

大模型提问策略与提示语框架影响学习效果研究

钟柏昌 裴永漫

(华南师范大学 教育信息技术学院, 广东广州 510631)

[摘要] 以 DeepSeek 为代表的推理大模型的出现, 颠覆了人机对话的模式, 提示语框架的重要性式微。与之对应, 基于推理大模型的提问策略摒弃了程序化的提示词框架, 采用开放的人类自然语言交流, 一定程度上更有利于人机交互。本研究基于问题解决理论, 设计面向推理大模型的两阶段三环节提问策略。推理大模型的提问策略、提示词框架, 以及是否显示推理步骤, 可能在不同人群和学习阶段产生不同的效应。为此, 本研究以逆向工程教学法为例, 设计 2(两阶段三环节提问策略与 TIE 提示语框架)×2(推理可视化与非可视化推理)×2(要素增减阶段与结构创新阶段)的混合实验设计, 检验不同提问支架、推理呈现及学习阶段对学生学习效果的影响。对 168 名学生开展教学实验发现: 1) 要素增减阶段, 两阶段三环节提问策略与非可视化推理结合更利于提升学生的创新能力; 2) 在两个学习阶段中, 使用两阶段三环节提问策略更利于提高学生的工程知识水平; 3) 两阶段三环节提问策略与非可视化推理结合、TIE 提示语框架与可视化推理结合更利于提高学生的工作纸填写质量和迁移能力; 4) TIE 提示语框架与推理可视化结合更利于提升学生对生成式人工智能的使用态度。这一发现能为学生合理使用推理大模型, 培养相关能力提供实践指导。

[关键词] 两阶段三环节提问策略; TIE 提示语框架; 生成式人工智能; 推理可视化; 非可视化推理

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2025)06-0093-11

2025 年 1 月, 中国深度求索公司发布开源模型 DeepSeek-R1, 引发全球关注。DeepSeek 是专注复杂推理的大语言模型, 提供基础模型、深度推理 (R1)、联网搜索功能, 拥有“低成本、高性能、开源生态”的优势。它的出现标志着中国人工智能领域跳跃式发展(李丽, 2025)。值得关注的是, DeepSeek 凭借深度思考功能实现推理过程可视化突破, 将原本“黑箱式”的内容转化为“可追踪与可解释”的逐步推理路径, 显著增强了大模型的深度推理能力(沈阳等, 2025)。

然而, 推理大模型对教育教学的有效性多停留于理论猜想, 如郭蕾蕾(2025)探讨了 DeepSeek 推

理路径可视化效果, 但缺乏实证检验。鲜有研究比较“推理过程可视化”与“结果导向”“提问策略”“传统提示词工程”对学生学习效果的差异。本研究尝试总结相关提问策略, 探讨以上变量对学生学习效果的影响, 以期为推理大模型的教育教学应用提供参考。

一、文献综述

(一) 大语言模型推理可视化研究

大语言模型的推理可视化指将模型内隐的推理过程外显化, 从而增强模型的可解释性和透明性(沈阳等, 2025)。其中, 推理大模型的“深度思考”

[收稿日期] 2025-09-19

[修回日期] 2025-10-16

[DOI编码] 10.13966/j.cnki.kfjyyj.2025.06.010

[基金项目] 国家社科基金教育学重点项目“人工智能教育学研究”(编号: ACA250024)。

[作者简介] 钟柏昌, 博士, 教授, 博士生导师, 华南师范大学教育信息技术学院学术分委员会主任, 研究方向: 跨学科教育、人工智能教育研究(zhongbc@163.com); 裴永漫, 硕士研究生, 华南师范大学教育信息技术学院, 研究方向: 人工智能教育。

[引用信息] 钟柏昌, 裴永漫(2025). 大模型提问策略与提示语框架影响学习效果研究[J]. 开放教育研究, 31(6): 93-103.

功能代表推理可视化的突破性进展,实现全流程的自主推理。当用户提出“深度思考”要求时,推理大模型会呈现详细的思考与推理过程及推理依据,使大语言模型的回答更像是思维对话,推动大语言模型从“黑箱”变为可理解的“白盒”,让推理过程变得可追踪、可反思、可干预(陈智,2025)。

诸多学者高度评价推理大模型可视化的教学效果。例如,郭蕾蕾(2025)认为大模型通过推理路径可视化,可帮助学习者反思知识整合过程,发展元认知能力。萨利巴等(Saliba et al., 2025)认为大语言模型推理可视化可帮助学生使用模型解释复杂问题时识别推理错误,从而优化学习过程。徐政等(2025)认为大模型推理可视化可引导学习者深度思考,促进批判性思维和问题解决能力的培养。陈智(2025)认为获取完整推理链条比只看到最终答案更有价值,能帮助学习者理解思维推导的逻辑依据,培养更严谨、更客观、更全面的思维方式。

