outcomes.

References

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and ethics, 2*(3), 431-440.

Björk, B. C., & Solomon, D. (2013). The publishing delay in scholarly peer-reviewed journals. *Journal of informetrics, 7*(4), 914-923.

Durante, Z., Huang, Q., Wake, N., Gong, R., Park, J. S., Sarkar, B., ... & Gao, J. (2024). Agent AI: Surveying the horizons of multimodal interaction. *arXiv preprint arXiv:2401.03568.*

Hamal, O., Faddouli, E., & Harouni, M. H. A. (2021). Design and implementation of the multi-agent system in education. *World journal on educational technology: Current issues, 13*(4), 775-793.

Holmes, W., & Miao, F. (2023). *Guidance for generative AI in education and research.* UNESCO Publishing.

Jiang, Y. H., Liu, T. Y., Zhuang, X., Hu, H., Li, R., & Jia, R. (2024). Enhancing educational practices with multi-agent systems: A review. *Enhancing educational practices: Strategies for assessing and improving learning outcomes,* 47-65.

Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International journal of artificial intelligence, 26*(1), 160-169.

Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. *Vicinagearth, 1*(1), 9.

Malik, M. A., Amjad, A. I., Aslam, S., & Fakhrou, A. (2024). Global insights: ChatGPT's influence on academic and research writing, creativity, and plagiarism policies. *Frontiers in research metrics and analytics, 9*, 1486832.

Martha, A. S. D., & Santoso, H. B. (2019). The design and impact of the pedagogical agent: A systematic literature review. *Journal of educators online,* 16(1), n1.

Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO Publishing.

OECD (2021), *Beyond academic learning: First results from the survey of social and emotional skills.* OECD Publishing.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic reviews, 10*(1), 89.

Pedró, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development,* UNESCO Publishing.

Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, 104564.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators?. *International journal of educational technology in higher education, 16*(1), 1-27.

Zhang, S., Jaldi, C. D., Schroeder, N. L., López, A. A., Gladstone, J. R., & Heidig, S. (2024). Pedagogical agent design for K-12 education: A systematic review. *Computers & Education, 223*, 105165.

Zhao, Z. (2025). A new paradigm of personalized education driven by multi-agent collaboration. *Academic journal of sociology and management, 3*(3), 9-17.

科学教育情境中好奇心向探索性学习行为转化的神经机制研究

Research on the Neural Mechanisms Underlying the Transformation of Curiosity into Exploratory Learning Behaviors in Science Education Contexts

郑盈盈[1]*，董伟[2]
1,2 天津大学教育学院
*zyy_25@tju.edu.cn

【摘要】本研究基于预测误差-评估-好奇-探索的神经认知框架，探讨科学教育情境下中学生面对知识缺口时的好奇心评估过程及其对探索行为的影响。研究旨在回答：(1)面对科学知识缺口时，学习者不确定性感知与好奇心强度的变化规律；(2)好奇心强度在不同时间成本条件下是否及如何促发展探索性学习决策；(3)前额叶皮层神经活动是否在好奇心与探索决策间起中介作用。本研究在理论上深化对好奇心与学习机制的理解，实践上为科学教育中探究情境的设计与评估提供可操作依据。
【关键词】好奇心；探索性学习行为；功能性近红外；科学教育

Abstract: *This study is grounded in the neurocognitive framework of prediction error, appraisal, curiosity, exploration(PACE), investigating how middle school students evaluate scientific knowledge gaps in science education contexts and how this evaluative process influences subsequent exploratory learning behaviors. Specifically, the study addresses three research questions: (1)How do learners' uncertainty perception and curiosity intensity change when encountering scientific knowledge gaps? (2) Whether and under what time cost conditions does curiosity intensity trigger subsequent exploratory learning decisions? (3) Does prefrontal cortical activity mediate the relationship between curiosity and exploratory learning decisions as a value-based evaluation mechanism? Theoretically, this research deepens the understanding of the mechanisms underlying curiosity and learning. Practically, it provides actionable insights for the design and assessment of inquiry-based learning environments in science education.*
Keywords: curiosity; exploratory learning behaviors; fNIRS; science education

1. 引言

科学教学应引导学生投入到知识建构和评估过程中，推动科学学习转向在科学实践中学习(Berland et al., 2016)。学生主动参与科学知识建构有助于激发和维持学生的好奇心(Ali et al., 2025)；科学探究心理意愿对其科学素养发展具有重要作用(Chi et al., 2025)。然而，现有研究更多关注知识掌握、概念理解和知识整合等(Kidron, 2025; Zhou & Moon, 2025)，较少关注学习者面临知识缺口时是否及如何将对知识的好奇心转化为实际的探索性学习行为。好奇心被广泛视为驱动学习者探索性行为的内在动力(Grossnickle, 2016)。大量研究表明，好奇心能够促进信息寻求(Korniluk et al., 2025)、增强学习投入(Van Schijndel & Jansen, 2025)，对知识记忆与理解产生积极影响(Fandakova & Gruber, 2021)。然而，在真实学习情境中，好奇心并不一定转化为探索性行为(Gruber & Ranganath, 2019) 。当学习者在面对知识缺口时，往往同时感知到不确定性、潜在失败风险或时间与精力成本，从而选择行为回避并非主动探索(Lau et al., 2020)。这说明仅用信息寻求或者学习结果作为学习指标，难以判断实际学习是否可能发生。理解好奇心产生后的评估与决策机制对于理解人类学习和探索至关重要(Poli et al. , 2024)。

认知性好奇心是获得知识与信息的需求和渴望(Grossnickle, 2016)，是探究行为发生的关键。现有关于认知好奇心的研究主要采用冷知识问题范式(Kang et al., 2008)，通过操控信息缺口诱发认知性好奇心，并考察其对学习和记忆的影响。该范式为揭示好奇心的动机价值与神经基础提供了重要证据，但其局限在于：一方面，冷知识范式大多脱离具体学科情境，学习者难以判断某项活动对学习目标的有用性，与科学学习中真实的知识建构过程的相关性不大(Poli et al., 2024)；另一方面，相关研究往往以主观好奇评分或学习结果为指标，较少关注好奇心是否以及在何种条件下转化为实际的探索性学习行为。

基于此，本研究尝试在科学教育情境中引入冷知识研究范式，通过构建科学知识问题情境，诱发学习者对科学好奇心，并进一步分析好奇心评估过程如何影响后续探索性学习决策，从过程视角揭示好奇心由动机状态向探索性学习行为转化的关键机制，为更好地促进好奇心向探索性学习行为转化提供科学证据。

2. 理论背景与问题提出

2.1. PACE 理论框架

Gruber 和 Ranganath (2019)从神经认知层面提出了 PACE 框架，用来解释好奇心的产生机制及其促进学习与记忆的作用。该框架认为，好奇心的唤醒从预测误差开始，但预测误差本身并不会自动引发好奇心，需要进入评估阶段经过一系列信息的潜在价值及风险评估(Silvia, 2005)。大脑前额叶皮层的活动反映了学习者对信息价值与探索成本的整合计算(Gruber & Ranganath, 2019)。当个体认为信息具备价值且可获取时，会进入探索阶段，进一步寻求新的信息。在该框架中，评估是连接知识空白和知识探索的关键环节，但是该阶段在学习情境中的具体作用机制仍缺乏直接的实证检验。尤其是在真实的科学知识情境下，好奇心是否以及如何转化为实际的探索性学习行为尚不清楚。基于此，本研究以 PACE 理论框架为分析视角，聚焦评估和探索关键转化环节，旨在考察学习者在面对科学知识缺口时，好奇心评估过程如何影响其后续探索性学习决策。

2.2. 好奇心评估的神经机制

PACE 框架中指出，好奇心评估阶段，由前额叶皮层负责评估与海马体及前扣带回皮层相关的预测误差信息，并调控由预测误差诱发的好奇心或焦虑情绪的作用路径(Gruber & Ranganath, 2019)。当好奇心激活增强时，前额叶皮层的表现也会增强(Gruber et al., 2014)。具体来说，当面临是否探索信息或者选择放弃信息等需要更多决策评估时，外侧前额叶皮层(Lateral PFC)的神经活动会显著增强，LPFC 参与了探索行为的发起决策(Gruber & Fandakova, 2021; Gruber & Ranganath, 2019)；当面临加工不确定性和评估信息时，眶额叶皮层(orbitofrontal cortex, OFC)会显著激活(Korniluk et al., 2025) 。在学习情境中，相较于其他深部脑区，前额叶皮层更适合在自然任务条件下进行监测。因此，本研究假设前额叶皮层在好奇心驱动的探索决策中发挥关键作用，其神经活动可能反映学习者对信息价值与探索成本的整合评估过程。

2.3. 探索成本与探索行为决策

好奇心驱动的探索是指，探索行为出于求知的动机驱动，而非依赖外在奖赏(Poli et al., 2022)。该过程将知识获取作为一个内在奖励，通过奖励学习过程，强化人们信息寻求行为(Murayama, 2022)。已有研究发现，一方面，好奇心与探索欲与正相关，好奇心越高，学习者的探索意愿越强，越渴望深度的探索相关话题(Van Schijndel & Jansen, 2025)，但参与者的探索行为受到学习进展与感知新颖性的共同引导，会规避极端不确定的状态(Korniluk et al., 2025)。另一方面，好奇心激发后，个体会去探索、提问和操作有趣的物体(Kidd & Hayden, 2015)，个体甚至愿意付出一定时间或资源成本(Bennett et al., 2016; Lau et al., 2020)，但高成本可能抑制探索行为。这表明好奇心向探索行为转化过程中，探索成本会影响探索行为的决策。

由于时间成本具有较高的生态效度，能够真实反映课堂与课外学习中学习者在持续探究与放弃探索之间的权衡。因此，本研究引入时间成本作为情境变量，考察在不同成本条件下，好奇心评估是否以及如何转化为后续的探索性学习行为。

2.4. 问题提出

已有的研究表明认知好奇心在信息寻求和学习过程中发挥着重要作用，并初步探索了好奇心与探索行为的之间的关系。但是现有研究多停留在好奇心水平与学习结果之间的相关分析，较少从过程视角探讨好奇心评估阶段如何影响后续探索性学习决策，也缺乏对探索成本调节作用的直接检验。因此，本研究基于 PACE 理论框架，聚焦科学教育情境中，采用认知神经科学方法，监控学习者面临知识缺口时的好奇心评估过程，通过行为数据和前额叶皮层神经活动指标，系统性考察好奇心向探索性学习行为转化的过程机制。具体研究问题如下：

(1)学生在面对科学知识缺口时，其答案不确定性感知与好奇心强度呈现什么样的特征？

(2)学生在好奇心评估阶段所表现出的好奇心强度，是否以及在何种时间成本条件下，能够促发后续的探索性学习决策？

(3)评估阶段前额叶皮层的神经活动是否作为价值评估机制，在好奇心与探索性学习决策之间发挥中介作用？

3. 研究方法

3.1. 实验目的

本研究采用以探索成本为操作变量的两阶段实验设计。第一阶段通过冷知识范式诱发科学知识缺口，重点考察学习者对信息不确定性与主观信息价值的评估，以刻画科学好奇心主观评估过程；第二阶段则在好奇心已形成的基础上，引入探索成本，考察学习者是否将该动机状态转化为实际的探索性学习行为。

3.2. 实验设计

3.2.1. 实验任务与材料

研究首先编制科学好奇心材料库，以精准地操控好奇心诱发，并开发好奇心诱发-探索性行为两阶段实验任务范式。第一阶段：好奇心诱发阶段。基于经典冷知识研究范式题库(Gruber et al , 2014)对试题进行改编，构建符合科学教育情境的科学知识题库(如物理、工程、科技等相关基础科学问题)。每个试次依次呈现：(1)科学知识问题，被试需作答；(2)被试对当前问题引发的好奇心强度进行主观评分，题目：你对该问题的答案有多好奇？(1=非常不好奇，7=非常好奇)；(3)被试对答案不确定性感知进行主观评分，题目：你对该问题的答案有多确定？(1=非常不确定，7 非常确定)。第二阶段：探索性学习决策阶段。通过操控探索所需付出的时间成本水平(高/低)，构建不同的探索决策情境。每个试次依次呈现：(1)探索意愿选择，题目：你是否原因花费 30 分钟的时间对这些问题进一步探索？(是/否)，被试需思考后选择放弃或进入探索；(2)最小行为探索，若被试选择探索，则呈现一段 150-200 字结构化补充科学知识材料，被试可以选择随时结束。

3.2.1. 测量指标与工具

第一阶段测量指标：(1)好奇心强度；(2)答案不确定性感知；(3)反应时；神经指标：同步采集功能性近红外脑成像数据，重点关注前额叶皮层氧合血红蛋白(oxygenated hemoglobin, HbO)()与脱氧血红蛋白变化(deoxygenated hemoglobin, HbR)。第二阶段：(1)是否选择探索(探索/放弃)；(2)探索决策反应时；(3)实际探索持续时间。

3.3. 实验设备与生理数据采集

本实验使用的功能性近红外脑成像设备是深圳英智科技有限公司生产的(型号 NIRS Revo)。该设备通过发射特定波长的近红外光并测量其在脑组织中的吸收变化，反映大脑皮层

HbO 和 HbR 的浓度变化。其对近红外光(波长 785 和 830)的采样率为 12.5Hz。使用的是 4×4 的光极板，包含 8 个发射光源(sources)和 8 个探测器(detectors)，光极之间的间隔为 3 厘米，共构成 24 个通道。鉴于前额叶皮层被认为是好奇心评估与探索行为调节的关键脑区(Gruber & Fandakova, 2021; Gruber & Ranganath, 2019)，其功能与本研究理论框架高度契合，依据国际脑电图的 10-20 系统，以 Fpz 点为中线参考，确保光极板覆盖双侧前额叶皮层区域。通过虚拟定位的方法来确定记录通道与测量点在大脑皮层上对应的位置(Singh et al., 2005; Tsuzuki et al., 2007) 。

4. 研究阶段

本研究目前处于实验材料设计及伦理审查阶段，正在参照冷知识研究范式构建科学知识情境下的实验材料试题库，计划形成约 40 个试次，并对先验知识水平与探索成本进行系统操控。

5. 下一步研究工作

后续拟邀请科学教育领域专家、一线教师及学生对实验材料难度与情境适切性进行评估，并采用近红外设备进行预实验，评估刺激材料的数量和有效性、近红外数据的信号与质量。在完善实验设计后，开展正式实验，对实验数据进行采集与分析，并据此开展论文的撰写。

参考文献

Ali, F., Wang, Y., Wang, S. J.-W., & Zhu, G. (2025). Triggers of curiosity in social constructivist classroom discourse. *Npj Science of Learning*, *10*(1), 33. https://doi.org/10.1038/s41539-025-00330-5

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLOS Computational Biology*, *12*(7), e1005020. https://doi.org/10.1371/journal.pcbi.1005020

