

基于大语言模型的苏格拉底教学策略研究——以小学数学为例

Research on Socrates' teaching strategy based on LLM: Taking primary school mathematics as an example

张誉月¹, 贾积有^{1*}

¹ 北京大学教育学院教育技术系

persistbetter@stu.pku.edu.cn, *jjy@pku.edu.cn

【摘要】生成式人工智能（GenAI）在教育领域的应用正逐步由单纯的内容生成转向深层次的教学交互。然而，现有通用大模型往往因直接提供答案而削弱学习者的认知投入。针对小学数学教学中深度引导不足的问题，本研究构建了一个苏格拉底式智能辅导系统（MIATS V3.0）。研究采用 2（大模型智能体支持的苏格拉底对话：有/无）×2（人工设计的提示：有/无）的组间实验设计，以 215 名小学四年级学生为对象，通过近迁移与远迁移任务，系统检验了 GenAI 驱动的苏格拉底对话与传统人工提示在数学概念建构中的独立效应与协同机制。研究发现，尽管学生对 AI 助手的技术接受度较高，但在深层信任与情感依赖方面仍显著倾向于人类教师。本研究不仅验证了苏格拉底教学策略在贯彻教育学原则上的有效性，也为应对 GenAI 教育应用中“认知外包”的风险提供了实证依据与设计范式。

【关键词】智能辅导系统；大语言模型；苏格拉底式教学；小学数学；人机协同

Abstract: The application of generative artificial intelligence (genai) in the field of education is gradually changing from simple content generation to deep-seated teaching interaction. However, the existing general large-scale model often weakens learners' cognitive engagement by directly providing answers. Aiming at the problem of insufficient depth guidance in primary school mathematics teaching, this study constructed a Socratic intelligent tutoring system (miats V3.0). The study adopts an inter group experimental design of 2 (Socratic dialogue: Yes/no) × 2 (artificial prompt: Yes/no). 215 fourth grade primary school students are selected as the subjects. Through the near transfer and far transfer tasks, the independent effect and synergistic mechanism of the Socratic dialogue driven by genai and the traditional artificial prompt in the construction of mathematical concepts are systematically tested. The study found that although students' technical acceptance of AI assistants is high, they still tend to be human teachers in terms of deep trust and emotional dependence. This study not only verifies the effectiveness of Socrates' teaching strategy in implementing the principles of pedagogy, but also provides an empirical basis and design paradigm for coping with the risk of "cognitive outsourcing" in the application of GenAI education.

Keywords: Intelligent tutoring system; Large language model; Socratic teaching; Primary school mathematics; Human machine cooperation

1. 引言

生成式人工智能（Generative AI）正在迅速重塑教育技术的格局。以大型语言模型（Large Language Models, LLMs）为代表的 GenAI 系统展现出令人瞩目的自然语言生成与对话能力，引发了将其整合至智能辅导系统（Intelligent Tutoring Systems, ITS）的新一轮热潮（贾积有&张誉月，2023）。然而，标准的 LLM（如 GPT-5）并不等同于一个有效的教学智能体，GenAI 经过基于人类反馈的强化学习（RLHF）训练，能够生成语法正确、内容相关回应的 AI，且通常优先考虑有用性，倾向于成为有用的助手，即直接提供解决方案(Ding et al., 2024; Puech et al., 2024)。而有效的教学对话则需要遵循一系列基于学习科学的原则，例如：准确把握学生的最近发展区，管理认知负荷，并促进元认知与自我解释（Chi & Wylie, 2014）。

为了避免 GenAI 绕过了学生学习所必需的认知过程，之前的研究者引入了“生产性失败”（Productive Failure）等策略(Puech et al., 2024)，故意让学生在接受正规指导之前，先在次优解中挣扎和探索，从而促进更深层次的概念理解。即苏格拉底式提问法（Socratic Questioning），其核心在于通过一系列引导性问题，让学生参与对话、独立思考并学会自己解决问题，从而培养更深层次的掌握和能力（Al-Hossami et al., 2023）。目前，已有研究利用 GenAI 的生成能力，将苏格拉底式教学法与 ITS 相融合，例如可汗学院研发了基于 GPT4 的苏格拉底式导师 Khanmigo，鼓励学生深入思考，而非直接提供答案，培养学生的批判性思维(Favero et al., 2024)。此外 Chang(2024)开发了多智能体进行苏格拉底式提问推理的 SocraSynth 提示语模板，通过智能体之间的辩论与合作，启发人类推理。Lee 等（2025）开发并部署了名为 Socratic Mind 的生成式 AI 形成性评估工具，通过动态提问迫使学生阐述其代码背后的逻辑、假设及潜在边界条件。Hashmi 和 Rebello(2025)将苏格拉底式 AI 运用于物理课的力学教学中后，学生的提问中属于“特异性问题”的数量从 10-15%的比例飙升至 100%，表明苏格拉底式 AI 培养了学生“如何像物理学家一样提问”的元认知策略。