也有研究指出大语言模型推理可视化可能带来的弊端。例如,学习者易养成过度依赖技术获取答案的习惯,这种依赖性不仅会助长学习惰性,还导致学习主动性逐渐衰退(郜洁,2025)。推理可视化还可能带来信息过载,加重学习负担,知识的准确性也难以得到保证(Xu et al., 2025)。

上述梳理表明,推理大模型推理可视化有利于学习者厘清解题思路,也存在使学习者思维惰化的风险,然而鲜有研究者开展实证研究检验。

(二)大语言模型提示语框架的实证研究

越来越多的研究者将大语言模型应用于教育领域,检验其实践效果。其中,提示语框架的设计与应用因其基础性和关键作用,成为有效发挥大语言模型教育潜能的重要方向。此类研究可概括为两个方面:利用已有提示语框架和自主设计提示语框架开展教学实验。

部分学者利用已有提示语框架开展教学实验。陈颢鹏等(2023)提出的BROKE(Background-Role-Objective-Key Points- Expected Output)提示语框架受到广泛应用。赵培培(2025)通过对普通提示语与基于BROKE框架的提示语在文本分析、读中问题设计及读后活动设计的输出效果发现,BROKE提示语框架能显著提升生成式人工智能的输出质

量,促进学生深度学习和思维发展。陈小梅等(2025)以48名大一学生为研究对象,引入BROKE提示语框架设计学习提示单,发现其能显著提高学生学习成绩,能培养学生自主学习和创新能力。康等(Kang et al., 2025)采用BROKE提示语框架,并以高校研究生为研究对象开展教学实验表明,该框架能有效提升学生对生成内容的整合效率与关键信息的识别能力。王等(Wang et al., 2023)探讨了由OpenAI开发的CRISPE(Context - Role - Instruction - Style - Purpose - Expected Output)框架在输出方面的有效性,表明该框架能生成更完整、更深入的答案,提升学生获取信息的质量。

也有学者尝试自主设计提示语框架,并检验其成效。塔索蒂(Tassotti, 2024)设计5S提示语框架并将其应用于化学教育,结果显示采用结构化的提示语框架能显著提升学生对生成内容的感知能力,获得更准确的输出。钟柏昌等(2025)设计面向工程教育的TIE提示语框架,结果显示其能显著提升学生的创新能力、工程知识水平等。李希等(2025)围绕学习情况、学习要求和学习习惯设计提示指令,发现可以提升学生对知识的理解和运用能力、计算思维能力及自学能力等。此外,沈倩(2025)设计问题式、指令式和情景式提示语发现,提示语普遍能提高学生的阅读理解能力,提升学习体验感。

整体而言,已有研究肯定提示语框架的教学价值。同时,现有研究也存在不足:其一,大部分研究集中于提示语框架的理论探讨与设计,未开展实证研究以验证其应用效果;其二,少数研究尝试利用已有提示语框架开展实践教学,但实验周期短、实验设计与数据处理简单、结果讨论不深入,制约了研究的可靠性和说服力;其三,当前能自主设计提示语框架,开展实证研究检验其应用效果的研究过少,部分研究的提示语框架过于固定,缺乏灵活性与开放性。因此,如何重组和设计灵活开放的提示语,并检验其应用价值成为本文重点解决的问题。

(三)大语言模型的提问策略研究

推理大模型具备自动补全提示词、优化推理路径的能力。这种技术进步显著降低了人机交互对提示词工程的依赖,使用户能以更加自然和简洁的方式进行表达(沈阳等,2025)。因此,掌握向推理大模型提问的有效策略至关重要。本研究从提

问策略出发,尝试建立适用推理大模型的开放性提问策略。

学者针对大语言模型提问策略研究大致如下:一是从提问元素开展研究。何珊云等(2025)将表述明确、目的清晰作为有效的基础提问策略,并归纳出角色提问、材料提问和方案提问等要求。陈等(Chen et al., 2025)认为清晰而精确的指令对指导模型生成准确且相关的输出至关重要,建议将复杂的问题分解为系列简单的子问题。梅斯科(Meskó, 2023)建议向大语言模型提问要:问题尽可能清晰具体、设定切合实际的期望、提供示例、迭代优化等。坎因(Cain, 2024)提出有效的提示设计包括三个部分:全面的内容知识、批判性思维和迭代设计方法。博兹库尔特等(Bozkurt et al., 2023)认为问题清晰简洁、提供示例、指定输出、迭代优化、考虑伦理道德等策略有助于获取有意义且准确的回应。张等(Chang et al., 2025)通过提示语解决复杂问题以提高创造力,提出包含目标(goal)、提示(prompt)及策略(strategy)的GPS框架,其中策略包括分步骤提问、角色扮演、类比、强调、反思和自我优化等。此外,还有研究者认为使用反问、祈使等句式会极大影响大模型回应的风格和输出内容(Hwang et al., 2025)。