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082-1112. https://doi.org/10.1002/tea.21257

Chi, M., He, P., Li, G., & Zheng, C. (2025). Development and validation of an instrument to investigate secondary school students' desire to engage in scientific inquiry. *International Journal of Science Education*, 1-25. https://doi.org/10.1080/09500693.2025.2546460

Fandakova, Y., & Gruber, M. J. (2021). States of curiosity and interest enhance memory differently in adolescents and in children. *Developmental Science*, *24*(1), e13005. https://doi.org/10.1111/desc.13005

Grossnickle, E. M. (2016). Disentangling Curiosity: Dimensionality, Definitions, and Distinctions from Interest in Educational Contexts. *Educational Psychology Review*, *28*(1), 23-60. https://doi.org/10.1007/s10648-014-9294-y

Gruber, M. J., & Fandakova, Y. (2021). Curiosity in childhood and adolescence—What can we learn from the brain. *Current Opinion in Behavioral Sciences*, *39*, 178-184. https://doi.org/10.1016/j.cobeha.2021.03.031

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486-496. https://doi.org/10.1016/j.neuron.2014.08.060

学习科学视角下拔尖创新人才早期 MAPS 素养模型构建——基于深度访谈的实证研究

Construction of the Early MAPS Literacy Model for Top Innovative Talents from the Perspective of Learning Sciences: An Empirical Study Based on In-depth Interviews

王淑文[1]，刘誉[1]，侯兰[2]，尚俊杰[1*]

[1]北京大学教育学院学习科学实验室，北京 100871

[2]北京市海淀区教育科学研究院，北京 100029

[*] jjshang@pku.edu.cn；swwang25@stu.pku.edu.cn

【摘要】 拔尖创新人才早期识别是当前教育领域的重要议题。当前中国拔尖创新人才选拔仍依赖经验判断，缺乏科学、系统的素养模型。本研究立足学习科学视角，从认知与非认知维度梳理拔尖创新人才早期特征；以 20 名具有创新潜能的中学生为对象开展深度访谈，采用主题分析法构建 MAPS 早期素养模型，涵盖底层动力、学习素养、实践技能、社会影响与领导力四大维度及十大核心素养。研究结果可为拔尖创新人才早期识别、培养与评价提供理论指导与实践借鉴，推动创新人才由经验式选拔向素养导向转型。

【关键词】 学习科学；拔尖创新人才；主题分析；核心素养

Abstract: *Early identification of top innovative talents is a significant issue in education. Current talent selection in China still relies on empirical judgment and lacks a scientific and systematic literacy model. From the perspective of learning sciences, this study identifies early characteristics of top innovative talents from cognitive and non-cognitive dimensions; it conducts in-depth interviews with 20 middle school students with innovative potential and uses thematic analysis to construct the MAPS core literacy, which covers four dimensions and ten core literacy including motivational foundation, academic literacy, practical skills, social impact and leadership. The findings provide a theoretical and practical basis for the early identification, cultivation, and evaluation of top innovative talents and promote the transformation of selection from empirical to competency-oriented.*

Keywords: Learning sciences, Top innovative talents, Thematic analysis, Core literacy

1.引言

拔尖创新人才的早期识别与培养是回应“钱学森之问”、破解关键领域人才瓶颈的核心议题。当前我国拔尖创新人才识别仍依赖经验判断，缺乏科学系统的素养框架，且对早期成长规律的研究不够系统，难以实现精准识别与个性化培养（常宝宁，2025）。

学习科学整合认知、非认知与情境因素，为解构拔尖创新人才早期特征提供了理论支撑。建构主义、元认知理论、自我决定理论等学习科学理论，共同揭示了学习的主动性、反思性与内在驱动机制，与拔尖创新人才多维发展的规律高度契合。然而现有研究多聚焦单一维度

评估，较少将学习科学理论与拔尖创新人才早期素养体系深度结合，难以形成可操作的识别框架。

基于上述背景与现实困境，本研究立足学习科学视角，整合认知、非认知因素，提炼出拔尖创新人才早期特征；通过对 20 名具有拔尖创新潜能的中学生进行深度访谈与主题分析，构建 MAPS 早期核心素养模型，为早期识别与培养提供理论依据与实践参考。

2.理论基础及特征归纳

拔尖创新人才是在特定领域具备突出表现与发展潜能、可通过系统培养服务于国家战略的重要人才群体（白学军等，2023）。学习科学作为整合心理学、教育学与认知科学的跨学科领域，强调从认知、非认知与情境互动的整体视角理解学习规律，为系统解构拔尖创新人才早期发展特征提供了科学分析框架（尚俊杰，2023）。

本研究以学习科学核心理论为支撑，从认知与非认知两大维度梳理拔尖创新人才早期特征。在认知维度，依托建构主义、元认知理论（Flavell, 1976；Vygotsky, 1978），提炼出学习能力、结构化知识、元认知调控、高阶批判性思维与问题解决等关键特征。在非认知维度，基于自我决定理论、成长型心态、创造力投资理论与社会认知理论（Deci & Ryan, 2000；Hong et al., 1999；Sternberg & Lubart, 1991），归纳出内在动机、自我概念、心理韧性、创新人格、社会价值取向及人际协作等重要特质。

这些从学习科学视角凝练的拔尖创新人才早期特征，为后续构建 MAPS 核心素养模型奠定了扎实的理论基础。

3.研究设计

3.1.研究范式 本研究采用质性研究范式，以主题分析法为核心分析方法（Braun & Clarke, 2006）。该方法适合从访谈文本中归纳核心主题、提炼模型，契合本研究探索早期素养框架的目标。

3.2.研究对象 研究采用目的性抽样，在 B 市某区域经教师推荐、家长与学生知情同意后，选取 20 名具有拔尖创新潜能的青少年为研究对象。样本覆盖初中、高中及低年级大学生，其中男生 14 名、女生 6 名，涉及数理、科学、工程、AI、人文五大领域以保证研究的多样性与代表性。

3.3.数据收集与分析 通过半结构化深度访谈收集资料，单场访谈时长 60 至 90 分钟，全程录音并转录文本。采用 Braun & Clarke（2006）六步主题分析法进行编码，并通过多研究者编码、三角互证提升研究信度与效度，最终提炼核心素养维度。

4.MAPS 素养模型构建

前期基于学习科学视角从认知与非认知维度凝练的拔尖创新人才早期特征，为本研究核心素养提炼奠定了扎实基础。特征侧重外在表现描述，而素养更强调内在能力与品格的整合，因此，本研究进一步结合对早期拔尖创新人才的深度访谈与主题分析，完成从特征到素养的系统性转化。在主轴编码基础上，研究结合学习科学理论框架与研究目标，进行选择性编码，提炼核心素养。通过筛选具有统领性、代表性的次级主题，剔除重复或关联性较弱的内容，研究从底层动力（Motivational Foundation）、学习素养（Academic Literacy）、实践技能（Practical Skills）、社会影响与领导力（Societal Impact & Leadership）四个方面提炼出拔尖创新人才成

长与发展的十大核心素养，即志趣与意义认同、韧性与成长型心态、自我认知与自我实现、元认知、审辩式系统思维、自主学习能力、创造力、问题发现与解决能力、人际协作能力与领导力、社会价值与责任担当。

基于上述分析，研究构建了一个完整 MAPS 核心素养模型，如图 1 所示。底层动力是持续探索的心理基石，内在志趣与意义认同、韧性与成长型心态、清晰的自我认知与自我实现定位，共同构成长期深耕创新领域的核心动力，使个体能够抵御外部功利导向、在挫折中保持探索勇气（陆一和史静寰，2015）。学习素养是支撑高效学习与高阶思维的认知引擎，元认知调控、自主学习能力与审辩式系统思维协同发力，帮助个体规划学习过程、整合多元资源、系统性处理复杂信息，实现高质量知识建构（孙懿宏，2024）。实践技能是将内在认知转化为创新成果的关键载体。拔尖创新人才凭借敏锐的问题发现能力与突出创造力，在真实实践中把创意转化为可行方案与具体成果，完成从认知到实践的落地转化（程黎等，2023）。社会影响与领导力则实现创新价值的升华，人际协作与领导力推动创新从个体创造走向集体共创，而强烈的社会责任感与使命担当，让个人探索与社会需求、国家发展相结合，实现个人价值与社会价值的统一（周丹等，2025）。这四大维度相互支撑、层层递进，形成闭环式素养结构，为拔尖创新人才早期识别与培养提供了系统框架。



图 1 拔尖创新人才 MAPS 早期素养模型

5.结论与启示

本研究基于学习科学理论与访谈实证，提炼出包含底层动力、学习素养、实践技能、社会影响与领导力四大层面的拔尖创新人才应当具备的 MAPS 核心素养：志趣与意义认同、韧性与成长型心态、自我认知与自我实现、元认知、审辩式系统思维、自主学习能力、创造力、问题发现与解决能力、人际协作能力与领导力以及社会价值与责任担当。

研究表明，早期拔尖创新人才的发展是动机、认知、实践与社会价值协同演进的过程，这为超越传统成绩导向的选拔模式、构建素养本位的识别体系提供了依据。在实践层面，教

育者可依托本模型，从兴趣激发、元认知训练、真实问题实践、社会价值引领等方面系统化培养早期拔尖创新人才，推动拔尖创新人才的识别从经验判断走向科学、可操作的素养评价。

6.研究局限

本研究仍存在一些局限。第一，研究采用教师推荐的目的性抽样，未设置对照组，难以严格区分拔尖创新人才与普通优秀学生的素养差异，一定程度上影响结论的区分效度。第二，样本存在性别比例失衡，年龄跨度较大，可能带来抽样偏差，未能充分体现不同性别、不同发展阶段的创新人才发展路径差异。第三，本研究为探索性定性研究，样本范围与地域有限，结论的外部效度与普适性有待检验。未来可扩大样本规模、优化抽样结构，并结合量化方法开展验证研究，以提升模型的科学性与推广价值。

参考文献

白学军,刘颖,卢柳柳,吕佳,刘羽,郭秀丽... 和王永丽.(2023).拔尖创新人才早期培养的国际经验.**基础教育参考**,(04),3-21.

常宝宁.(2025).我国拔尖创新人才早期甄别的经验、问题与建议.**湖南师范大学教育科学学报**,24(01),103-109.https://doi.org/10.19503/j.cnki.1671-6124.2025.01.013.

程黎,陈啸宇,刘玉娟,等.(2023).我国拔尖创新人才成长模型的建构.**中国远程教育**,43(12):10-20. https://doi.org/10.13541/j.cnki.chinade.20231106.002

陆一,史静寰.(2015).拔尖创新人才培养中影响学术志趣的教育因素探析——以清华大学生命科学专业本科生为例.**教育研究**, 36(05):38-47.

尚俊杰.(2023).**学习科学导论**.北京大学出版社.

孙懿宏.(2024).我国英才教育研究综述及未来展望.**上海教育科研**,(10),8-19.https://doi.org/10.16194/j.cnki.31-1059/g4.2024.10.004.

周丹,赵子俊,龚超,等.(2025).人工智能拔尖创新人才的成长路径与影响因素——基于22位中小学生的访谈.**教育科学研究**,(06):90-96.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of goal pursuits: human needs and the Self-Determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/s15327965pli1104_01

Flavell, J. H. (1976). *Metacognitive aspects of problem solving.* In L. B. Resnick (Ed.), The Nature of Intelligence. Lawrence Erlbaum.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. J*ournal of Personality and Social Psychology*, *77*(3), 588–599. https://doi.org/10.1037/0022-3514.77.3.588

Sternberg, R. J., & Lubart, T. I. (1991). An investment theory of creativity and its development. *Human Development*, *34*(1), 1–31. https://doi.org/10.1159/000277029

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

基于多智能体辩论的多视角反思支架对职前教师教学设计能力的影响研究

Effects of Multi-Perspective Reflection Scaffolds Based on Multi-Agent Debate on Preservice Teachers' Instructional Design Competence

王依琳[1,2]，朱佳[1,2*]，
[1]浙江全省智能教育技术与应用重点实验室
[2]浙江师范大学教育学院
[*] jiazhu@zjnu.edu.cn

【摘要】 教学设计能力是教师专业发展的核心素养，而职前教师在成长过程中常面临理论与实践脱节的问题，且缺乏开展深度教学反思的有效支撑。针对这一现实困境，研究开发了基于多智能体辩论模式的多视角反思支架系统。该系统设置主持人智能体、学习者中心智能体和内容中心智能体，借助多理论视角的辩论式分析，助力职前教师审视并优化教学设计方案。研究拟采用三组准实验设计，从教学设计知识掌握、教学自我效能感、教案设计质量、反思行为深度等方面，评估该系统对职前教师的影响。目前，已完成系统的前期开发和小范围的测试，初步结果表明多视角辩论有助于教师加深对教学设计的认识和反思，提升其反思能力与实践水平。

【关键词】 多智能体；教学设计能力；反思支架；职前教师

Abstract: *Instructional design competence is central to teacher professional development, yet pre-service teachers often face theory-practice gaps and insufficient reflection support. This study developed a multi-agent debate system comprising a moderator, learner-centered, and content-centered agent to help teachers examine instructional design from multiple theoretical perspectives. A three-group quasi-experimental design is planned to evaluate the system's effects on instructional design knowledge, self-efficacy, lesson plan quality and reflection depth. Initial system development and small-scale testing have been completed, with early results showing that multi-perspective debate effectively bridges theory and practice, deepening teachers' understanding and reflection on teaching design.*

Keywords: multi-agent systems, instructional design competence, reflection scaffolds, pre-service teacher

基金项目：本文系浙江省尖兵领雁项目“跨领域多模态数据融合驱动下的多智能体智能决策服务研发及示范应用”（2026C02A1236）、国家自然科学基金面上项目“知识融合增强下基于多元教学事理图谱的生成式教学路径规划及应用研究”（62577050）的研究成果。

2. 前言

教学设计作为连接教学目标与教学实践的核心环节，直接决定教学质量与学生学习效果，其科学性与创新性是教师专业能力的集中体现（丁安琪等，2025）。然而，教学设计要求教师在复杂情境中权衡多种教学理念，平衡不同教学目标，做出专业的决策。研究表明，专家教师与新手教师的关键差异不仅在于教学技能的熟练度，更在于其能否深入理解不同教学选择背后的教育学原理，并根据具体情境做出适当的判断（Berliner, 2001）。这种专业判断能力需要教师持续进行批判性反思才能逐渐形成。然而，当前教师在教学设计中获得的支持显然不足，师范生在面对学校的具体状况时，往往产生理论与实践的隔阂（杨燕燕，2012）；在职教师则因繁重的教学压力和琐碎的管理工作而忽略了前提性反思（曹永国，2014），导致教学设计停留在表面或经验层面，难以实现真正的专业成长。