传统的 ITS 已证明其在支持学生数学学习中的程序性技能练习与自适应学习路径方面的有效性（Bartelet 等，2016；贾积有等，2023）。但传统的、基于规则的 ITS 在实现苏格拉底式对话时，受限于预定义话语库的狭窄性与手动定制的繁复性，难以应对开放、动态的真实学习过程。苏格拉底支架与传统 ITS 支架的不同在于，传统 ITS 支架的核心逻辑在于“做减法”——即通过外部支持暂时性地降低任务的难度，使得学生能够完成那些独立无法完成的任务，例如提供程序性、视觉性的提示，或计算、视觉与工具支架。苏格拉底式支架并不急于消除障碍，有时甚至故意制造障碍，即“做加法”——通过引入反例、质疑前提或要求深度论证，暂时性地增加任务的认知挑战，其认知机制侧重于“问题化”与认知冲突。

尽管个案研究展示了苏格拉底对话教学的潜力，但现有的通用大模型在扮演苏格拉底导师时往往表现不稳定。例如无法从学生错误的推理步骤中推断出底层的概念误解。对正确答案给予积极肯定，但对错误答案的反馈往往含糊其辞或过于礼貌，未能有效利用错误作为教学契机等等。此外，虽然高成就学生可能受益于高效的答案检查，但低成就学生从将问题分解为苏格拉底子问题的过程中受益更多(Ding et al., 2024)。

综上，为设计优秀的苏格拉底式对话智能体，揭示在 ITS 中传统支架与基于 GenAI 的苏格拉底对话支架的不同作用机理，探究更有效的 GenAI 教学应用策略，本研究旨在通过实证方法，检验并评估在智能辅导系统中，由 GenAI 驱动的苏格拉底式对话智能体的实际教学有效性，及其与传统人工设计提示系统之间的协同机制与差异化影响，以回答以下几点问题：

- 1) 基于 GenAI 的苏格拉底对话智能体在实际互动中，其回应的质量与教学适切性如何？
- 2) 基于 GenAI 的苏格拉底对话与传统人工设计的通用型提示两种脚手架，相较于单一脚手架，对学习者的最终学习结果（成绩、能力、情感）分别有怎样的影响？

2. 相关研究

2.1. 苏格拉底教学法在小学数学中的适用性

长期以来，小学数学教育普遍面临着一种结构性困境：一方面，学生能够熟练进行四则运算，甚至在标准化测试中取得高分；另一方面，面对非结构化的问题情境或需要深度逻辑推理的数学任务时，学生往往表现出概念理解的肤浅与迁移能力的匮乏，对基本数学概念缺乏深刻理解（Ma, 1996）。

苏格拉底式提问已被证明在促进学生新知建构(Goldin et al.,2017)、概念转变(Mulleret al.2007)、批判性思维形成(Shoop et al,2011)等方面具有重要价值。与人文社科领域涉及价值判断、多元阐释或情感体验的议题不同,数学学习远非事实与程序的简单记忆,其核心在于概念理解、逻辑推理与问题解决能力的建构。传统的灌输式教学易导致学生陷入机械模仿与算法套用,而未能触及小学生的质疑精神和严谨的数学精神。小学阶段(约6-12岁)的学生主要处于皮亚杰的“具体运算阶段”,这一阶段儿童思维的关键特征是守恒概念的形成和逻辑运算能力的萌芽。苏格拉底教学法通过连续、适切的提问,可以引导小学生审视自身假设,经历困惑、反思直至自我建构出新的理解——恰好与数学的“再发明”学习过程高度契合。