二是从提问阶段开展研究。许川根(2025)提出学习者与大语言模型对话的首问和追问策略:首问策略即阐明具体请求和信息,追问策略包括扩展追问、重复追问和重启追问等。赵晓伟等(2025)鼓励学生通过主动发问、递进追问、系统求问等策略与大模型对话。主动发问指鼓励学生向大模型提出多样化问题;递进追问指在获得大模型的响应后,针对已有问题、已解决方案提出后续问题。追问逻辑包括:将复杂问题分解为若干部分逐一探讨、对比追问、追问反事实观点、提供追问指南与公式。系统求问指引导学生发现衍生的新问题,构建新的问题架构。

整体而言,以上研究存在以下不足:一是大多数研究仅从理论层面探讨提问策略,未能系统考察以上策略教学应用及其对学生核心素养的影响;二是部分提问策略缺乏层次性和结构性,如各提问策略孤立分散,缺乏有效的逻辑关联和层次递进,未形成贯穿提问全过程的方法链,易使人机交互陷入

低效。为此,本研究尝试寻找逻辑主线,将提问策略有机串联,并验证其教育应用价值。

二、提问策略设计

与大模型交互,目的是通过不断提出问题和解决问题推动认知与能力的发展。理论研究普遍认为,问题解决是由一定情景引起的,需按照一定目标,应用各种认知活动、技能等,经过一系列的思维操作,使问题得以解决的过程(燕良轼, 2015)。

问题解决有两种代表性观点:一是桑代克的试误说,二是苛勒的顿悟说(代晔, 2021)。信息加工理论将问题解决看作信息加工的过程,将人看作信息加工者,认知过程就是对信息收集、加工处理,主张对问题进行程序化解决。该过程常分四个阶段:理解与表征问题、寻求解答策略、执行策略或尝试某种解答、评价结果(熊应等, 2019)。我国心理学界一般将问题解决的思维过程分为四个阶段:发现问题、明确问题、提出假设和检验假设(何犇等, 2020)。总体来看,无论是国外还是国内,大多对问题解决持“阶段论”观点,认为问题解决包括若干具有逻辑顺序的阶段与环节(莫雷, 2007)。

本研究尝试将学习者与大模型交互解决问题过程解构为三个环节:提出与表征问题、分析与讨论问题、获得方案与评价结果。在提出与表征问题环节,学习者需对问题进行结构化表征,即清晰描述问题、明确核心需求;在分析与讨论环节,通过运用多种提问策略,寻求可能的解决方案;在获得方案与评价结果环节,通过规范输出,获得理想结果,进而开展实践验证并反思优化方案。

(一) 不同环节的提问策略与整体框架

在提出与表征问题环节,精准定义问题是帮助学习者向推理大模型有效表征问题的关键。这要求学习者不仅要清晰准确地描述问题,更要深入把握问题的本质。只有精准定义问题,大语言模型才能理解任务需求,生成符合要求的输出,反之会产生宽泛或模棱两可的回答。因此,学习者需深入理解问题的背景、边界和核心诉求,并将模糊、零散的信息与想法转化为清晰、结构化的语言表达。这不仅能帮助学习者厘清问题的本质,而且有助于其与大语言模型有效沟通。值得注意的是,对初学者而言,精准定义问题往往是迭代优化的过程,学

习者可根据模型的反馈不断调整和优化问题表述,逐步提升提问的精准度。

在分析与讨论问题环节,问题拆解、深思互动和额外技巧是关键的提问策略。其中,问题拆解指将复杂问题分解为问题链(逻辑关联的一组子问题),然后按顺序求解,并将前一问题的结果作为后续问题的输入,系统整合各子问题的答案并形成最终答案。这种结构化的问题拆解方式能帮助学习者与大语言模型对话聚焦,避免因问题过于笼统而陷入信息过载的困境。

在完成问题拆解后,深思互动策略发挥着优化提问的作用。一方面,深思互动强调学习者以辩证思维审视大模型的输出内容,识别输出中可能存在的事实性错误、逻辑漏洞或潜在偏见,并针对性地调整提问策略和表述,持续提升输出质量。另一方面,批判性反思指鼓励大模型呈现支持与反驳性论据论点,即通过批判性反馈引导大模型进行批判性推理,优化输出内容。学习者尝试加入“请回答问题时加入批判性思考或进行多轮自我验证”等要求,促使大模型多角度推演和自我验证。