智能技术工具正逐步变革传统教学设计模式，这虽有效提升了教学效率，却也可能导致教师过度依赖智能生成内容，弱化其独立思考与批判性反思能力。为此，本研究开发了基于多智能体辩论模式的多视角反思支架，通过呈现内容结构与学习活动两个互补视角之间的对话与碰撞，激发教师的批判性思维，支持其深入开展前提性反思。目前，系统已完成构建与小范围试测，初步反馈显示其有效推动了理论与实践的结合，研究团队将据此完善系统设计并制定了完整的实证研究方案。

3. 相关工作

2.1. 教学设计中的人工智能

人工智能应作为教师的协作伙伴，帮助其在复杂情境中做出更科学的教学决策并推动自身的发展。然而，当前人工智能在教学设计中的应用大多聚焦于效率提升，如快速生成结构完整的教案（周洪宇等，2025）。若人工智能仅呈现输出结果，却未展现背后的推理过程，教师将难以理解其依托的教育学原理，甚至可能从主动的课程设计者沦为被动的系统使用者（Blömeke et al., 2022）。由此可见，人工智能在教学中的应用应更多支持教师的思考与判断，而非替代决策，从而帮助教师在实践中不断反思与成长。

2.2. 教育中的多智能体系统

已有研究表明，多个智能体围绕同一问题展开辩论，能显著提升推理质量和事实准确性（Du 等，2023）。这一辩论机制可推动智能体反复审视自身论点、识别逻辑漏洞，进而生成优于单一智能体的输出结果。然而，现有研究大多聚焦于提升智能系统自身能力，鲜少探讨将多智能体辩论机制应用于教师专业发展场景。从理论机制看，当教师同时接触来自内容结构与学习活动两个维度的分析时，两种视角间的差异形成认知冲突，打破原有认知平衡，驱动教师主动审视自身教学假设，进而激活元认知监控，推动其从描述性反思走向批判性反思。

4. 系统设计

本研究自主开发了多视角反思支架平台，包含三个智能体：主持人智能体负责协调对话全程，学习者中心智能体与内容中心智能体则分别代表课程论中两种经典立场，通过模拟多视角辩论引导教师深化对教学设计的专业理解。

3.1. 智能体角色设计

主持人智能体扮演协调者角色：分析教案并提炼核心议题，引导两个智能体展开平衡而有深度的对话，最终整合双方观点以启发教师反思。

学习者中心智能体围绕学生立场审视教案，关注学生已有经验与认知发展水平、自主探究空间的充分性以及学习情境的真实性，并据此提出促进学生深度参与的改进建议。

内容中心智能体忠实于学科知识体系的内在逻辑，审查教学内容的组织是否遵循学科顺序、知识结构是否严谨，并据此提出增强学科系统性的优化建议。

3.2. 辩论流程设计

系统采用多轮辩论机制，每轮辩论分为四个阶段：①问题聚焦：主持人提炼教案核心议题；②观点陈述：两个智能体分别从各自立场表态；③交互辩论：双方相互回应与反驳；④综合引导：主持人整合观点，促进教师全面反思。图 1 呈现了前两个辩论阶段的示例。

3.3. 提示词设计策略

系统采用结构化的提示词设计策略，分为五个层次：基础设定层定义智能体角色与风格、理论框架层嵌入核心理论原则、分析维度层明确关注的教学要素，对话策略层指导建设性辩论方式、情境适应层根据辩论阶段动态调整，确保各智能体在保持立场的同时生成有深度的对话内容。

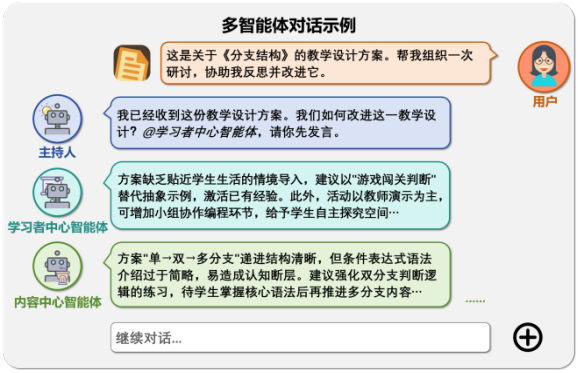


图 1 多智能体对话示例

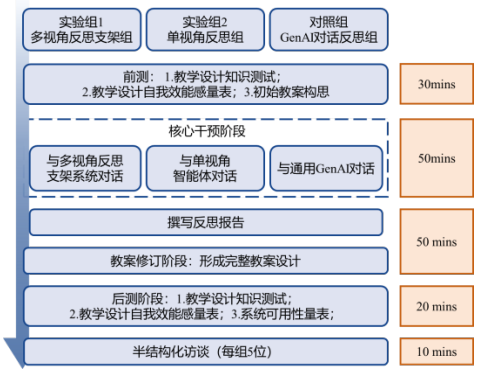


图 2 实验流程图

5. 研究方法与设计

基于前期系统开发与小规模试测（详见第 5 节），本研究计划开展准实验，系统评估多视角反思支架对职前教师教学设计能力的影响。研究核心假设为多视角辩论机制具有独立效应，优于单一视角对话。

4.1. 研究对象

本研究将以华东地区某所师范院校教育技术学专业和现代教育技术专业在读研究生作为研究对象，计划招募 75 名参与者，随机平均分配到多视角反思支架组、单视角反思组和 GenAI 对话反思组这三个组别，确保各组在性别和专业方向上的基本平衡。

4.2. 研究流程

研究将采用单次实验设计，在约 3 小时内完成前测、干预和后测的完整流程，如图 2 所示。前测阶段，所有参与者签署知情同意书后，完成前测量表并就针对算法的基本概念与应用这一高中信息技术课程主题提交初始教案构思。前测结束后随机分组，各组接受为期 10 分钟的系统培训或任务说明。干预阶段，多视角反思支架组将教案提交至本研究开发的系统后，观看学习者中心智能体与内容中心智能体的 3-4 轮辩论；单视角反思组提交教案后，随机与仅持一种立场的智能体对话获取分析建议；通用 GenAI 组提交教案后与通用大语言模型对话获取改进建议。干预结束后，所有参与者有 50 分钟时间修订教案与撰写反思报告。后测阶段：所有参与者完成教学设计知识测试、教学设计自我效能感量表及系统可用性量表，并对随机抽取的部分参与者开展半结构化访谈。

4.3. 研究工具

本研究综合运用定量测量工具和定性数据收集方法。教学设计知识测试改编自 Koh 等（2013）的量表，包含 15 道选择题，涵盖学习理论基础、教学策略知识、教学设计原则和学习者分析四个维度。教学设计自我效能感量表改编自 Holden 等（2011）的工具，包含 12 个项目，采用 7 点 Likert 量表。教案质量评分采用研究团队开发的评分量表，从教学目标清晰度、教学活动适切性、教学策略多样性、学生中心性和评估设计五个维度评分，每个维度 0-4 分，总分 20 分，由两名评分者独立评分并计算 ICC。反思深度编码框架（RDCS）改编自 Hatton 等（1995）的模型，将反思分为描述性反思（1 分）、比较性反思（2 分）和批判性反思（3 分）三个层次。系统可用性量表（SUS）采用 Brooke（1996）的标准工具评估系统接受度。半结构化访谈探索参与者的反思体验和教学设计思维变化。

4.4. 数据分析计划

定量分析采用单因素协方差分析，以前测得分为协变量，比较三组在教学设计知识增益、自我效能感增益、教案质量和反思深度等因变量上的差异。事后多重比较采用 Bonferroni 校正，并计算效应量（偏 η^2 或 Cohen's d）。质性分析方面，反思日志使用 RDCS 框架进行演绎

编码，访谈数据采用主题分析法。最后采用三角互证策略整合定量和定性发现，增强研究结论的可信度。

6. 初步试测与下一步研究计划

本研究于 2025 年 12 月邀请 3 位职前教师进行试用，参与者均为教育技术学专业研二学生，已修读教学设计相关课程但无实际教学经验。试用反馈显示：（1）系统通过多视角辩论帮助参与者意识到教学设计的多种可能性，有效促进了理论与实践的连接；（2）部分内容较为学术化，辩论轮次较多时存在认知负荷；（3）建议增加更具操作性的改进方案。

后续研究将围绕以下三方面重点推进：首先，结合前期试用反馈优化系统设计，对智能体的语言表达进行调整，完善辩论流程设计，并为教师补充更具实操性的教学改进建议。其次，遵循本研究第四章的研究设计，招募 75 名职前教师分为三组开展准实验研究，评估系统对教师教学设计能力与教学反思水平的实际影响。最后，在验证系统有效性的基础上，进一步探索多智能体辩论机制在其他教师专业发展场景中的应用潜力，如课堂管理、学生评价等，进一步探讨如何通过认知冲突、观点融合等机制，推动教师开展深度专业反思，助力其在更广域的教学实践中实现专业能力的持续成长与提升。

参考文献

丁安琪,蒙小凤.(2025).从模仿到重构:AI 生成教案对国际中文教育新手教师教学设计能力的影响.贵州师范大学学报(社会科学版),(05),42-52.

杨燕燕.(2012).论教师职前实践教学的取向转换.教育研究,33(05),84-89.

周洪宇,余江涛.(2025).迈向通用人工智能时代的教师工作：挑战与出路.电化教育研究,46(07),5-12.

曹永国.(2014).在做什么,抑或知道在做什么——教师的前提性反思的危机与重建.华东师范大学学报(教育科学版),32(01),41-49.

Berliner, D. C. (2001). Learning about and learning from expert teachers. International journal of educational research, 35(5), 463-482.

Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. Learning and Instruction, 79, 101600.

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189(194), 4-7.

Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., & Mordatch, I. (2023). Improving factuality and reasoning in language models through multiagent debate. In Forty-first International Conference on Machine Learning.

Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. Teaching and Teacher Education, 11(1), 33-49.

Holden, H., & Rada, R. (2011). Understanding the influence of perceived usability and technology self-efficacy on teachers' technology acceptance. Journal of research on technology in education, 43(4), 343-367.

Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2013). Examining practicing teachers' perceptions of technological pedagogical content knowledge (TPACK) pathways: A structural equation modeling approach. Instructional Science, 41(4), 793-809.

人机协同视域下教师跨学科教学设计胜任力提升研究：模型构建与 GAI 平台干预

Empowering Teachers' Interdisciplinary Instructional Design Competency through Human-AI Collaboration: Model Construction and Generative AI Platform Intervention

陈莉莉[1,2] 董艳[3]
北京师范大学未来教育学院，珠海，519087；
深圳市宝安区燕山学校，深圳，518105
北京师范大学教育学部，北京，100875
lilychenferc@163.com

【摘要】 为响应新课标对跨学科主题学习的要求，破解一线教师因长期分科教学背景而面临的跨学科设计胜任力瓶颈，本研究提出了“人机协同”的赋能路径。研究旨在：（1）构建一个整合生成式人工智能（GAI）维度的教师跨学科教学设计胜任力模型（AI-ITPACK）；（2）以自主开发的 GAI 平台“跨科三问”为干预载体，通过准实验设计揭示其赋能教师胜任力提升的具体机制与效果；（3）最终形成一套可推广的实践模式。本研究将"赋能机制"操作化为认知卸载程度、人机交互质量与元认知激活水平三类可测中介变量，通过中介检验与调节分析，揭示人机协同赋能教师跨学科设计胜任力的具体路径。
【关键词】 AI-ITPACK；生成式人工智能；跨学科教学设计胜任力；人机协同；基于设计的研究

Abstract: *In response to the national Compulsory Education Curriculum Plan (2022 Edition)'s mandate for interdisciplinary thematic learning, and to address the critical bottleneck in interdisciplinary instructional design competency faced by frontline teachers whose professional backgrounds are predominantly rooted in single-discipline pedagogy, this study proposes a human-AI collaborative empowerment pathway. The research pursues three interconnected objectives: (1) to construct a teacher interdisciplinary instructional design competency model (AI-ITPACK) that integrates a Generative Artificial Intelligence (GAI) dimension; (2) to employ a self-developed GAI platform—the Three-Question Interdisciplinary (TQI) platform—as the intervention vehicle within a quasi-experimental design, thereby elucidating the specific mechanisms and effects through which it empowers teachers' competency development; and (3) to ultimately yield a replicable and scalable practice model. This study operationalizes the "empowerment mechanism" into three measurable mediating variables: the degree of cognitive offloading, the quality of human-computer interaction, and the level of metacognitive activation. Employing mediation and moderation analyses, this research aims to unveil the specific pathways through which human-computer collaboration empowers teachers' interdisciplinary design competence.*
Keywords: AI-ITPACK, Generative AI, Interdisciplinary Instructional Design Competency, Human-AI Collaboration, Design-Based Research

1.前言

《义务教育课程方案（2022 年版）》将跨学科主题学习设定为刚性要求，如何通过教学实践培养学生解决复杂问题的核心素养，已成为国家课程改革的战略焦点。然而，植根于长期分科教学的专业背景，一线教师在落实这一要求时普遍遭遇了显著的“跨学科教学设计胜任力”瓶颈（杜文彬，2023），具体表现为整合多学科知识的困难、创设真实探究情境的困难以及设计结构化学习活动的困难（詹泽慧等，2024）。这种“政策高要求”与“教师实践能力不足”之间的结构性矛盾，构成了当前深化素养导向课程改革所必须破解的核心实践难题。

生成式人工智能的迅猛发展，为应对上述挑战、赋能教师专业发展提供了新的技术路径。已有研究初步揭示了 GAI 在辅助教学资源生成、启发教学设计思路方面的潜力（董艳等，2023；卢宇等，2023）。然而，现有探索存在两个明显局限：其一，多数研究依赖于通用大型语言模型，其生成内容往往难以精准契合新课标所要求的系统性、深度性教学设计逻辑，缺乏对

教育垂直领域的专门化适配（夏亮亮等，2025）；其二，相关研究多集中于验证 GAI 工具的“是否有效”，对于其“如何赋能”的内在机制——即人机协同过程中具体的认知交互与能力发展过程——缺乏深入的过程性剖析（穆肃等，2025）。这导致 GAI 赋能教师跨学科设计能力的“黑箱”仍未打开，也难以为开发精准有效的干预工具提供清晰的理论指引。

为弥补上述研究缺口，本研究遵循“理论建构-实证检验”的整合路径，探究“人机协同”赋能教师跨学科教学设计胜任力的具体模式。

2.研究方法

为系统性探究生成式人工智能（GAI）赋能中小学教师跨学科教学设计胜任力的内在机制与实践有效性，本研究以基于设计的研究作为统领的方法论范式。在本研究中，“设计”的核心对象是基于“跨科三问”GAI 平台的教师跨学科教学设计赋能路径即干预方案。整个研究遵循两轮完整的 DBR 循环（如图 1 所示），每一轮循环均整合多种具体研究方法，以协同解答三个研究问题。