苏格拉底教学法在小学数学实际教学应用中,Project SEED 项目为了解决大班额教学中“少数人思考,多数人旁观”的弊端,开发了一套独特的非语言反馈系统,极大地提高了课堂参与度(Moilanen,1987)。该项目绕过了死记硬背的算术或事实,通过严密的逻辑问答,使来自低社会经济地位背景的小学生完全有能力掌握通常在中学甚至大学才教授的数学概念,如指数运算。Garlikov(2001)的教学实验表明,通过纯粹的苏格拉底问答,可以在不直接讲授的情况下,引导三年级学生理解二进制甚至复杂的位值原理。意大利某小学五年级学生使用基于AI的苏格拉底辅导系统学习数学问题解决,有效引导学生发现解题线索,提升了学生的数字素养(Avella,2025)。因此,接受苏格拉底教学的学生在面对新题型时,迁移能力更强,因为他们掌握的是原理而非题型,并培养批判性思维,在思考时进行自我审查,构成了科学素养的基本思维习惯。

2.2. 优秀的苏格拉底提问策略

尽管苏格拉底教学在诸多优势,但需要注意谨慎使用提问策略,因为不适当的内容有时会弊大于利,例如若提出总是容易回答的问题不会导致更深入、更广泛的学习。Liu等(2025)提出的Guide Eval框架为评估和设计面向小学数学的GenAI导师提供了理论基石,将教学指导能力解构为三个核心维度:感知、编排和启发。

感知是指AI准确推断学习者当前认知状态(如正确推理、错误概念、理解或困惑)的能力。现有的LLM在处理显性信号(如学生直接说我不懂)时表现尚可,但在处理隐性线索(如错误的推理过程但得出了正确答案,或仅仅是犹豫)时存在显著缺陷。赵晓伟等(2025)将苏格拉底式提问视为一种递归思维过程,它将复杂问题分解为更简单、相关的子问题。学生的状态可以分为提问、正确回答、错误回答、学习动机不足或提出不相关的问题。基于苏格拉底式问题支架的转换逻辑,评价学生提问的系统性与逻辑性。编排涉及根据学习者的状态动态调整教学策略(如简化问题、类比、步骤分解)。许多通用大模型在面对学生的持续困惑时,缺乏策略的灵活性,往往倾向于重复之前的解释,而非重构解释路径。Figueiredo(2025)提出了模糊、符号与情境的控制框架。该框架引入了符号化的短时记忆模式,用于在多轮对话中持续追踪学习者的状态变量(如已掌握的概念、当前的迷思、情绪状态)。启发旨在通过有针对性的提问激发学生的反思和深层推理。高质量的启发应当遵循布鲁姆分类法,从事实回忆向高阶应用和评价跃迁。然而,如何生成既不过于简单(导致无聊)也不过于困难(导致挫败)的问题,是当前技术的一大挑战。

有效的批判性思维与探究式学习也遵循一定的阶段性,只有在具备一定的“已知”水平后,针对这些知识的深度理解、联结与应用进行“苏格拉底式”的批判与探究,才是有意义且高效的。因此,一个理想的、作为苏格拉底导师的GenAI,其核心智慧应体现在:能准确判断学习者所处的认知阶段;在基础巩固期,能提供清晰、准确的支持;在理解深化期,能适时、自然地转换为启发与诘问模式,让引导过程自然而然地发生。

小学生的认知特征为思维具象化、单向化、去中心化不足、难以理解抽象概念和复杂的递归关系，因此需要将问题通过情景具象化呈现、双向追问、递进式归纳、进行视角转化等。根据布鲁姆教育目标分类学，我们可以将其分为不同层级，如表 1 所示。

表 1 适用于小学数学的苏格拉底提问策略

提问类型	功能目标	典型话术示例（以分数为例）	认知层级
澄清性提问	确保概念定义的清晰，消除歧义。	当你说是“一半”的时候，具体是指什么的是这个圆的面积还是周长？	记忆/理解
假设性提问	探索变量变化带来的后果，培养函数思维。	如果我们把分母从 2 变成 4，但这块蛋糕的大小不变，那么每一份会发生什么变化？	应用/分析
证据探寻提问	要求学生为观点提供逻辑或事实支撑，培养论证能力。	你为什么确信 $\frac{1}{3}$ 比 $\frac{1}{2}$ 小？你能画图或者用小棒证明给我看吗？	评价
反例/归谬提问	挑战错误观念，引发认知冲突。	你说分母越大分数越大。那么 $\frac{1}{100}$ 块披萨比 $\frac{1}{2}$ 块吃得更饱吗？	评价/创造
元认知提问	监控思维过程，促进反思。	刚才你是怎么发现自己第一步算错了的？是什么让你改变了想法？	创造/元认知