同时,额外技巧也是该环节的重要提问策略,能为深化问题理解助力。额外技巧指学习者使用反问、暗示或祈使等语句影响大模型的输出内容或风格。以反向提问(逆向逻辑或反向视角提出问题)为例,学生可向大模型询问“为什么我们不能采用这个方案”。学习者还可以使用追问指南和公式,将思考过程转化为逻辑表达,以获得尽可能多的解决方案。

最后,在获得方案与评价结果环节,学习者需明确输出格式、目标受众和应用范围等,并强调输出应符合伦理道德要求。值得一提的是,使用推理大模型要使学习者专注于领域知识和问题本身,而非提示词设计。因此,人机交互的最终目的在于建立类人际的对话,使学习者能以“精炼提问、深度交互”的方式与大模型沟通。

(二) 推理大模型提问策略与传统提示词工程的差异

本研究提出面向工程教育的TIE提示语框架,并通过实验证该框架的有效性(钟柏昌等,2025)。TIE提示语框架以传统ICIO (Input - Context - Instruction - Output)提示语框架为基础,融入

TRIZ(Theory of Inventive Problem Solving)等要素,其核心特征在于为学习者提供详细的操作指南和结构化的问题表述模板。与之相比,推理大模型的提问策略采用“启发而非限制”的设计理念,主要提供方向指引。该策略强调培养学习者自主提炼问题的能力,最终实现“精炼提问与深度交互”目标,实际指向的是学习者主体内的深度思考及其与人工智能(拟主体)的“人机共思”,体现更强的灵活性与开放性。

基于上述差异,本研究拟开展比较研究,探讨TIE提示语框架与推理大模型提问策略的差异。需说明的是,TIE提示语框架主要面向工程教育领域,因此本实验选择在工程教育情境下开展。工程教育是培育学生实践能力和创新能力的重要载体,生成式人工智能应用于工程教育主要集中于辅助工程知识学习和助力工程设计,提升学生的工程知识水平、工程设计能力和创新能力(钟柏昌等,2025)。同时,依据逆向工程教学法,本研究将实验过程分为要素增减和结构创新两个阶段,并在其中各嵌入三个问题解决环节,构建面向推理大模型的两阶段三环节提问策略(见图1)。

三、实验设计

(一) 实验目标与问题

基于以上分析,本实验构建了2(两阶段三环节提问策略与TIE提示语框架)×2(推理可视化与

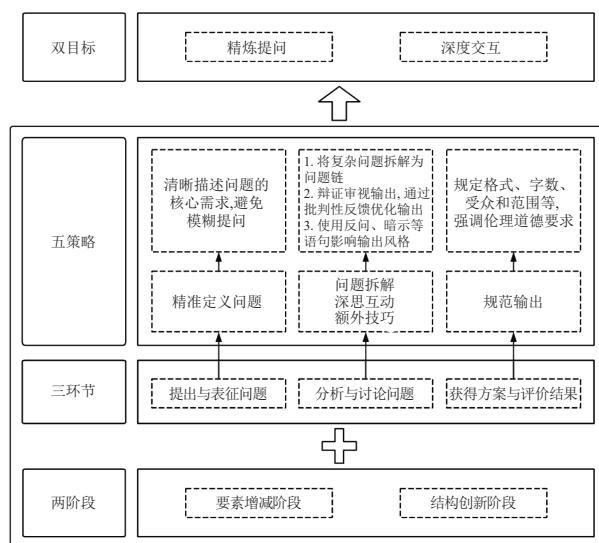


图1 两阶段三环节提问策略

非可视化推理)×2(要素增减与结构创新)的混合实验设计,讨论TIE提示语框架与开放式两阶段三环节提问策略在不同学习阶段与推理呈现下是否存在差异。参考前期研究,本研究从学生的创新能力、工程知识水平、工程设计能力、生成式人工智能使用态度及近迁移能力等维度评价学生学习效果。研究问题包括:

研究问题1:不同提问支架、推理呈现及学习阶段对学生创新能力有何影响?

研究问题2:不同提问支架、推理呈现及学习阶段对学生工程知识水平有何影响?

研究问题3:不同提问支架、推理呈现及学习阶段对学生工程设计能力有何影响?

研究问题4:不同提问支架、推理呈现及学习阶段对学生生成式人工智能使用态度有何影响?

研究问题5:不同提问支架、推理呈现及学习阶段对学生近迁移能力有何影响?