研究核心包含两个关键环节：首先，致力于构建一个整合了 GAI 维度的教师跨学科教学设计胜任力新模型（AI-ITPACK）。不同于 TPACK 所强调的“技术—学科—教学法”三元整合，以及 ITPACK 在此基础上加入跨学科整合维度的扩展，AI-ITPACK 的核心创新在于：聚焦于描述教师在与 GAI 持续交互过程中，其跨学科设计能力如何经历“激活—重组—提升”的动态生成机制，GAI 作为认知协作伙伴，通过即时反馈、方案生成与概念整合，触发教师原有学科知识网络的跨界连接（激活）；通过提示词迭代与方案批判性修订，促使教师在元认知层面重构设计逻辑（重组）；最终在持续人机对话中将跨学科设计能力内化为可迁移的专业图式（提升）。这一动态机制是 AI-ITPACK 区别于既有模型的理论核心，也是本研究框架的立论基础。

其次，以自主研发、深度内嵌“追求理解的教学设计”（Understanding by Design， UbD）逻辑的教育垂直大模型平台——“跨科三问”为核心干预载体（夏亮亮等，2025）。研究采用**准实验设计**，并融合量化（如胜任力问卷、教学设计方案专家评分）与质性（如半结构化访谈、平台交互日志分析）的混合方法，系统检验该平台对教师胜任力的提升效果，并以**分布式认知理论**（Distributed Cognition）为核心解释框架，整合认知卸载与智能脚手架概念，深入揭示人机协同的作用机制。本研究的理论主轴确立如下：分布式认知理论主张认知过程并非局限于个体内部，而是分布于人、工具与环境构成的系统之中（Hutchins, 1995）。在此框架下，教师与"跨科三问"平台构成一个**人机认知系统**：GAI 承担外部认知功能，通过结构化提示与方案草稿生成实现**认知卸载**，从而为教师腾出有限的工作记忆资源，专注于高阶的跨学科整合判断；与此同时，平台内嵌的 UbD 逻辑以**智能脚手架**的形式动态支撑教师的设计思维，引导其从学科孤立走向目标统整。这一框架不仅作为事后解释工具，更直接指导干预方案的设计逻辑：平台任务结构的分层设计、提示顺序安排，均依据分布式认知系统中“卸载—协作—内化”的运作原理加以建构。

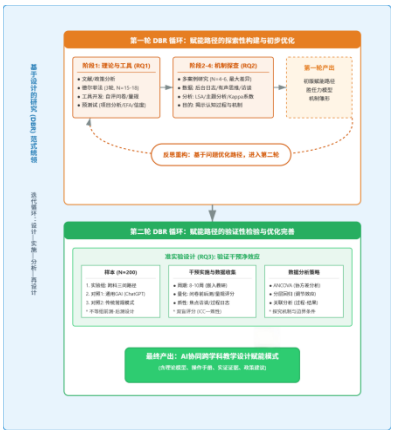


图 1 研究技术路线图

2.1. 第一轮 DBR 循环：赋能路径的探索性构建与初步优化

本轮 DBR 旨在通过小范围、深度的探索性研究，初步构建并优化赋能路径，为核心问题一和二提供初步解答，并为第二轮验证性研究奠定基础。

2.1.1 阶段一：理论模型构建与测评工具开发

首先，通过文献研究法与政策文本分析，初步提炼教师跨学科教学设计胜任力的关键要素。随后，采用德尔菲法构建并验证其结构化模型。

依据权威性与经验性原则，邀请 15-18 位专家参与咨询。专家组包括：1）教育技术或课程与教学论领域学者（4-5 位），需具备高级职称或博士学位，并持续关注技术赋能教学研究；2）中小学跨学科教学名师或资深教研员（6-8 位），需具有五年以上跨学科教学实践或指导经验，并曾主持/参与相关教研课题。实施三轮匿名咨询。首轮提供基于文献的初步框架，进行开放式修订；次轮使用李克特五点量表进行重要性评价；终轮反馈统计结果（均值、标准差、变异系数）以达成最终共识。采用肯德尔和谐系数（W）评估共识度，以 $W>0.5$ 且 $p<.05$ 为标准。

基于最终模型，编制《教师跨学科教学设计胜任力自评问卷》与《教学设计方案评价量规》。随后，进行大样本预测试以检验工具信效度。

2.1.2 阶段二至四：赋能机制微观探查与路径初版设计

在“跨科三问”平台初步应用场景中，深入理解人机协同的认知过程，为设计赋能路径提供依据。

研究采用多案例研究设计。依据“最大差异抽样”原则，选取 4-6 名参与教师作为案例，确保其在学科背景（文/理）、教龄（新手/熟手）、技术接受度前测（高/低）上具有异质性，以充分捕捉不同背景下的使用体验。

收集 1）过程性数据：通过平台后台自动记录教师的操作行为序列;2）认知过程数据：要求教师在进行设计任务时同步进行有声思维报告，并全程录音录像;3）回溯性理解数据：任务结束后，立即对教师进行半结构化访谈;4）负荷感知数据：在关键设计节点后，使用改编的《心智负荷量表》进行测量。

对有声思维转录文本与访谈文本进行主题分析。依据分布式认知理论构建初始编码框架，进行逐行编码并迭代修订。最后，将质性分析得出的关键主题（如“提示词迭代策略”）与后台日志中的具体行为序列进行时间线对齐与模式匹配分析，实现数据三角互证，从而揭示初步的赋能机制。

基于第一轮循环的发现，即初步的胜任力模型、微观机制洞察及教师反馈的问题，研究将设计出第一版较为系统的“跨科三问”赋能教师跨学科教学设计干预方案即赋能路径，进入第二轮循环。

2.2. 第二轮 DBR 循环：赋能路径的验证性检验与优化完善

本轮循环的核心目标是对优化后的赋能路径进行效果验证，并进一步收集数据以支持其持续完善。

研究设计采用不等组前测-后测准实验设计，旨在控制前测差异，检验干预的净效应。

参与者与抽样：通过合作区域教研网络，招募约 180-220 名中小学在职教师。依据其学科、教龄及前测胜任力得分进行匹配，将其分为三组。**个体特征变量的理论定位**：本研究将教师的教龄与学科背景纳入分析，其理论依据如下：在分布式认知框架下，人机认知系统的运作效能受到个体认知资源储备与知识结构的显著影响。具体而言：①**教龄**在本研究中定位为**调节变量**。依据专业发展阶段理论（Berliner, 1988），新手教师与熟手教师在已有学科图式的丰富程度上存在系统性差异：熟手教师具备更完善的学科知识网络，可能更有效地利用 GAI 输出进行批判性整合，从而在认知卸载的基础上产生更高质量的跨学科重构；而新手教师可能因缺乏足够的先备知识，对智能脚手架的依赖程度更高，从而表现出不同的能力提升轨迹。②**学科背景（文/理）同样定位为调节变量**。理工科教师与人文社科教师在知识结构的整合方式、对问题情境的建构逻辑上存在差异（Lattuca, 2001），这可能导致其与 GAI 协同进行跨学科整合时呈现不同的认知路径，进而调节干预的效果大小与方向。③**前测胜任力得分**作

为**背景控制变量**纳入协方差分析，旨在排除初始能力差异对结果的混淆效应，确保干预净效应估计的准确性。综上，个体特征变量在理论模型中的位置明确：教龄与学科背景为调节变量，前测得分为控制变量，三者在统计分析中承担不同的角色与功能。

干预实施："跨科三问"赋能路径的具体结构：本研究的核心干预载体为"跨科三问"平台，其内容逻辑与实施结构如下：（一）"三问"的内容逻辑直接对应 UbD 逆向设计的三阶段框架，构成跨学科教学设计的核心驱动线索：第一问"为何联结？"：引导教师识别两个及以上学科之间的本质性概念关联，锚定跨学科整合的真实价值点，对应 UbD 中"确定预期结果"阶段。教师在平台中通过与 GAI 的对话式交互，输入拟整合的学科与主题，GAI 依据大概念图谱生成候选联结方向，教师进行批判性筛选与确认。第二问"何以为证？"：引导教师设计能证明学生实现跨学科理解的表现性评估任务，对应 UbD 中"确定合适的评估证据"阶段。GAI 基于第一问确认的整合方向，提供评估任务脚手架，教师结合本班学情进行修订与情境化设计。第三问"如何实施？"：引导教师将前两问的成果转化为具体的学习活动序列，确保活动与目标、评估的逆向一致性。GAI 提供活动框架建议，教师进行任务分解与时间规划。（二）阶段安排与时间结构：干预历时 8-10 周，分为三个阶段：第 1-2 周为"定向与入门期"，包括平台功能培训、第一问的完整操练（以教研组为单位），重点建立人机协同的基本交互习惯；第 3-6 周为"深化设计期"，教师以小组合作形式依次完成三问的完整流程，形成一份跨学科教学设计方案初稿，每两周开展一次线上教研反思；第 7-10 周为"修订与内化期"，教师基于专家和同伴反馈对方案进行迭代修订，并完成个人独立设计任务以评估能力迁移。（三）教师在平台中的具体互动方式：教师通过自然语言对话框与 GAI 交互，平台提供三类结构化提示模板，并在每次对话后自动记录交互日志。平台还内置"设计决策追踪"功能，要求教师在每次修订后填写简短的决策理由，以促进元认知反思，同时为研究提供过程性数据。

数据收集：1）前、后测：使用《教师跨学科教学设计胜任力自评问卷》进行测量。2）过程性数据：收集实验组在平台上的所有设计过程日志。3）结果性数据：收集所有参与教师的最终教学设计方案，由 2-3 位不熟悉分组情况的外部专家，依据《教学设计方案评价量规》进行双盲评分，并计算评分者间一致性系数（如组内相关系数 ICC）。4）质性补充数据：在干预结束后，从每组中选取部分教师进行焦点小组访谈，以深入理解量化结果背后的原因。

数据分析计划：1）主效应检验：采用协方差分析，将前测得分作为协变量，检验三组在后测胜任力及教学设计方案得分上是否存在显著差异。2）中介机制检验：本研究将"赋能机制"操作化为可测量的中介过程变量，具体包括：①认知卸载程度（通过任务后心智负荷量表得分的降幅反映外在负荷的转移量）；②人机交互质量（通过平台日志中提示词迭代频次、修订深度等行为指标量化）；③元认知激活水平（通过有声思维编码中自我监控与反思性语句的频率表征）。干预变量通过上述中介变量（认知卸载程度、交互质量、元认知水平）作用于结果变量（跨学科教学设计胜任力得分与教学设计方案质量评分）。将构建包含自变量（干预类型：实验组/对照一/对照二）→中介变量（认知卸载程度/交互质量/元认知水平）→因变量（胜任力提升量/方案质量得分）的变量关系模型，并采用 Bootstrap 中介检验对上述路径进行统计验证。3）调节效应分析：采用分层回归分析，检验教师个体特征对上述中介路径的调节作用。4）过程行为与结果的关联分析：在中介模型框架内，将后台过程日志数据（提示词质量评分、方案修订频次、平台使用时长分布）纳入中介变量测量，与前后测差值进行路径分析，以识别有效人机协同的行为模式。

2.3. 研究质量保障与伦理

本研究中所有参与教师在研究前均充分知晓研究内容、过程、潜在风险与权利，并签署书面知情同意书，有权在研究中途无条件退出。所有收集的问卷、文本、行为数据均进行去标识化处理，仅用于本研究分析，并严格保密。通过采用上述混合方法、多轮迭代、三角互证、以及严格的信效度检验和编码者信度检验，共同保障本研究的科学严谨性。

3.现阶段研究进展与初步成果

本研究致力于探究生成式人工智能赋能教师跨学科教学设计胜任力的机制与路径。截至目前，研究工作严格遵循基于设计的研究范式，已完成第一轮完整的“理论与工具构建”周期，并于近期启动了第二轮“实证检验与迭代”工作。现阶段，研究的关注点已从纯粹的理论构想和平台设计，转入在真实教学情境中进行混合方法的实证验证与赋能路径优化。

3.1. 第一阶段：胜任力模型构建与测评工具开发

本阶段旨在解决研究的理论测量基础，回应“构建融合 AI 素养的教师跨学科教学设计胜任力的能力结构模型”（AI-ITPACK）的核心目标。

3.2. 测量工具库的初步开发

基于上述已确立的 AI-ITPACK 模型，研究目前正进入配套测量工具的开发与初步验证阶段。

正在进行伦理审查与小范围的焦点小组访谈，访谈对象主要为一线中小学教师和教育技术研究人员。访谈目的在于检视量表题项的表述是否符合一线语境，以及在模型理解上是否存在歧义，以确保量表具有良好的表面效度和内容效度。

3.3 下一步规划

计划在线发放便利样本进行探索性因子分析预调查（预计样本量≥150）。

3.4 总结

综上所述，本研究目前已完成了理论建构阶段的核心任务，即通过德尔菲法，构建并验证了具有高度内部一致性的 AI-ITPACK 跨学科教学设计胜任力三维结构模型。当前，研究正处于测量工具开发的初始阶段，初步结果已指向一个清晰、稳定且具备良好研究前景的理论框架，为后续基于 DBR 范式开展实证研究，即“跨科三问”平台的干预设计与机制探索。

参考文献

杜文彬. (2023). 教师跨学科教学能力的关键要素与结构模型建构研究——基于混合研究方法. 全球教育展望，52(8), 70–86.

董艳, 夏亮亮, 王良辉. (2023). 新课标背景下的跨学科学习：内涵、设置逻辑、实践原则与基础. 现代教育技术，33(2), 24–32.

教育部. (2022). 教育部关于印发义务教育课程方案和课程标准（2022 年版）的通知. 取自 https://www.gov.cn/zhengce/zhengceku/2022-04/21/content_5686535.htm

卢宇, 余京蕾, 陈鹏鹤, 等. (2023). 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例. 中国远程教育， (4), 24–31+51.

穆肃, 陈孝然, 周德青. (2025). 生成式人工智能赋能教学设计分析：需求、方法和发展. 开放教育研究，31(1), 61–72.

夏亮亮, 姚菁晨, 陈莉莉, 等. (2025). GenAI 赋能教师跨学科教学素养的培养——以“跨科三问”智能平台为例. 现代教育技术，35(7), 54–63.

詹泽慧, 吕思源, 周嘉慧, 等. (2024). 基于新课标的教师跨学科教学素养模型研究. 现代远距离教育， (6), 3–15.