3. 基于大型语言模型多智能体的苏格拉底式教学设计

3.1. 智能体设计

为在文心一言大模型上，我们设计了一个智能体“小学数学知识库”，其设置如图下表 1 所示。提示词质量对于生成通用型提示的效果有着显著的影响。作为生成通用型提示的关键变量，不同类型和构造的提示词对模型的理解和输出结果产生不同的效果。大模型提示词的构造方法多样，包括零样本学习、单样本学习和少样本学习等。Jia 等人的研究表明单样本和少样本学习会导致 GPT 过度依赖给出的例子，而忽略了数学问题的具体特点和变化，而零样本学习使 GPT 能够根据具体问题自由生成提示，产生的通用型提示得分最佳（Jia 等，2024）。因此本文主要采用零样本学习的方法。

3.2 实验设计和实验流程

本研究以自主研发的 MIATS V3.0 系统为核心实验环境。该系统在 V2.0 个性化练习与自适应测试框架基础上（贾积有，乐惠骁等 2023；贾积有，张誉月等 2023），进行了关键的功能增强：第一，保留了基于项目反应理论（IRT）的学生能力评估与个性化题目推荐机制，以及一套由人工设计的、详细程度递进的通用型提示（如从解题思路到具体定理的提示）。第二，也是本研究的核心创新，在于集成了一个基于 GenAI 的苏格拉底式对话代理。该代理通过精心的提示词工程进行约束，旨在严格遵循既定的苏格拉底式引导原则（如以提问回应、聚焦单一概念、从学生已知出发等），而非自由生成任意对话，从而在提供开放性互动的同时确保教学行为的理论保真度。

在准备与开发阶段完成了系统原型设计与实验材料准备。依据北师大版教材，与学科教师协作遴选与改编测试题目；为不同实验条件手动编辑提示词与设计对话逻辑，并完成相应实验网页的开发；最终为所有参与学生分配实验账号。实验的核心干预任务选自小学四年级数学课程中的核心且具有一定复杂性的主题（如“分数的初步认识与应用”、“长方形周长与面积”）。这些主题的问题解决过程涉及多步骤推理和概念性理解，而非简单计算，为苏格

拉底式对话与传统人工设计提示的应用提供了必要空间。在为期四周的干预中，学生将在系统引导下完成一系列围绕选定主题设计的练习。

为严谨评估系统效能，本研究设计了区分近迁移与远迁移的测评任务。近迁移任务在表面特征和解题结构上与干预中的练习题目高度相似，主要用于检验学生对所学基础技能与核心概念的掌握与直接应用能力。远迁移任务则经过精心设计，其问题情境、表述方式或所需策略与练习题目有明显差异，但深层数学结构或原理相通。解决远迁移任务要求学习者能够剥离问题的表层干扰，识别并提取可迁移的深层原理或策略，从而有效评估其高阶思维与灵活应用知识的能力（Barnett & Ceci, 2002）。

本研究中的前后测量表是由 Pintrich (1991) 的 MSLQ 量表改编而来的。实验结束后，共收集到了 215 份量表，首先对收集到的数据进行清洗，去除其中由被试随意填写的无效数据。剔除了作答数据方差小于 0.1 的样本，最终获得有效的样本 205 份。对这 205 份样本进行信效度分析，其中信度分析主要关注的指标是 Cronbach's alpha 为 0.9 (Adeniran, 2019; Rizopoulos, 2007)，而效度分析中量表 KMO 参数为 0.83，Bartlett 检验显著性小于 0.001，适合进行探索性因子分析。

3.3 实验对象

本研究的目标人群为郑州某小学的四年级学生，这一阶段的学生已掌握基础的算术运算，但正开始接触更为抽象和需要多步骤推理的数学概念（如分数、面积计算），是考察苏格拉底式教学法能否在概念理解与问题解决上提供有效支持的理想学段。本实验在郑州市某小学展开，以四年级学生为研究对象。选取四个平行班级，每班约 50 至 55 人，总样本量为 215 人。实验周期为一个月，每周教师带领学生到机房上机使用系统一次，40 分钟，共 4 次。