(二)实验对象

本实验于2025年春季学期在某中学以常规课程形式开展。研究人员从初一20个班级随机选择四个平行班级168名学生作为实验对象(见表1)。多数学生未学过工程设计课程,也没有用过生成式人工智能工具。前测显示,四组学生的创新能力($F(3, 164)=1.163, p=0.326>0.05, \eta^2=0.021$)、工程知识水平($F(3, 164)=1.791, p=0.151>0.05, \eta^2=0.032$)不存在显著差异,即四组被试的先验知识水平相似,不会对数据分析结果产生干扰与影响。

(三)实验处理

本研究采用混合实验设计。其中,提问支架和推理呈现作为组间变量,学习阶段为组内变量。研究人员根据分组条件将实验对象分配到4个实验条件组:分别是两阶段三环节提问策略+非可视化推理、两阶段三环节提问策略+推理可视化、TIE

提示语框架+非可视化推理、TIE提示语框架+推理可视化(见图2)。实验课时与上一轮实验相似,研究人员参考《义务教育科学课程标准(2022年版)》工程设计与物化模块、《普通高中通用技术课程标准》等新课标和新教材适当改编教学内容。教学过程中,研究人员鼓励学生两人自由配对完成任务。实验选用DeepSeek平台。

(四)测量工具

定量工具包括:1)创新能力问卷源自本研究编制的跨学科创新能力量表(钟柏昌等,2022),由32个李克特五级量表题项(“非常同意”至“非常不同意”)构成,量表信度良好(Cronbach's $\alpha=0.940$);2)工程知识水平测试包括桥梁基础知识、结构稳定性等;3)工程设计能力评分参照上一轮实验的评分标准;4)生成式人工智能使用态度问卷参考Ngo(2023)编制的量表,该量表Cronbach's α 系数0.856,信度较好;5)近迁移任务总分为100分,评分标准由研究团队编制,研究者和三位助教依据评分标准共同评分,取Kendall和谐系数最高的三位评分者的均值为最终分数。

定性工具包括:1)自我反思报告,旨在记录学生使用推理可视化的感悟,反思报告由使用该功能的学生完成,如需回答“推理可视化功能对你是否有帮助?并说明理由。”2)半结构访谈,旨在深入探究学生人机交互过程情况。访谈对象需回答“生成式人工智能对你的创新想法是否有帮助?你对生成式人工智能的评价如何?”等问题。

四、研究结果

本研究使用重复测量方差分析法(Repeated Measures ANOVA)分析实验数据。数据的球形性、方差齐性及正态分布检验满足该方法的要求。

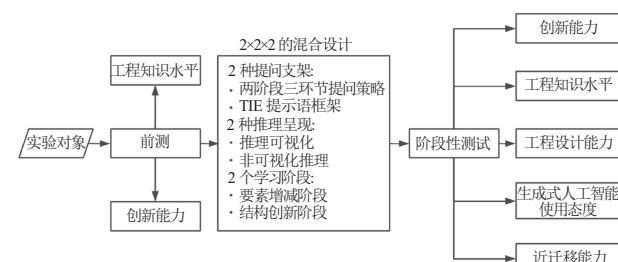


图2 实验设计

(一) 提问支架、推理呈现及学习阶段的交互效应对学生的创新能力产生显著影响

为考查不同提问支架、推理呈现及学习阶段对学生创新能力的影响,本研究以不同提问支架、推理呈现和学习阶段为自变量,以创新能力为因变量进行重复测量方差分析(见表2、3)。在创新能力上,学习阶段的主效应显著, $F(1, 164)=4.949, p=0.027, \eta^2=0.029$; 提问支架的主效应显著, $F(1, 164)=8.682, p=0.004, \eta^2=0.050$; 推理呈现的主效应显著, $F(1, 164)=4.630, p=0.033, \eta^2=0.027$ 。此外, 学习阶段 \times 提问支架 \times 推理呈现三因素交互作用显著 [$F(1, 164)=4.420, p=0.037, \eta^2=0.026$]。简单效应分析结果表明, 在任意两个学习阶段与推理可视化下, 使用两阶段三环节提问策略班级的创新能力与使用TIE提示语框架的班级不存在显著差异($p>0.05$); 在要素增减阶段与非可视化推理下, 使用两阶段三环节提问策略班级的创新能力显著高于使用TIE提示语框架的班级($p<0.05$); 在结构创新阶段与非可视化推理下, 使用两阶段三环节提问策略班级的创新能力与使用TIE提示语框架的班级不

表2 四个实验条件组创新能力分析

| 班级 | 提问支架 | 推理呈现 | 学习阶段 | 创新能力 ($M \pm SD$) |
|----|------------|------|------|------------------------|
| A | 两阶段三环节提问策略 | 非可视化 | 要素增减 | 118.38 \pm 25.92 |
| | | | 结构创新 | 115.48 \pm 26.71 |
| B | 两阶段三环节提问策略 | 可视化 | 要素增减 | 118.29 \pm 20.40 |
| | | | 结构创新 | 123.17 \pm 19.00 |
| C | TIE提示语框架 | 非可视化 | 要素增减 | 101.76 \pm 19.94 |
| | | | 结构创新 | 107.00 \pm 18.75 |
| D | TIE提示语框架 | 可视化 | 要素增减 | 112.79 \pm 21.98 |
| | | | 结构创新 | 115.98 \pm 22.77 |