致谢

本研究受全国教育科学规划课题“生成式人工智能‘跨科三问’赋能教师跨学科教学设计的研究”（EHA240435）支持。

高等教育环境下生成式人工智能何以影响学习结果？—基于 27 项实验与准实验的元分析

How Does Generative Artificial Intelligence Affect Learning Outcomes in Higher Education? A Meta-Analysis of 27 Experimental and Quasi-Experimental Studies

王佳伟* 董伟
天津大学教育学院
hbujiaweiwang@163.com

【摘要】针对学习者使用生成式人工智能（GenAI）可能引发的“认知卸载”风险，研究旨在厘清 GenAI 对高等教育学习结果的影响及边界条件。本研究采用元分析方法，纳入 27 项实验与准实验研究，检验保持、迁移与延迟测试结果及调节效应。GenAI 总体显著提升学习结果，但效果高度依赖条件；其显著促进保持与迁移的即时表现，而延迟效应不稳定。长期干预（>4 周）、脚手架引导与支架型交互能有效促进能力跨情境迁移与长效维持。未来应用需从“工具接入”转向“机制设计”，以规避“认知债务”风险。

【关键词】生成式人工智能；认知卸载；学习结果；元分析；教学支架

Abstract: *To address the risk of “cognitive offloading” that learners may face when using generative artificial intelligence (GenAI), this study aims to clarify the impact of GenAI on learning outcomes in higher education and its boundary conditions. A meta-analysis was conducted, incorporating 27 experimental and quasi-experimental studies to examine retention, transfer, and delayed test results, as well as moderating effects. The findings indicate that GenAI significantly improves learning outcomes overall, yet its effects are highly context-dependent. While it significantly enhances immediate performance in retention and transfer, its delayed effects are less stable. Long-term interventions (>4 weeks), scaffolding guidance, and scaffold-based interaction effectively promote cross-contextual transfer and long-term maintenance of competencies. Future applications should shift from mere “tool access” to systematic “mechanism design” in order to mitigate the risk of “cognitive debt”.*

Keywords: Generative Artificial Intelligence (GenAI); Cognitive Offloading; Learning Outcomes; Meta-analysis; Instructional Scaffolding

1.引言

随着 ChatGPT 为代表的生成式人工智能（Generative Artificial Intelligence，GenAI）深度嵌入高等教育，学习者的加工认知链条正在被深刻重塑（Zhang,2025）。相较于传统数字工具，GenAI 不仅提供信息检索，更通过生成解释、重组知识等方式直接介入推理与协作学习过程（Villar-Onrubia et al.,2025）。在此背景下，认知卸载（Cognitive offloading）风险日益凸显：虽然 GenAI 能释放认知资源并快速提升短期任务绩效，但也极易导致学习者将自我解释与监控等关键环节让渡给系统，进而形成“认知债务”（Kosmyna et al.,2025; Risko et al.,2016）。若教育价值评估仍主要依赖即时测验或课堂产出，短期表现与真实习得之间的界限将被严重弱化(Soderstrom et al.,2015)。

针对 GenAI 的学习效应，既有实证研究在结论上呈现显著分歧（Bastani et al.,2025; Kestin et al.,2025），这提示其效应具有高度的条件依赖性。尽管近期已有部分元分析探讨了 GenAI 对学生学习表现的总体影响（Wang et al.,2025），但这些研究多聚焦于即时或短期的产出指标，尚未在“保持—迁移—延迟”等更具深度的认知结果层级上进行细分检验。此外，现

有综合研究未能充分澄清工具介入的边界条件。例如，当交互导向“替代式”完成时，图式建构与策略内化随之受损；而当其作为强调反馈约束的“支架”时，则更可能促发生成性加工并转化为长期收益(Soderstrom et al.,2015;Deng et al.,2025)。

综上，亟需通过系统的量化综合估计，澄清 GenAI 学习效应的边界条件与风险区间，为“认知债务”的出现情境提供更具解释力的证据。本研究旨在回答以下核心问题：(1)GenAI 介入对高等教育学习结果的总体效应如何？其在保持、迁移与延迟测试三个结果层级上的效应大小与稳定性是否存在差异？（2）学习情境、材料特征与交互特征是否对效应异质性具有系统解释力？哪些条件会触发效应的增强、衰减或反转？

2.文献综述

随着 GenAI 深度嵌入高等教育，其功能已直接介入学习者的认知加工过程。当前学界围绕 GenAI 学习效应的机制争议，主要聚焦于其究竟是促进了“生成性加工”还是诱发了“认知卸载”（Bastani et al. 2025;Risko et al.,2016）。一方面，若交互能触发解释与反思，GenAI 便构成认知支架；另一方面，若关键推理长期由系统承担，则可能导致图式建构受损，使学习收益仅停留在即时表现，难以沉淀为跨情境的迁移能力或在延迟测验中维持（Kestin et al., 2025）。因此，教育价值的评估应超越单一的即时成绩，向保持、迁移与延迟测试等更深层的结果层级延伸。

尽管近期实证研究已初步揭示 GenAI 与学习表现的正向关联，但相关结论存在显著异质性（Wang et al., 2025）。更为关键的是，现有元分析在解释效应异质性时，多停留在宏观的学科或学习模式层面，未能深入探究交互限制性、干预时长及教学护栏等对“认知债务”风险的具体调节机制。

在机制层面的理论分歧之外，实证研究对 GenAI 学习效应的报告亦呈现出显著的异质性，提示其效应极可能受制于特定的边界条件。尽管多数研究观察到 GenAI 能正向提升即时学习表现或情感动机，但这些收益并未在所有情境中稳定出现（Wu et al., 2025）。现有证据表明，在缺乏教学护栏的通用交互中，学习者容易出现深层认知投入不足，进而削弱高阶思维及长时距评估中的表现（Bastani et al.,2025）；相反，遵循“必要难度”原则的支架式 AI 辅导则更有利于促进知识的深度内化（Nelson et al,2023）。此外，干预时长、任务复杂性及学习模式等方法学层面的差异，也导致了研究结论的进一步分化。这表明，要全面评估 GenAI 的教育价值，需要系统识别并检验影响其效应转化的调节因素。

因此，既有研究中的关键边界条件可系统归纳为三大调节维度。其一，学习环境特征。在真实教学中，教师的脚手架引导能促使学生将 AI 输出作为解释支架而非答案替代，从而彻底改变知识保持与迁移的轨迹。其二，材料特征。相较于易诱发直接采用路径的简单、良构任务，复杂、劣构任务更迫使学习者进行表征重构，进而触发跨情境的能力迁移。其三，交互特征。具备学习护栏的支架型交互能够维持必要的认知挑战，其对延迟保持稳定性的提升，显著优于无限制的替代型交互。综上所述，本研究将以保持成绩、迁移成绩与延迟测试成绩作为核心评估指标，并基于上述三大维度构建调节变量编码体系，旨在通过量化综合，深度澄清除即时表现外，GenAI 学习效应发生异质性的内在机制。

3.研究设计

3.1 数据来源

基于上述问题，本研究严格遵循 PRISMA 准则开展系统检索，时间范围限定为 2022 年 11 月至 2025 年 12 月。为兼顾学科覆盖与文献完整性，检索数据库包括 Web of Science、PubMed、ERIC 等引文与学科索引数据库，以及 ElsevierScienceDirect、Taylor & Francis、Sprin

ger Nature Link 等出版商数据库。检索利用关键词 Generative AI/GenAI/Large Language Model/LLM/ChatGPT/GPT-/Claude/Gemini/Chatbot/Conversational Agent/Copilot 和 Human-AI interaction/Human-AI collaboration/Practice mode/Pedagogical mode/Scaffold/Tutor/Learning companion/Teaching assistant 和 Learning outcome/Higher-order thinking/HOTS/Critical thinking/Creative thinking/Problem solving/Self-regulat/Skill acquisition 以及 Experiment/Quasi-experiment/Randomized controlled trial/RCT/Control group/Pre-test/Post-test/Comparison group 进行检索。此外，采用滚雪球法对纳入研究的参考文献进行回溯，以补充潜在遗漏的关键实证研究。

文献筛选流程见图 1。初检共获得 893 条记录。经题名、摘要与关键词初筛，剔除重复记录、非期刊学术论文及主题不相关研究 532 篇，纳入 232 篇进入下一轮。进一步按研究类型与设计标准筛除综述、非实证与非教育场景研究，以及不满足随机对照或准实验设计者 56 篇，余 188 篇。全文审查阶段依据元分析统计要求排除效应量不可提取、关键调节变量缺失或样本量差距过大者 132 篇，保留 56 篇。随后进行内容一致性核验，剔除不构成 GenAI 干预或与研究问题不匹配者，最终纳入 27 篇期刊论文。

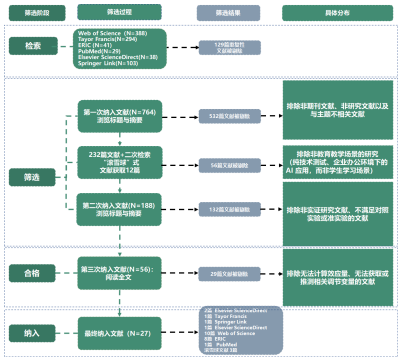


图 1 文献检索与纳入流程

3.2. 文献编码

元分析的文献编码维度包括环境特征、材料特征、交互特征及统计特征等四个核心维度。环境特征涵盖学习环境(实验室环境、在线环境、真实课堂)、干预时长（<1 周、1~4 周、>4 周）与引导方式（自主探索、教师引导/脚手架支持）；材料特征涵盖学科领域(自然学科和人文社科)、任务复杂性（简单/良构任务和复杂/劣构任务）和生成内容类型（文本生成、代码生成、多模态生成）；交互特征涵盖交互角色(助教、学伴)、交互限制性(替代型、支架型)和反馈精细度(结果反馈、揭示反馈)；统计特征涵盖文献作者及出版年份、样本量与实验对照组数据。为确保元分析编码的质量与信度，本研究采用双人独立编码。两名编码者先对随机抽取的 5 篇文献进行试编码，并据此协商完善编码框架，随后对其余文献开展正式编码。编码一致性以 Kappa 系数评估，K=0.88，表明一致性良好；分歧结果经复核与协商达成一致。

3.3. 研究方法

本研究采用元分析方法，使用 Comprehensive Meta Analysis 3.7 软件展开。首先，从纳入元分析的 27 篇文献中提取相关统计量。然后，计算合并效应值，根据合并效应值得出 GenAI 对高等教育学生学习成果的整体效应值，同时对纳入的文献进行发表偏倚检验。最后，进行调节效应检验，确定影响保持成绩、迁移成绩和延迟测试成绩的调节变量。本研究 27 篇样本文献中提取 105 项效应值。此外，为排除极端数值的影响，研究使用元分析中排除某一项研究的方法进行影响分析。经过检验，当前元分析中不存在异常值。

4.研究结果

4.1. 发表偏倚检验与异质性检验

发表偏倚是指统计结果显著的研究较不显著的研究更容易被发表的现象，这可能导致元分析夸大真实的效应估计。为了确保研究结论的稳健性，本研究采用漏斗图、Egger 线性回归检验以及失安全系数三种方法，对纳入效应量进行综合评估(Borenstein,2021)。通过多重证据交叉检验发现，元分析结果具有较高稳健性，不存在严重发表偏倚。尽管直观检视漏斗图在中小样本区域（漏斗中下部）存在微弱非对称性，但更为灵敏的 Egger 线性回归结果显示截距 t 值=0.520，p1=0.302,p2=0.604,t 值<1.96 且 p2>0.05，说明存在发表偏倚的可能性较小。

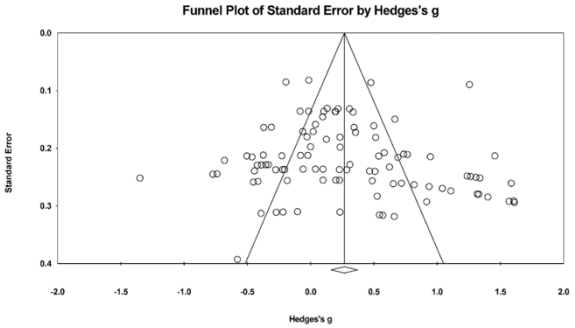


图 2 漏斗图

4.2. 总体效应异质性检验与分析

本研究整体效应检验结果表 1 所示所示。异质性检验结果显示，纳入的 105 个独立效应量之间存在高度显著的异质性（Q=795.735，p<0.001,I^2=86.930%）。这表明高达 86.93%的效应量变异源于研究间样本特征与干预方式的真实差异，而非单纯的抽样误差。鉴于此，本研究采用随机效应模型（Random-effects model）对合并效应量进行稳健估计。

由此可见，GenAI 对学习结果的整体促进作用虽具备统计学意义上的稳定性，但其实际提升幅度相对温和（g=0.267,p<0.001）。同时，极高的 I^2 值揭示了该平均效应背后存在显著的情境波动，表明 GenAI 的实际教学收益高度依赖于特定的边界条件。这一特征为后续开展调节效应分析、进一步识别诱发效应增强或衰减的核心教学要素提供了坚实的统计学基础。

表 1 整体效应检验结果

模型	k	g	95%CI		渐进		异质性检验			
			下限	上限	Z 值	P 值	Q	df	p	I^2
固定效应	105	0.245	0.209	0.282	12.074	0.000	795.735	104	0.000	86.930
随机效应	105	0.267	0.162	0.371	4.992	0.000				

为进一步检验 GenAI 介入效应在不同层级因变量上的差异，本研究分别对学习结果与过程性变量进行了子群分析（如表 2 所示）。总体而言，GenAI 对保持成绩、迁移成绩及部分过程性指标（如认知负荷）具有显著影响，但在延迟测验成绩、学习满意度与学习参与度上未观察到稳定效应。此外，多数结果变量的异质性检验显著且 I^2 值较高，表明效应量在不同研究间存在较大离散度，单一的合并效应不足以完全刻画跨研究的真实效应分布。

学习结果指标呈即时与长效的显著分化。在认知结果层面，GenAI 对保持成绩（g=0.338）与迁移成绩（g=0.381）均具有显著的正向促进作用，且子群内部异质性显著（保持成绩 Q(18)=172.994；迁移成绩 Q(37)=268.981）。相较之下，延迟测试成绩的合并效应未达显著水平（g=0.053），其置信区间跨越零点且异质性显著（Q(16)=65.270）。这充分说明，GenAI 介入带来的即时或跨情境学习增益，尚未在长时距的延迟测量中转化为稳定的总体优势。

过程性指标改善显著但高度依赖实施情境。在学习过程层面，GenAI 对认知负荷呈现显著的降低效应（g=-0.411），但组内异质性显著（Q(4)=16.511）。这种高异质性往往源于任务复杂度与支架设计的差异。当 GenAI 能够有效分解复杂概念时，可显著降低学习者的内在负荷；反之，若缺乏必要的教学护栏，学习者在甄别海量生成结果时极易增加无关的外部认

知负荷（Bastani et al.,2025）。同时，GenAI 显著提升了学习动机（g=0.347），但该指标的异质性同样处于高位（Q(20)=162.909）。这表明，基于技术新奇感与任务便利性带来的初期动机增益，在高强度评估或长期干预下能否转化为稳定的内在认知需求，深度受制于特定的研究情境。