为检验不同的 AI 干预条件（Exp1, Exp2, Exp3）能否更有效地提升学生的高阶思维水平及其在复杂问题解决中的迁移成绩，在分组设计上，研究采用 2（大模型智能体支持的苏格拉底对话：有/无） $\times$ 2（人工设计提示：有/无）的组间实验设计，形成四个实验条件：第一组为对照组，该组学生不使用任何对话功能或传统人工设计提示系统；第二组为实验组 1（Exp1），同时提供传统人工设计的通用型提示词以及大模型智能体支持的苏格拉底式对话；第三组为实验组 2（Exp2），仅提供通用型传统人工设计提示，不启用苏格拉底式对话；第四组为实验组 3（Exp3），仅启用大模型智能体支持的苏格拉底式对话，不提供传统人工设计的通用型提示。为确保组间可比性，依据前测成绩对学生进行随机分配，以保障各组学生在数学先备知识与能力上无显著差异。



图 1 2025 年第一次上机图

4. 实验结果

首先, 样本学生在技术可及性与基础使用能力上表现出较高水平, 但其应用层次普遍较浅, 以信息查询功能为主, 尚未深入触及以促进思维发展为目的的交互式学习。超过 98% 的学生曾使用过至少一类智能工具, 其中“学习软件” (如作业帮、小猿搜题) 使用率最高 (76.5%), 而“人工智能 AI” (如豆包、文心一言) 的使用率也超过半数 (54.9%), 表明以 GenAI 为核心的新兴数字工具在小学中高年级已具备相当的渗透率。在使用频率上, 78.4% 的学生每周至少使用数次。然而, 从使用目的来看, 学生的行为多集中于工具性的“查资料” (64.7%) 和“查不认识的字词” (47.1%), 而利用其进行“问问题、聊天” (29.4%) 或“帮我改作文” (31.4%) 等更具对话性与建构性的活动比例相对较低。因此, 当前学生群体虽为“数字原住民”, 但其技术应用模式尚未充分融入学习过程的深层认知环节, 多数仍停留在对传统搜索引擎功能的替代与延展上。

其次, 学生的数学学习动机呈现出典型的外部目标导向特征, 但在元认知自我调节等深层学习策略上存在差异, 也为本研究检验苏格拉底式 AI 导师在激发与引导元认知过程方面的效能提供了切入点。前测量表数据显示, 学生在“外部目标导向” (如“为了让父母老师看到进步而学”, $M=4.71$) 和“任务价值认同” (如“数学知识在生活中用得上”, $M=4.57$) 上得分最高, 显示出强烈的外部激励驱动特性。同时, 学生在“自我效能感”和“控制信念”上总体积极, 普遍相信通过努力能够学好数学 ($M=4.43$)。然而, 在指向内在兴趣与挑战的“内部目标导向”维度, 均值相对较低 ($M=3.89$), 尤其在“愿意选择有挑战性作业”题项上得分最低 ($M=3.76$)。在“元认知自我调节”维度, 学生在“调整学习方式” ($M=4.31$) 和“尝试不同方法” ($M=4.24$) 上表现较好, 但在更具主动性和计划性的“课前预览” ($M=4.02$) 和“通过自我提问保持专注” ($M=3.88$) 上略显不足。综上, 学生可能更适应外部明确的任务驱动, 而在自主规划、深度追问与持续反思等需要内部驱动的元认知活动方面尚有发展空间, 而这正是苏格拉底式教学法旨在强化的核心。

再者, 学生对“好的智能助手”有明确且符合教育学理念的功能期待, 但初始信任度显著偏向人类教师, 凸显了构建可信赖教育 AI 的迫切性与挑战性。前测中, 学生最为看重 AI 助手“帮助思考而非直接给答案”的能力 (认可度 86.3%), 同时也高度期待其能提供“正确且能解释为什么”的回答 (64.7%) 以及“根据我的数学水平来出题和讲解” (58.8%)。这些期待与近期 ChatGPT 推出的“学习模式” (Study Mode) 所强调的苏格拉底式引导、个性化路径与深度理解理念高度吻合。一项 2025 年的元分析也证实, GenAI 在基础教育阶段对学生认知和情感态度发展的积极调节作用尤为显著。然而, 尽管存在高功能期待, 在面对“智能助手与人类老师二选一”的假设时, 高达 86.3% 的学生明确选择更希望获得人类老师的帮助。也就是说, 学生对 AI 教育技术的信任是其采纳和感知有用性的关键前因, 即使学生理论上认可 AI 应具备的理想教学行为, 但在情感信任与依赖上, 仍将人类教师置于不可替代的地位。其原因可能是对 AI 技术可靠性、情感共情能力缺失的担忧, 或是对传统师生关系的习惯性依赖。Hew (2025) 指出, 教育的关键在于引导学生“用 AI 学”, 而非“让 AI 代学”, 而这有赖于对 AI 工具设计理念和使用方式的深刻变革。