表3 四个实验条件组创新能力后测重复测量结果

| 效应 | F | p | η^2 |
|----------------------------------|-------|---------|----------|
| 学习阶段 | 4.949 | 0.027* | 0.029 |
| 提问支架 | 8.682 | 0.004** | 0.050 |
| 推理呈现 | 4.630 | 0.033* | 0.027 |
| 学习阶段 \times 提问支架 | 1.903 | 0.170 | 0.011 |
| 学习阶段 \times 推理呈现 | 1.505 | 0.222 | 0.009 |
| 提问支架 \times 推理呈现 | 0.936 | 0.335 | 0.006 |
| 学习阶段 \times 提问支架 \times 推理呈现 | 4.420 | 0.037* | 0.026 |

注: * $p<0.05$, ** $p<0.01$, *** $p<0.001$ 。

存在显著差异($p>0.05$)。

(二) 提问支架和学习阶段的交互效应对学生的工程知识水平产生显著影响

本研究通过重复测量方差分析发现, 工程知识水平在学习阶段 [$F(1, 164)=45.282, p=0.000, \eta^2=0.216$]、提问支架 [$F(1, 164)=52.516, p=0.000, \eta^2=0.243$] 和推理呈现 [$F(1, 164)=5.031, p=0.026, \eta^2=0.030$] 的主效应均存在显著差异, 其中推理可视化优于非可视化推理。此外, 学习阶段和提问支架的交互作用显著 [$F(1, 164)=4.322, p=0.039, \eta^2=0.026$] (见表4、5)。简单效应分析显示, 在两个学习阶段中, 使用两阶段三环节提问策略班级的工程知识水平显著高于使用TIE提示语框架的班级($p<0.05$)(见图3)。

(三) 提问支架和推理呈现的交互效应对学生工作纸填写质量与近迁移能力产生显著影响

工程设计能力主要通过工作纸和作品体现。近迁移指学生完成工作纸上的问题后, 接着完成类似的近迁移任务。

表4 四个实验条件组工程知识水平分析

| 班级 | 提问支架 | 推理呈现 | 学习阶段 | 工程知识水平 ($M \pm SD$) |
|----|------------|------|------|--------------------------|
| A | 两阶段三环节提问策略 | 非可视化 | 要素增减 | 61.43 \pm 11.54 |
| | | | 结构创新 | 71.19 \pm 14.56 |
| B | 两阶段三环节提问策略 | 可视化 | 要素增减 | 68.21 \pm 8.18 |
| | | | 结构创新 | 77.62 \pm 7.67 |
| C | TIE提示语框架 | 非可视化 | 要素增减 | 53.21 \pm 14.22 |
| | | | 结构创新 | 59.88 \pm 15.04 |
| D | TIE提示语框架 | 可视化 | 要素增减 | 55.95 \pm 14.02 |
| | | | 结构创新 | 59.40 \pm 17.36 |

表5 四个实验条件组工程知识水平后测重复测量结果

| 效应 | F | p | η^2 |
|----------------------------------|--------|----------|----------|
| 学习阶段 | 45.282 | 0.000*** | 0.216 |
| 提问支架 | 52.516 | 0.000*** | 0.243 |
| 推理呈现 | 5.031 | 0.026* | 0.030 |
| 学习阶段 \times 提问支架 | 4.322 | 0.039* | 0.026 |
| 学习阶段 \times 推理呈现 | 0.673 | 0.413 | 0.004 |
| 提问支架 \times 推理呈现 | 2.520 | 0.114 | 0.015 |
| 学习阶段 \times 提问支架 \times 推理呈现 | 0.431 | 0.512 | 0.003 |