至于学习参与度与满意度，鉴于独立研究数量过少且置信区间过宽，当前证据尚无法支持得出稳健的总体结论。综合上述结果可见，GenAI 对高等教育学习效果的影响呈现出显著的维度敏感性与时间尺度差异。即时性的保持与迁移指标呈显著正效应，而延迟测验未见稳定增益；过程性体验虽在负荷与动机上有所改善，但缺乏跨研究的普遍稳定性。此外，各子群极高的 I^2 值再次印证，这些效应分化绝非偶然抽样误差，而是由更深层的系统性边界条件所驱动。这为本文后续开展调节效应检验指明了方向，即需要进一步剖析何种学习情境、材料属性与交互特征能够真正促使 GenAI 的短期表现红利转化为长期的稳定习得。

表 2 因变量检验结果

结果变量	K	g	95%CI		渐进		异质性检验			
			下限	上限	Z 值	P 值	Q	df	P	I2
认知负荷	5	-0.411	-0.786	-0.035	-2.144	0.032	16.511	4	0.002	75.774
学习参与度	2	0.652	-0.682	1.986	0.958	0.652	17.070	1	0.000	94.142
学习动机	20	0.347	0.069	0.626	2.446	0.014	162.909	19	0.000	88.337
学习满意度	4	0.058	-0.218	0.333	0.409	0.682	8.148	3	0.043	63.180
保持成绩	19	0.338	0.098	0.579	2.763	0.006	172.994	18	0.000	89.595
迁移成绩	38	0.381	0.212	0.550	4.420	0.000	268.981	37	0.000	86.244
延迟测试成绩	17	0.053	-0.177	0.284	0.452	0.652	65.270	16	0.000	75.487

4.3. 不同调节变量对学生学习结果的影响

4.3.1. 能力迁移维度的调节效应分析：干预时长与支架引导的关键作用

能力迁移维度受干预时长与引导方式的显著调节。子群调节分析结果显示，GenAI 对迁移能力的促进作用深度受制于特定教学条件的系统性调节（如表 3 所示）。具体而言，干预时长表现出显著的组间差异（Q_B=13.742,p=0.001）。短期介入（少于 1 周）未能产生稳定的迁移增益，中期介入（1 至 4 周）效应虽转为正向但置信区间仍跨越零点（g=0.272,p=0.090），唯有长期介入（大于 4 周）才获得了显著且幅度可观的迁移增益（g=0.503,p<0.001）。这表明迁移能力的形成高度依赖持续性的练习反馈循环与跨任务应用机会，短期的工具接入不足以转化为稳定的跨情境应用能力。此外，引导方式对迁移效应亦具有高度显著的调节作用（Q_B=14.605,p<0.001）。在教师引导或脚手架支持条件下，迁移成绩呈现显著正效应（g=0.449），而在缺乏护栏的自主探索条件下，效应则转为负向且不显著（g=-0.171）。这一对比充分证实，迁移收益应在结构化的教学设计中才能被稳定激发。只有当学习者获得明确的任务要求、过程性提示与监控支持时，GenAI 的输出才更可能被转化为推理与整合的素材，进而服务于跨情境应用。值得注意的是，学习环境、学科领域及任务复杂度等其余变量未表现出显著的组间差异。鉴于部分子组纳入的研究数量偏少且统计功效受限，这种不显著不应直接等同于无调节作用，后续研究仍需通过扩大样本量或采用元回归方法对这些潜在边界进行精细检验。

表 3 GenAI 对迁移成绩的调节效应分析

一级维度	二级维度	调节变量分类	k	g	95%CI		渐进		组间效应值
					下限	上限	Z 值	P 值	Q
环境特征	学习环境	实验室环境	7	0.277	-0.118	0.673	1.375	0.169	$Q_{B=2.099}$ $P_{=0.350}$
		在线环境	3	0.135	-0.232	0.503	0.721	0.471	
		真实课堂	28	0.437	0.227	0.646	4.088	0.000	

表 3 GenAI 对迁移成绩的调节效应分析(续)

	干预时长	<1 周	5	-1.101	-0.327	0.125	-0.878	0.380	$Q_{B=13.742}$ $P=_{0.001}$
		1~4 周	6	0.272	-0.042	0.585	1.696	0.090	
		>4 周	27	0.503	0.274	0.731	4.312	0.000	
	引导方式	自主探索	4	-0.171	-0.432	0.091	-1.278	0.201	$Q_{B=14.605}$ $P=_{0.000}$
		教师引导/脚手架支持	34	0.449	0.269	0.629	4.879	0.000	
材料特征	学科领域	自然学科	12	0.288	0.022	0.001	1.965	0.049	$Q_{B=0.568}$ $P=_{0.451}$
		人文社科	26	0.425	0.011	0.217	4.015	0.000	
	任务复杂度	简单/良构任务	1	0.311	-0.137	0.759	1.360	0.174	$Q_{B=0.088}$ $P=_{0.767}$
		复杂/劣构任务	37	0.384	0.211	0.556	4.346	0.000	
	生成内容类型	文本生成	30	0.359	0.182	0.537	3.969	0.000	$Q_{B=0.225}$ $P=_{0.893}$
		代码生成	7	0.494	-0.126	1.114	1.561	0.118	
		多模态生成	1	0.311	-0.137	0.759	1.360	0.174	
交互特征	交互角色	助教	7	0.156	-0.278	0.590	0.705	0.481	$Q_{B=1.318}$ $P=_{0.251}$
		学伴	31	0.433	0.246	0.619	4.550	0.000	
	交互限制性	替代型	11	0.226	-0.049	0.501	1.610	0.107	$Q_{B=1.548}$ $P=_{0.213}$
		支架型	27	0.445	0.236	0.655	4.169	0.000	
	反馈精细度	结果反馈	2	0.014	-0.474	0.501	0.055	0.956	$Q_{B=2.163}$ $P=_{0.141}$
		揭示反馈	36	0.403	0.226	0.580	4.450	0.000	

4.3.2. 延迟测试维度的调节效应分析：引导方式与交互限制性作为关键条件

针对延迟测试维度（如表 4 所示），虽然整体合并效应未达统计学显著，但调节效应分析揭示了引导方式与交互限制性存在显著的组间差异（Q_B=9.871,p=0.002）。具体而言，在缺乏结构化支持或呈现完全替代型交互的条件下，延迟测试表现出显著的负向效应（g=-0.680）。而在具备支架属性或护栏引导的条件下，效应点估计虽转为正向但未达显著（g=0.102）。这表明 GenAI 介入带来的即时增益在长时距测量中存在极大的衰减风险，长效收益的沉淀与维持深度依赖于系统交互能否强制保留并激活学习者关键的深层加工与自我监控环节。

表 4 GenAI 对迁移成绩的调节效应分析

一级维度	二级维度	调节变量分类	k	g	95%CI		渐进		组间效应值
					下限	上限	Z 值	P 值	Q
环境特征	学习环境	实验室环境	2	0.161	-0.587	0.909	0.423	0.672	QB=5.048 P=0.080
		在线环境	1	0.654	0.141	1.167	2.500	0.012	
		真实课堂	14	-0.000	-0.252	0.252	-0.003	0.998	
	干预时长	<1 周	4	0.222	-0.265	0.708	0.893	0.372	QB=0.754 P=0.686
		1~4 周	4	0.064	-0.614	0.742	0.184	0.854	
		>4 周	9	-0.027	-0.312	0.258	-0.186	0.852	
	引导方式	自主探索	1	-0.680	-1.114	-0.247	-3.076	0.002	QB=9.871 P=0.002
		教师引导/脚手架支持	16	0.102	-0.122	0.325	0.890	0.373	
材料特征	学科领域	自然学科	—						QB=0.000 P=1.000
		人文社科	17	0.053	-0.177	0.284	0.452	0.652	
	任务复杂度	简单/良构任务	2	0.086	-1.008	1.180	0.154	0.878	QB=0.004 P=0.951
		复杂/劣构任务	15	0.051	-0.189	0.291	0.417	0.677	

表 4　GenAI 对迁移成绩的调节效应分析（续）

	生成内容类型	文本生成	16	0.017	-0.217	0.251	0.142	0.887	QB=4.911 P=0.027
		代码生成	—						
		多模态生成	1	0.654	0.141	1.167	2.500	0.012	
交互特征	交互角色	助教	6	0.140	-0.322	0.601	0.594	0.552	QB=0.261 P=0.609
		学伴	11	0.002	-0.254	0.258	0.017	0.986	
	交互限制性	替代型	1	-0.680	-1.114	-0.247	-3.076	0.002	QB=9.871 P=0.002
		支架型	16	0.102	-0.122	0.325	0.890	0.373	
	反馈精细度	结果反馈	—						QB=0.000 P=1.000
		揭示反馈	17	0.053	-0.177	0.284	0.452	0.652	

5.讨论与启示

5.1. 研究讨论

本研究的元分析结果表明，GenAI 对高等教育学习结果总体呈现显著但幅度温和的正向促进作用。然而，高度的研究间异质性揭示了该效应具有极强的条件依赖性。

在学习结果的维度分化上，即时增益尚未稳定转化为长效习得。认知结果方面，GenAI 在干预阶段能有效提升保持与迁移表现，但延迟测试的总体合并效应不显著且置信区间跨越零点。这一结果印证了学习科学的经典论断，即技术赋能下的高表现并不等同于真实的认知内化(Soderstrom et al.,2015;Deng et al.,2025)。当推理、整合与监控等关键加工过程被工具部分替代时，撤除支持后的延迟测量往往会暴露能力沉淀的不足。过程指标方面，GenAI 虽能显著降低认知负荷并提升学习动机，但学习参与度与满意度并未获得稳定支持，表明过程性体验的改善仍受制于任务压力与交互质量等复合因素。

在效应的边界条件上，长效收益深度依赖系统化的教学设计。调节分析证实，迁移成绩存在显著的“时长效应”趋势，短期介入难以形成稳定增益，而超过 4 周的长期干预则能通过持续的练习与反馈循环激发更为显著的跨情境迁移。同时，引导方式起到决定性调节作用。教师引导与脚手架支持条件下的迁移与延迟表现显著优于自主探索，且无限制的替代型使用在延迟测验中释放出明确的“认知卸载”风险信号。需要指出的是，部分变量未显现统计学差异主要受限于当前子群样本量，不应被过度推论为 GenAI 具备跨情境的绝对普适性。综上所述，GenAI 在高等教育中带来即时收益的前提，是需要将其置于充足的干预时长、结构化的支架引导以及严格的交互限制等特定教学设计架构之内。

5.2. 研究启示

为规避认知卸载风险并提升知识内化概率，GenAI 赋能教学应从单纯的“工具接入”转向系统化的“机制设计”，首要在于重构生成性交互机制与动态干预时序。依据“必要难度”理论(Nelson,2023; Bjork et al.,2020)，教学实践应明确限制 GenAI“成品化输出”的使用边界，避免直接供给答案诱发表现与学习的混淆及能力幻觉。GenAI 学习系统应转为以苏格拉底式追问、反例提示等形式提供认知辅助，强制保留学习者的核验与修正步骤，将外部智能可靠地转化为内部心理表征。同时，为防范长期无差别依赖诱发的元认知惰性(Fan et al.,2025)，干预设计需遵循脚手架原则实施动态褪除。从新手期的细粒度支持，逐步向发展期的诊断反思过渡，并在巩固复盘环节最终撤除外部代理，通过认知责任的动态交接，推动学习者完成从外部调节向自我调节的实质性跨越(ATKINSON R K,2003;Ryan et al.,2000)。

在评价范式与主体建构层面，教学体系需协同发力以规避即时表现的遮蔽效应与技术依赖引发的动机置换。鉴于单一的产出导向极易高估真实的学习效果，有必要构建“协同与独立”并重的双轨评价链条。在评估人机协作任务分解与推理透明度的同时，可以通过变式任务强

化脱机状态下近/远迁移能力的独立评估，促使学习者深度重组生成信息，消解协同绩效对独立习得能力的测量遮蔽(Bastani et al.,2025)。此外，过度自动化极易导致学习者在关键环节处于“被替代”地位，引发主体性弱化与心理所有权衰退(Pierce et al.,2001)。因此，整合应用应确立“人为主导、机为辅助”的规范边界，依托理由陈述与选择解释等约束机制(Chatzichristofis,2026; Iranmanesh,2026)，确保学习者在证据判断与最终决策中保有一定控制权，借此维持深层加工的内在动机，使技术真正成为促进独立思维发展的长效支撑。

参考文献

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. Journal of Educational Psychology, 95(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. Journal of Applied Research in Memory and Cognition, 9(4), 475. https://doi.org/10.1016/j.jarmac.2020.09.003

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). Introduction to meta-analysis. John Wiley & Sons.