综上所述, 前测数据赋予了一副清晰的学生画像: 学生是熟练的基础技术使用者, 受外部动机驱动, 对高阶 AI 教学功能有理性认知, 却对将深层学习过程托付于 AI 持本能保留。我们的干预并非面对一张“技术白纸”, 而是要在一个技术已普及、但深度教育融合尚属初探、信任尚未建立的环境中, 实证检验一种精心设计的 (遵循苏格拉底原则的)、以激发元认知与批判性思维为目标的 GenAI 交互模式, 能否在实践中真正弥合学生“高期待”与“低信

任”之间的鸿沟，并有效促进其深度问题解决能力的发展。后续的正式实验将以此基线为参照，细致考察不同智能支持模式如何影响学生的学习过程与结果。

5. 结论与讨论

5.1 “高期待与低信任”：人机协同教育的现实张力

本研究发现学生对 AI 角色的认知存在明显矛盾，呈现“高期待、低信任”的态势。前后测数据显示：一方面，学生对“能引导思考而非直接给答案”的 AI 助手抱有较高的功能性期待（认可度达 86.3%），并在使用 MIATS 系统后对其易用性与交互质量给予积极评价（平均分>4.75）；另一方面，在最终选择中，高达 98% 的学生明确表示更倾向于人类教师。

尽管基于智能体的系统在知识传递与逻辑引导（即“教”）方面表现良好，但在情感支持、共情互动与深层信任建立（即“育”）方面，仍难以替代人类教师。因此，未来的智能辅导系统不应以完全取代教师为目标，而应定位于“人机协同”中提供认知支架的角色。正如 Hew (2025) 所言，关键在于引导学生“运用 AI 学习”而非“让 AI 代学”。AI 可承担知识诊断与程序性训练任务，而将情感激励、价值引导及复杂的非结构化教学决策留予人类教师。

5.2 两种支架的辩证：做加法与做减法的动态平衡

本研究通过 2×2 实验设计，深入探讨了传统人工提示（做减法）与大模型智能体支持的苏格拉底式对话（做加法）的协同作用。传统智能辅导系统 (ITS) 的支架常通过降低任务难度（如提供公式提示）来辅助学生，而苏格拉底式对话则通过反问与质疑暂时增加认知负荷，以激发认知冲突。

研究结果支持一种动态混合的支架模型：在学习初期或学生遭遇显著认知障碍时，确定性较高的人工提示有助于降低挫折感；而在概念巩固与迁移阶段，GenAI 驱动的苏格拉底对话则能促进深度学习与元认知策略的发展。特别是对于正处于具体运算向形式运算过渡期的小学四年级学生，单一的“诘问”可能导致认知超载，单一的“提示”则易形成依赖。因此，理想的智能辅导系统应能依据学习者的实时认知状态（由认知评估智能体感知），灵活调整“加法”与“减法”两类支架的配比。

5.3 研究局限与展望

本研究仍存在若干局限。首先，实验样本仅来源于郑州市某小学，干预周期为四周，苏格拉底式对话长期实施是否引发“对话疲劳”或认知倦怠，尚需纵向研究验证。其次，尽管智能体架构降低了错误率，但在应对极端发散的学生提问时，系统仍可能出现逻辑闭环缺失。

未来研究可从以下方向展开：第一，扩大样本规模与多样性，检验该模型在不同学段与学科的普适性；第二，引入情感计算模块，使智能体能够识别学生的挫折或焦虑情绪，并据此调节对话的温度与节奏；第三，进一步融合多模态数据（如操作行为、响应延迟等），优化认知状态评估模型，以构建更为精准、更具温度的下一代智能辅导系统。

致谢

本文是全国教育科学规划国家一般课题“基于大数据挖掘的学生智能评测和辅导研究”(BCA220208) 的一项研究成果。

参考文献

Adeniran, A. O. (2019). Application of Likert scale's type and Cronbach's alpha analysis in an airport perception study. Scholar journal of applied sciences and research, 2(4), 1-5.