注: * $p<0.05$, ** $p<0.01$, *** $p<0.001$ 。

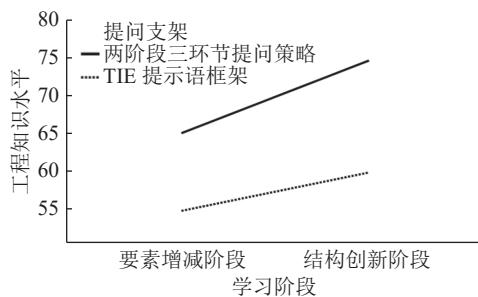


图3 不同学习阶段和提问支架的工程知识水平得分

1. 重复测量结果显示, 工作纸填写质量在学习阶段 [$F(1, 164)=37.761, p=0.000, \eta^2=0.321$]、提问支架 [$F(1, 164)=4.625, p=0.035, \eta^2=0.055$] 和推理呈现 [$F(1, 164)=5.427, p=0.022, \eta^2=0.064$] 的主效应存在显著差异, 其中结构创新阶段高于要素增减阶段。此外, 提问支架和推理呈现的交互作用显著 [$F(1, 164)=42.834, p=0.000, \eta^2=0.349$] (见表 6、7)。简单效应分析显示, 在推理可视化的情况下, 使用 TIE 提示语框架班级的工作纸填写质量显著高于使用

两阶段三环节提问策略的班级 ($p<0.05$) ; 在非可视化推理的情况下, 使用两阶段三环节提问策略班级的工作纸填写质量显著高于使用 TIE 提示语框架的班级 ($p<0.05$) (见图 4)。在作品上, 只有学习阶段的主效应存在显著差异 [$F(1, 164)=7.302, p=0.008, \eta^2=0.084$], 且要素增减阶段高于结构创新阶段。

2. 重复测量结果显示, 在近迁移能力上, 推理呈现 [$F(1, 164)=23.971, p=0.000, \eta^2=0.231$] 的主效应存在显著差异。此外, 提问支架和推理呈现的交互作用存在显著差异 [$F(1, 164)=62.616, p=0.000, \eta^2=0.439$] (见表 8、9)。简单效应分析表明, 在推理可视化情况下, 使用 TIE 提示语框架班级的近迁移能力显著高于使用两阶段三环节提问策略的班级 ($p<0.05$) ; 在非可视化推理情况下, 使用两阶段三环节提问策略班级的近迁移能力显著高于使用 TIE 提示语框架的班级 ($p<0.05$) (见图 5)。

(四) 提问支架和推理呈现的交互效应对学生生成式人工智能使用态度产生显著影响

结果显示, 在生成式人工智能使用态度上, 提问支架 [$F(1, 164)=4.558, p=0.034, \eta^2=0.027$] 和推理呈现 [$F(1, 164)=5.167, p=0.024, \eta^2=0.031$] 的主效应

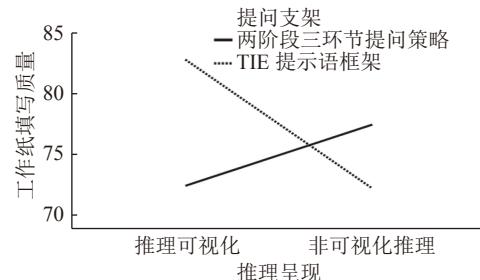


图4 不同提问支架和推理呈现方式的工作纸填写质量

表6 四个实验条件组工程设计能力分析

| 班级 | 提问支架 | 推理呈现 | 学习阶段 | 工作纸 ($M \pm SD$) | | 作品 ($M \pm SD$) | |
|----|------------|------|------|--------------------|----------|-------------------|------------|
| | | | | F | p | η^2 | |
| A | 两阶段三环节提问策略 | 非可视化 | 要素增减 | 75.14±6.45 | 0.000*** | 0.321 | 81.70±3.12 |
| | | | 结构创新 | 79.71±5.09 | 0.035* | 0.055 | 79.06±2.91 |
| B | 两阶段三环节提问策略 | 可视化 | 要素增减 | 70.97±9.06 | 0.022* | 0.064 | 81.59±2.70 |
| | | | 结构创新 | 73.79±5.86 | 0.009 | 0.113 | 81.60±4.75 |
| C | TIE 提示语框架 | 非可视化 | 要素增减 | 69.56±8.77 | 0.000*** | 0.099 | 80.86±5.79 |
| | | | 结构创新 | 74.78±5.00 | 0.008 | 0.045 | 79.59±2.41 |
| D | TIE 提示语框架 | 可视化 | 要素增减 | 80.49±5.42 | 0.000*** | 0.0349 | 81.27±3.03 |
| | | | 结构创新 | 85.10±2.88 | 0.000*** | 0.013 | 79.17±3.77 |