Chatzichristofis, S. A. (2026). Reframing AI in education: A pedagogical opportunity, not a technocraticthreat. Journal of Research in Science Teaching, 63(1), 97–103. https://doi.org/10.1002/tea.70024

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., … & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

Iranmanesh, A. (2026). Reflection in co-action: AI joins the crit. International Journal of Technology and Design Education, 1–14. https://doi.org/10.1007/s10798-026-10079-6

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V.,& Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. arXiv preprint arXiv:2506.08872. https://doi.org/10.48550/arXiv.2506.08872

Nelson, A., & Eliasz, K. L. (2023). Desirable difficulty: Theory and application of intentionally challenging learning. Medical Education, 57(2), 123–130. https://doi.org/10.1111/medu.14916

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

Pierce, J. L., Kostova, T., & Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. Academy of Management Review, 26(2), 298–310.https://doi.org/10.5465/amr.2001.4378028

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. Perspectives on Psychological Science, 10(2), 176–199. https://doi.org/10.1177/1745691615569000

Villar-Onrubia, D., Cachia, R., Rietz, C., Feltrero, R., Niemi, H., Hallissy, M., & Reuter, R. (2025). Generative artificial intelligence in secondary education: Uses and perceptions from the perspective of early adopters across five EU member states. https://doi.org/10.2760/8636621

Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. Humanities and Social Sciences Communications, 12(1), 1–21. https://doi.org/10.1057/s41599-025-04787-y

Wu, S., Liu, Y., Ruan, M., Chen, S., & Xie, X. Y. (2025). Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation. Scientific Reports, 15(1), 15105. https://doi.org/10.1038/s41598-025-98385-2

Zhang, A. (2025). Information retrieval in the age of generative AI: A mismatch that matters. Legal Reference Services Quarterly, 44(3), 297–306. https://doi.org/10.1080/0270319X.2025.2536920

教育智能体从模型驱动走向意图驱动的编排范式

From Model-Driven to Intention-Driven: An Orchestration Paradigm for Educational Agents

杨海亮[1]，黄慧鹏[1]，吴永和[1]*
[1]华东师范大学教育信息技术学系
*yhwu@deit.ecnu.edu.cn

【摘要】随着生成式人工智能由内容生成走向任务执行，教育智能体开始进入教学设计、课堂支持与学习评价等更复杂的教育场景。但现有应用仍多停留于答疑、推荐、评分等局部环节，缺乏面向课堂全过程的协同机制，难以支撑教学目标、学习活动与评价证据的一体化联动。基于此，本文提出从模型驱动到意图驱动的教育智能体编排范式，构建目标，规划，行动，反馈的编排主链，并从目标层、机制层与治理层阐释其教育意涵。研究认为，教育智能体的核心不在于增强模型输出能力，而在于将教学意图转化为可解释、可执行、可审计的课堂过程。本文可为教育智能体走向教学编排提供理论参考。

【关键词】 教育智能体；意图驱动；模型驱动；编排范式；智能课堂

Abstract: *As generative artificial intelligence evolves from content generation to task execution, educational agents are increasingly entering more complex educational scenarios such as instructional design, classroom support, and learning assessment. However, most existing applications still remain limited to localized functions such as question answering, recommendation, and grading, lacking a collaborative mechanism that supports the entire classroom process and thus failing to enable integrated alignment among instructional goals, learning activities, and assessment evidence. In response, this paper proposes an orchestration paradigm for educational agents that shifts from model-driven to intent-driven logic. It further constructs an orchestration chain of goal–planning–action–feedback and explains its educational implications from the perspectives of the goal layer, mechanism layer, and governance layer. The study argues that the core value of educational agents lies not in enhancing model output capabilities, but in transforming pedagogical intent into classroom processes that are interpretable, executable, and auditable. This paper provides a theoretical reference for moving educational agents from tool-based application toward instructional orchestration.*

Keywords: Educational Agents; Intent-driven; model-driven; Orchestration paradigm; Intelligent classroom

1.引言

生成式人工智能的发展推动智能系统由生成文本走向完成任务，智能体开始具备更强的任务理解、流程分解与工具调用能力。在教育领域，教育智能体也由早期的答疑助手、教学代理和数字化助教，逐步扩展为生成式人工智能支持下的多样化教育支持形态（Bozkurt, 2023）。然而，从当前实践来看，绝大多数教育智能体仍然停留在局部功能增强阶段，其核心仍是围绕模型能力提供服务，即通过更强的语言生成、内容推荐与交互反馈来提升单点任务的完成效率。但教学活动并不是若干孤立任务的简单叠加，而是一个围绕课程目标展开的连续组织过程。它既涉及目标设定、任务实施、课堂互动，也涉及证据收集、形成性评价与结果反馈。若教育智能体仅仅作为一个更强的对话工具存在，那么它固然能够提升信息供给效率，却难以真正进入课堂流程，也难以回应教学中对过程可见、结果可证、责任可溯的现

实要求（Eradze et al., 2019）。因此，教育智能体的发展应当完成从模型驱动走向意图驱动的重要转向。在该转变中，前者面向单点任务优化、关注模型本身的能力边界和智能体输出，后者面向教学全过程支撑、关注教育活动的目标组织和智能体协同。因此，教育智能体不应被理解为独立于教学过程之外的工具集合，而应被理解为能够对齐教学意图、调度课堂流程并沉淀过程证据的编排系统。

2 从模型驱动到意图驱动：教育智能体的范式转向

2.1 模型驱动的教育局限

模型驱动是当前教育智能体最常见的运行逻辑。在这种逻辑下，系统的核心价值主要体现为语言生成、内容归纳、问题回答、资源推荐等能力的提升，教学应用往往以把某一项任务做得更快、更好为目标。这类应用提升了教育活动中的部分效率，但其内在结构依然是从任务输入到模型输出的线性模式。该模式存在一个根本问题，它默认教学任务是可以被切割和孤立处理的，却较少考虑这些任务之间的前后依赖、角色关系和目标一致性。课堂中的教学目标并不天然等于模型的任务目标，学生的学习轨迹也不天然等于系统的对话上下文。如果缺乏对教学意图的识别、对课堂步骤的组织以及对过程证据的回收，教育智能体就只能在局部环节发挥作用，而难以形成真正意义上的教—学—评联动（Newton et al., 2021）。更重要的是，模型驱动强调的是能回答、能生成、能建议，而课堂真正需要的是能对齐、能执行、能协同。在教育场景中，单次高质量回答并不必然带来教学成效，真正有意义的是系统能否在恰当时间、恰当环节、以恰当方式介入，并且为后续评价和改进留下可追溯依据。如果模型驱动关注的是模型能输出什么，那么意图驱动关注的则是系统如何围绕教学目标组织过程并达成结果。

2.2 意图驱动的教育内涵

与模型驱动不同，意图驱动并不以模型输出本身为中心，而是以教学目标、课堂任务和学习证据之间的组织关系为中心。所谓意图，并非一般意义上的用户表达偏好，而是指教学情境中由课程目标、任务要求、学习者状态与评价标准共同规定的行动指向，因而具有目标约束、任务推进与评价导向三重属性。因此，意图驱动的教育智能体，不再只是对输入做出响应，而是要能够围绕教学目标理解任务、分解步骤、组织资源、调用工具并回收反馈。也就是说，系统不仅要知道说什么，更要知道为何说、何时说、对谁说、说完之后如何形成下一步行动。在教学设计支持场景中，这种转向意味着教育智能体开始由内容输出工具转变为任务组织系统（Wang et al., 2025）。在当前教育语境下，意图驱动至少体现为三个方面。第一，它要求教学目标先于工具调用，所有技术介入都应服务于课程标准、学习任务与评价要求。第二，它要求行动过程可组织，系统能够根据课堂进程对任务进行分解与调度，而非仅在被提问时被动响应。第三，它要求过程结果可回收，系统在支持教学的同时保留学习证据，为形成性评价和后续改进提供依据，尤其在虚拟探究等过程性评价场景中更为关键（郑毅，2024）。

2.3 意图驱动为何必然走向编排

意图驱动之所以必然走向编排，根本原因在于教学意图本身并不能直接产生教学成效，只有当其被转译为可执行的活动流程、角色关系与证据链条时，才可能真正进入课堂；因此，当教育智能体被置于教学全过程之中时，其实现形式就不可能只是对话增强，而必然走向系统编排。这是因为教学活动本质上是一种多主体参与、跨环节推进、受规则约束的组织过程。课堂中的目标设定、活动切换、资源调用、学习评价与反馈调整，并不是平铺展开的，而是存在明确的时序关系、角色分工和证据要求（Dillenbourg, 2013）。课堂编排的意义就在于，将教学法、活动流程、角色互动与评价证据组织为一个有序系统。若缺乏这种组织视角，教

学设计、课堂观察与学习评价之间就容易彼此分离，难以形成过程闭环（Eradze et al., 2019）。教育智能体一旦从会回答变成会支持，再从会支持变成会协同，其底层逻辑就不再是单轮对话，而是围绕意图组织流程的编排机制。因此，意图驱动不是模型驱动的简单升级版，而是教育智能体的一次范式转向；而编排，则是这种范式在教育场景中的具体落地形式。基于此，下文进一步从目标层、机制层与治理层对教育智能体的编排范式展开阐释。

3.教育智能体编排范式

为了系统阐述教育智能体如何从单纯的技术工具转化为深度融入教学过程的组织机制，本文构建了教育智能体意图驱动的编排范式理论框架（如图 3 所示）。该框架采用自上而下的三层递进结构：顶部为目标层，明确从教学意图到成功标准的转化，确立价值与方向；中部为机制层，构建目标—规划—行动—反馈的最小编排闭环，实现从输出内容到组织过程的范式转向；底部为治理层，通过可解释、可执行与可审计三大原则对系统运行进行制度性约束。同时，该范式在各层级中重新厘清了教师作为价值裁决者、学生作为反思主体、智能体作为执行单元的协同关系。

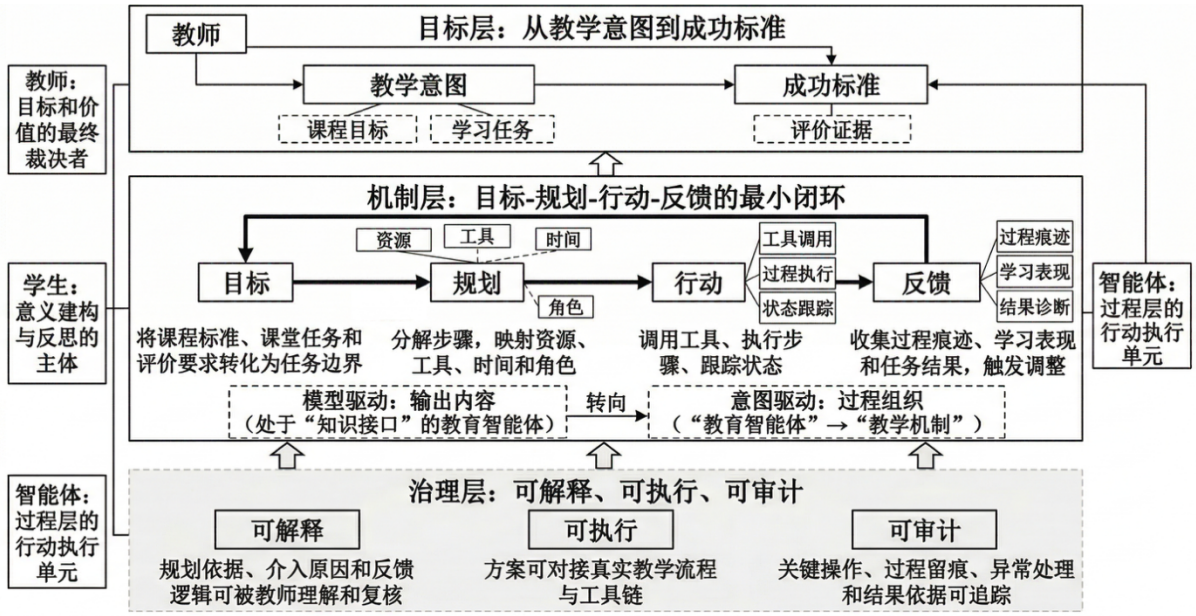


图 1 教育智能体意图驱动的编排范式

3.1 目标层：从教学意图到成功标准

意图驱动的起点不是技术模块，而是教学目标。任何教育智能体若要真正进入课堂，都必须首先回答两个问题：其一，要服务什么样的学习目标；其二，何种学习结果可以被视为达成。也就是说，教学意图必须先被形式化为可以识别、可以检验的成功标准，教育智能体的所有后续行为才能获得方向依据（Newton et al., 2021）。该层面中，教师并非被智能体替代，而是仍然承担目标设定与价值判断的核心职责。教师决定课堂要达成什么、哪些证据能够说明学生发生了有效学习、哪些环节可以交由智能体支撑、哪些决策必须保留人为裁量。也正因如此，教育智能化不能脱离教育目的本身。

3.2 机制层：目标—规划—行动—反馈的闭环

在目标明确之后，教育智能体的关键不再是直接给出结果，而是围绕教学意图形成过程闭环。基于此，本文提出目标—规划—行动—反馈的最小编排闭环，以说明教育智能体如何由内容输出工具转变为课堂过程的组织机制。其中，目标是编排的起点，强调将课程标准、课堂任务与评价要求转化为清晰的成功标准；规划是中介环节，强调将教学目标进一步转译为可执行的活动路径与支持策略；行动是智能体真正进入教学过程的关键，意味着系统能够在

恰当时机介入任务推进；反馈则负责回收学习证据与过程痕迹，为形成性诊断与后续调整提供依据。该闭环改变了教育智能体的存在方式。在模型驱动阶段，系统主要输出内容；在意图驱动阶段，系统开始组织过程。前者为可调用的知识接口，后者为可参与的教学机制。 当目标、规划、行动与反馈形成最小闭环之后，教育智能体就不再只是一个在课堂边缘提供帮助的工具，而成为能够嵌入教学流程、介入活动推进并支撑评价判断的过程性角色。

3.3 治理层：可解释、可执行与可审计

教育智能体之所以不同于通用智能体，关键还在于它必须接受教育治理逻辑的约束。课堂并不是任由系统自由试错的技术环境，而是具有明确责任边界、伦理要求与评价标准的制度场域。围绕教育场景中负责任人工智能与可接受性的讨论，已经为教育智能体治理提供了重要参考（Karran et al., 2025）。因此，意图驱动的编排范式除了要解决如何组织，还要回答如何负责。这意味着，教育智能体的运行必须至少满足三个要求。其一是可解释，即系统为何这样规划、为何这样介入、为何形成这样的反馈，应当能够被教师理解和复核；其二是可执行，即编排方案不能停留在建议层面，而应能够对接真实教学流程与工具链条；其三是可审计，即系统的关键操作、过程留痕、异常处理与结果依据都应保留记录，以便进行责任追踪和效果评估（Rupp et al., 2012）。进一步看，治理层的提出还意味着要重新界定人机关系。教师是目标与价值的最终裁决者，学生是意义建构与反思的主体，教育智能体则是流程层的行动执行单元。这说明，教育智能体的编排范式并不是用机器替代教学，而是在坚持教育主体性的前提下，让智能体承担可量化、可重复、可回溯的支持性流程。

4 结语

面对教育智能体日益广泛的应用趋势，仅仅围绕模型能力展开优化，已难以回应课堂教学的真实需求。教育场景不是若干独立任务的集合，而是一个由目标、活动、评价与反馈共同构成的连续系统。从知识传递走向认知建构与情境认知的教育转型，也说明教学支持系统必须嵌入更完整的学习过程之中。因此，教育智能体的发展不应停留在模型更强的技术逻辑上，而应转向意图如何落实的教育逻辑上。本文提出从模型驱动到意图驱动的教育智能体编排范式，试图说明教育智能体的关键价值，不在于它能提供多少内容，而在于它能否围绕教学目标组织过程、支持协同并沉淀证据。就现实推进而言，未来教育智能体若要走向规模化课堂应用，还必须进一步回应标准、规范与部署治理等问题。只有当教育智能体能够将教学意图稳定转译为可执行、可解释、可审计的课堂流程时，其应用价值才可能由工具可用真正迈向教学可用。

资助：本研究由 2021 年国家社会科学基金重大项目支持（项目编号：21&ZD238）

参考文献

郑毅. (2024). 智能技术赋能虚拟科学探究学习过程评价与自适应反馈研究. 电化教育研究, 45(3). https://doi.org/10.61369/ETR.11355

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. Asian Journal of Distance Education, 18(1). https://doi.org/10.5281/zenodo.7716416

Dillenbourg, P. (2013). Design for classroom orchestration. Computers & Education, 69, 485–492.

Eradze, M., Rodríguez-Triana, M. J., & Laanpere, M. (2019). A conversation between learning d esign and classroom observations: A systematic literature review. Education Sciences, 9(2), Article 9 1.