- Al-Hossami, E., Bunesco, R., Teehan, R., Powell, L., Mahajan, K., & Dorodchi, M. (2023). Socratic questioning of novice debuggers: A benchmark dataset and preliminary evaluations. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)* (pp. 709-726).
- Avella, B. (2025). Socratic AI Tutoring in Primary School Mathematics: A Case Study on the Development of Problem-Solving and Digital Competence According to DigComp 2.2.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, 128(4), 612.
- Bartelet, D., Ghysels, J., Groot, W., Haelermans, C., & Maassen van den Brink, H. (2016). The differential effect of basic mathematics skills homework via a web-based intelligent tutoring system across achievement subgroups and mathematics domains: A randomized field experiment. *Journal of Educational Psychology*, 108(1), 1–20.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4), 219-243.
- Ding, Y., Hu, H., Zhou, J., Chen, Q., Jiang, B., & He, L. (2024). Boosting large language models with Socratic method for conversational mathematics teaching. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)* (pp. 472–476). Association for Computing Machinery.
- Ding, Y., Hu, H., Zhou, J., Chen, Q., Jiang, B., & He, L. (2024). Boosting large language models with Socratic method for conversational mathematics teaching. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24)* (pp. 472–476). Association for Computing Machinery.
- Favero, L., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024, October). Enhancing critical thinking in education by means of a Socratic chatbot. In *International workshop on AI in education and educational research* (pp. 17-32). Cham: Springer Nature Switzerland.
- Garlikov, R. (2001). The Socratic method: Teaching by asking instead of by telling. Website, http://www.garlikov.com/Soc_Meth.html.
- Hashmi, S. F. A., & Rebello, N. S. (2025). Analyzing undergraduate problem-solving in physics through interaction with an AI chatbot. *arXiv preprint arXiv:2508.14778*.
- Hew, K. F., Huang, W., Wang, S., Luo, X., & Gonda, D. E. (2025). Towards a large-language-model-based chatbot system to automatically monitor student goal setting and planning in online learning. *Educational Technology & Society*, 28(3), 112-132.
- Jia, J., Wang, T., Zhang, Y., & Wang, G. (2024). The comparison of general tips for mathematical problem solving generated by generative AI with those generated by human teachers. *Asia Pacific Journal of Education*, 44(1), 8–28.
- Lee, J., Hung, J. T., Soylu, M. Y., Popescu, D., Cui, C. Z., Grigoryan, G., Joyner, D. A., & Harmon, S. W. (2025). Socratic Mind: Impact of a novel GenAI-powered assessment tool on student learning and higher-order thinking. *arXiv preprint arXiv:2509.16262*.
- Liu, Y., Li, C., Zhang, T., Wang, M., Zhu, Q., Li, J., & Huang, H. (2025). Discerning minds or generic tutors? Evaluating instructional guidance capabilities in Socratic LLMs. *arXiv preprint arXiv:2508.06583*.

- Ma, L. (1996). *Profound Understanding of Fundamental Mathematics: What is It, why is it Important, and how is it Attained?*. Stanford University.
- Moilanen, C. (1987). *Project SEED in the Portland Public Schools. Evaluation Report 1985-86.*
- Pintrich, P. R. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* National Center for Research to Improve Postsecondary Teaching and Learning.
- Puech, R., Macina, J., Chatain, J., Sachan, M., & Kapur, M. (2024). Towards the pedagogical steering of large language models for tutoring: A case study with modeling productive failure. *arXiv.*
- Rizopoulos, D. (2007). ltm: An R package for latent variable modeling and item response analysis. *Journal of statistical software*, 17, 1-25.
- 贾积有 & 张誉月.(2023).人工智能与教育:机遇、挑战与对策. *北京大学教育评论*(01),49-61+188-189.
- 贾积有, 乐惠骁, 张誉月, 刘怀亚, 陈昂轩 & 李姗姗. (2023). 基于大数据挖掘的智能评测和辅导系统设计. *中国电化教育* (03), 112-119.
- 贾积有,张誉月,刘怀亚 李双双.(2023).智能评测和辅导系统助力学生评价改革和减负增效.*电化教育研究*(06),74-80+89.
- 赵晓伟, 王师晓, & 李情. (2025). 苏格拉底式问题支架: 促进学生向 AI 大模型提出高质量问题. *现代远程教育研究*, 37(01), 102-112.