Karran, A. J., Charland, P., Trempe-Martineau, J., Ortiz de Guinea Lopez de Arana, A., Lesage, A. M., Senecal, S., & Leger, P. M. (2025). Multi-stakeholder perspective on responsible artificial i ntelligence and acceptability in education. npj Science of Learning, 10(1), Article 44.

Newton, S., Alemdar, M., Rutstein, D., Edwards, D., Helms, M., Hernandez, D., & Usselman, M. (2021). Utilizing evidence-centered design to develop assessments: A high school introductory co mputer science course. Frontiers in Education, 6, Article 695376.

Rupp, A. A., Levy, R., Dicerbo, K. E., Sweet, S. J., Crawford, A. V., Calico, T., & Behrens, J. T. (2012). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. Journal of Educational Data Mining, 4(1), 49–110.

Wang, J., Xiao, R., Hou, X., & Stamper, J. (2025). Enabling multi-agent systems as learning de signers: Applying learning sciences to AI instructional design. arXiv. https://arxiv.org/abs/2508.16659

适应性支架教学的运作机制：小学数学学习中支持策略与学生学习表现的交互作用

How Adaptive Scaffolding Works: The Interaction Between Support Strategies and Student Learning Performance in Elementary Math Learning

刘靖熙[1],孙丹儿[1]
*dsun@eduhk.hk

【摘要】 分数概念长期被认为是小学数学学习中的关键难点，传统学习系统多采用静态提示或固定支架，难以在真实课堂中匹配不同学生的学习需要。本研究设计并评估自适应分数学习系统 ScaffoldiaMyMaths，将支架策略具体化为四条可对比的自适应路径（A/B/C/D），以检验不同支持逻辑与学生原有能力水平之间的交互作用。研究采用课堂情境下的准实验设计，对 149 名四、五年级学生开展为期 3–4 周的干预，通过前后测比较学习增益，并进一步比较不同支架路径在不同能力组学生中的差异表现。初步结果显示整体成绩存在显著提升，且更具表现驱动调整的路径对中等水平学生更有效。研究为智能学习环境中的自适应支架设计提供可复现的策略框架与证据基础。

【关键词】 适应性支架教学法；小学数学；智能辅导系统；支持策略

Abstract: *Fractions remain a persistent challenge in elementary mathematics. Many technology-based learning systems still rely on static scaffolding, which may not fit learners with diverse prior achievement. This study develops and evaluates ScaffoldiaMyMaths, an adaptive learning system that operationalizes four distinct scaffolding paths (A/B/C/D) to examine how support strategies interact with student proficiency in fraction learning. A classroom-based quasi-experiment (N=149) was conducted with Grade 4–5 students over 3–4 weeks using comparable pre/post tests. Results indicate a statistically reliable overall improvement, and performance-sensitive scaffolding paths yield stronger gains for medium-achieving students. The study contributes an inspectable and testable design blueprint for adaptive scaffolding in intelligent learning environments.*

Keywords: Adaptive scaffolding, Elementary mathematics, Intelligent tutoring systems, **S**upport strategies

1.研究背景

分数学习被认为是小学数学走向更高阶推理的重要门槛，其困难主要来自概念结构的转变：学生需要从整数直觉转向“部分—整体”“比率关系”等理解，因此容易形成长期性误解并影响后续学习（National Mathematics Advisory Panel, 2008）。计算机辅助学习环境为分数教学提供了支持契机，它能够整合多表征并提供即时反馈，促进学生在互动中修正错误概念（Moreno & Mayer, 2007）。然而，现实中许多学习系统仍以静态提示或统一化支架为主，难以适配学生能力差异，且支架过多可能带来无关负荷或抑制自主推理（Sweller et al., 2011）。

因此，关键问题并非“是否提供支架”，而是“如何依据学习者状态动态调整支架”。目前研究仍缺乏在同一平台下对不同支架策略进行系统比较，并检验其与学生前置水平交互机制的证据（VanLehn, 2011）。基于此，本研究以小学分数学习为情境，构建可对比的自适应支架

策略框架，探究不同支架逻辑在不同能力学生中的差异化效果，为智能学习环境的精准支持设计提供依据。

2.研究背景

小学数学中的学习困难往往并非源于能力缺陷，而是教学支持与学习者需求之间存在失配，因此“如何提供恰当支持”始终是关键议题。在此背景下，计算机辅助教学（CAI）被认为能够通过可重复练习、即时反馈与个性化节奏，为数学薄弱学生提供有效补偿，并在实证研究中显示出对数学学习表现的促进作用（Kroesbergen & Van Luit, 2003）。然而，分数学习作为小学阶段典型的高难度概念主题，学生常出现顽固性误解与迁移困难，这使得单纯的练习强化难以解决深层的概念建构问题，学习支持需要更具结构性与适配性。

支架理论指出，当学习者无法独立完成任务时，外部支持能够帮助其在可达范围内推进理解与问题解决（Wood et al., 1976）。后续研究进一步区分了预设、相对稳定的“硬支架”与基于学习状态动态调整的“软支架”，强调支架应随学习进展逐步渐退，以避免依赖并维持适度挑战（Saye & Brush, 2002）。但在真实课堂中，教师难以持续为不同起点学生提供高频、个别化的动态支架，因此技术化支架成为可扩展的解决路径。与此同时，现有分数领域的技术干预研究仍存在不足：许多研究样本较小或干预组件混杂，导致难以分离并验证“具体哪一种支架策略更有效、对哪类学生更有效”。因此，有必要在同一学习系统中构建可对比的支架策略，并检验其在不同能力水平学生中的差异化作用机制，为智能学习环境的自适应支持设计提供更清晰的证据基础。

3. 研究方法

3.1. 研究问题

本研究旨在在真实课堂情境中检验 ScaffoldiaMyMaths 所内置的四类支架路径（A/B/C/D）在分数学习中的差异化成效，因此研究围绕三个核心问题展开：

RQ1. 学生在干预前后的整体学习成绩是否存在显著变化？

RQ2. 不同支架路径之间的学习增益是否存在差异？

RQ3. 支架策略的成效是否因学生原有学业水平而异？

通过比较路径效果与学习者水平的交互关系，本研究尝试在小学数学分数知识点学习中揭示自适应支架的条件性机制。

3.2. 研究方法

本研究采用准实验设计，在自然课堂中实施“前测—平台学习—后测”的教学干预流程（如图 1 所示）。研究对象来自大湾区四所小学，共获得 149 份完整数据（四年级 53 人、五年级 96 人）。学生以便利取样方式参与，由学校依据课程安排自愿加入学习活动。干预前已取得伦理审批，并向家长与学生分别完成书面知情同意与同意书，确保参与过程的自愿与透明性。

干预持续三到四周，学生在数学课堂或自习时段使用平台完成 3-5 个分数核心单元学习，每次学习时间限定在 30 分钟以内，以贴近学校日常教学节奏。学生首先完成前测以建立学习基础水平，随后在平台中以不同支架路径开展学习活动；每条路径通过调整提示数量、提示节奏与反馈策略构成不同的支架逻辑。干预结束后一周内实施与前测等值的平行后测，用以测量学习变化。

研究使用的主要测量工具为研究团队基于课程标准与教师经验编制的分数测验，分别作为前测与后测使用。试卷由经验丰富的数学教师审阅，经专家评估后定稿，以确保其效度。为避免练习效应，前后测采用平行卷结构，覆盖相同知识点与同等难度，但更换具体数字与情

境。每份试卷包含 13-15 道多类型题目（选择、填空、运算），满分 50 分，作答限时 30 分钟，以反映学生在课堂情境下的真实数学能力。

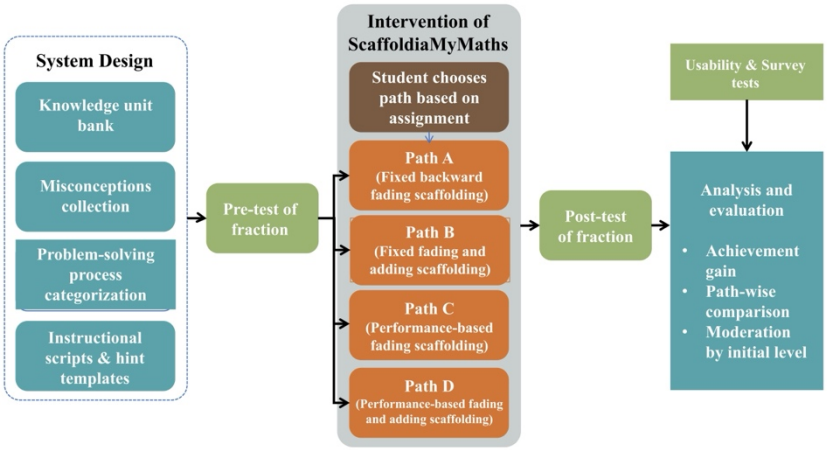


图 1 ScaffoldiaMyMaths 实验干预流程

整体而言，本研究通过课堂情境下的准实验方法、四类型支架路径的对比设计，以及前后测量化数据的收集，检验 ScaffoldiaMyMaths 在分数学习中的成效差异与条件性作用模式，为理解自适应支架在小学数学中的学习调节机制提供实证基础。

4. 实验研究结果

本研究已完成课堂情境下的试验实施与初步数据分析。整体结果显示，学生在使用 ScaffoldiaMyMaths 完成分数学习单元后，其数学成绩在后测相较前测出现小幅但显著的提升，平均分由 62.07 提升至 64.51（$p < .05$），说明该平台在分数学习情境中具有一定的促进作用（如图 2 所示）。

进一步比较四种支架路径（A/B/C/D）的学习增益后发现，不同支架策略的效果呈现明显差异趋势（如图 2 所示）。其中，路径 D 的学习提升最为突出，路径 C 次之；路径 A 的提升幅度较小，而路径 B 出现轻微下降。具体而言，路径 C 与路径 D 的平均增益最为明显（约+5.36 与+6.19 分），路径 A 仅呈现温和增长（约+1.35 分），路径 B 则略有负向变化（约−1.16 分），该结果提示“支持强度更高”并不必然带来更好的学习效果，支架策略与学习者需求之间的匹配可能更为关键。

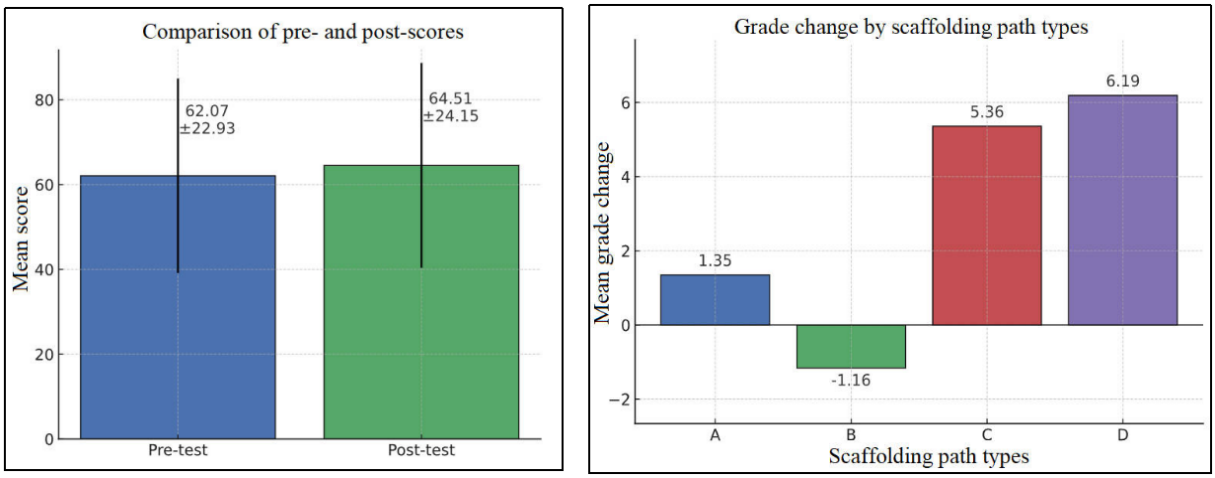


图 3 干预前后测成绩对比图，及不同支架路径对应的平均增益柱状图

结合学习者前测水平的分层分析也显示，自适应程度更高、对表现更敏感的路径（尤其是 C/D）在部分学生群体中更容易体现优势，而高水平学生的整体提升相对有限（如图 3 所示）。

这意味着自适应支架的有效性可能具有条件性，其作用机制需要进一步结合学习过程数据与个体差异来解释与优化。

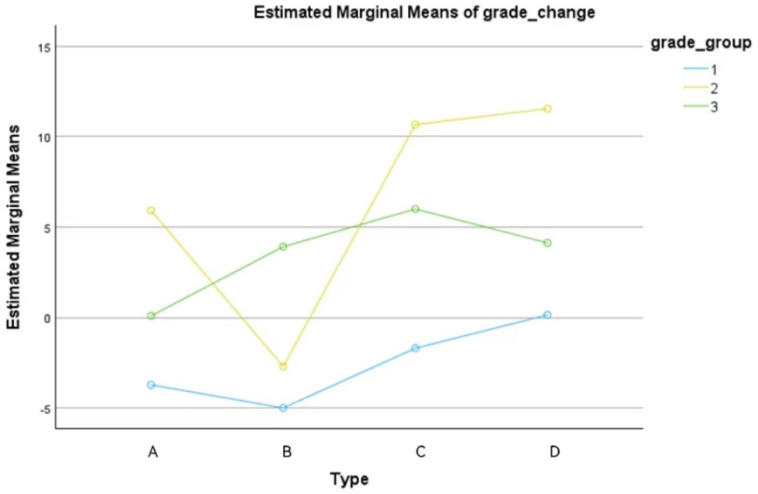


图 3 高分组、中分组和低分组学生成绩变化折线图

5. 下一步将展开的研究工作

基于本研究已完成的课堂试验与初步分析，下一阶段将进一步从“结果验证”走向“机制解释”，以提升研究的解释力与可推广性。首先，本研究将引入更细粒度的学习过程数据（如作答时长、连续错误模式、提示调用轨迹与平台交互行为），从学习轨迹层面解释不同支架路径为何在不同学业水平学生中呈现差异化效果，并识别哪些关键行为信号能够更准确触发“渐退”或“补充提示”的自适应决策。

其次，本研究将对现有自适应规则进行优化，使支架调整不仅依赖正确率，还能结合反应时、错误类型与帮助寻求等指标形成更稳定的多信号判断机制，从而提升系统对学习者即时需求的识别精度与支持匹配度。

最后，为增强外部效度与实践价值，后续研究将扩大样本与实施场域，在更多学校与不同年级开展验证，并探索将该框架迁移至其它数学主题，以检验自适应支架策略在不同知识领域中的稳健性与可复用性。

参考文献

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*(3), 309-326.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24*(2), 97-114.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89-100.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development, 50*(3), 77-96.

ISBN 978-626-93238-1-4