表7 四个实验条件组工程设计能力后测重复测量结果

| 效应 | 工作纸 | | | 作品 | | |
|----------------|--------|----------|----------|-------|--------|----------|
| | F | p | η^2 | F | p | η^2 |
| 学习阶段 | 37.761 | 0.000*** | 0.321 | 7.302 | 0.008* | 0.084 |
| 提问支架 | 4.625 | 0.035* | 0.055 | 1.663 | 0.201 | 0.020 |
| 推理呈现 | 5.427 | 0.022* | 0.064 | 1.045 | 0.310 | 0.013 |
| 学习阶段×提问支架 | 0.751 | 0.389 | 0.009 | 0.113 | 0.737 | 0.001 |
| 学习阶段×推理呈现 | 0.712 | 0.401 | 0.009 | 0.679 | 0.412 | 0.008 |
| 提问支架×推理呈现 | 42.834 | 0.000*** | 0.349 | 1.045 | 0.310 | 0.013 |
| 学习阶段×提问支架×推理呈现 | 0.162 | 0.689 | 0.002 | 2.464 | 0.120 | 0.030 |

注: * $p<0.05$, ** $p<0.01$, *** $p<0.001$ 。

表8 四个实验条件组近迁移能力分析

| 班级 | 提问支架 | 推理呈现 | 学习阶段 | 近迁移能力 ($M \pm SD$) |
|----|------------|------|------|----------------------|
| A | 两阶段三环节提问策略 | 非可视化 | 要素增减 | 77.54±3.52 |
| | | | 结构创新 | 75.73±3.99 |
| B | 两阶段三环节提问策略 | 可视化 | 要素增减 | 73.75±6.14 |
| | | | 结构创新 | 74.03±4.26 |
| C | TIE 提示语框架 | 非可视化 | 要素增减 | 70.71±7.42 |
| | | | 结构创新 | 69.32±5.44 |
| D | TIE 提示语框架 | 可视化 | 要素增减 | 81.29±3.24 |
| | | | 结构创新 | 82.06±4.00 |

表9 四个实验条件组近迁移能力后测重复测量结果

| 效应 | F | p | η^2 |
|----------------|--------|----------|----------|
| 学习阶段 | 0.859 | 0.357 | 0.011 |
| 提问支架 | 0.411 | 0.523 | 0.005 |
| 推理呈现 | 23.971 | 0.000*** | 0.231 |
| 学习阶段×提问支架 | 0.153 | 0.697 | 0.002 |
| 学习阶段×推理呈现 | 3.409 | 0.069 | 0.041 |
| 提问支架×推理呈现 | 62.616 | 0.000*** | 0.439 |
| 学习阶段×提问支架×推理呈现 | 0.001 | 0.973 | 0.000 |

注: * $p<0.05$, ** $p<0.01$, *** $p<0.001$ 。

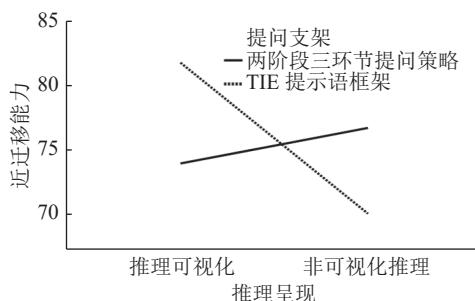


图5 不同提问支架和推理呈现方式近迁移能力

值存在显著差异。此外,提问支架和推理呈现的交互作用显著 [$F(1, 164)=4.858, p=0.029, \eta^2=0.029$] (见表 10、11)。简单效应分析显示,在推理可视化情况下,使用 TIE 提示语框架班级的生成式人工智能使用态度显著优于使用两阶段三环节提问策略的班级($p<0.05$);在非可视化推理下,两种提问支架下生成式人工智能使用态度不存在显著差异($p=0.961$)(见图 6)。

表10 四个实验条件组生成式人工智能使用态度分析

| 班级 | 提问支架 | 推理呈现 | 学习阶段 | 生成式人工智能使用态度 ($M \pm SD$) |
|----|------------|------|------|----------------------------|
| A | 两阶段三环节提问策略 | 非可视化 | 要素增减 | 70.86±14.38 |
| | | | 结构创新 | 68.93±11.70 |
| B | 两阶段三环节提问策略 | 可视化 | 要素增减 | 69.45±6.70 |
| | | | 结构创新 | 70.55±9.27 |
| C | TIE 提示语框架 | 非可视化 | 要素增减 | 70.79±11.79 |
| | | | 结构创新 | 68.79±10.86 |
| D | TIE 提示语框架 | 可视化 | 要素增减 | 75.74±13.85 |
| | | | 结构创新 | 77.71±12.70 |

表11 四个实验条件组生成式人工智能使用态度中后测重复测量结果

| 效应 | F | p | η^2 |
|----------------|-------|--------|----------|
| 学习阶段 | 0.056 | 0.814 | 0.000 |
| 提问支架 | 4.558 | 0.034* | 0.027 |
| 推理呈现 | 5.167 | 0.024* | 0.031 |
| 学习阶段×提问支架 | 0.050 | 0.824 | 0.000 |
| 学习阶段×推理呈现 | 3.712 | 0.056 | 0.022 |
| 提问支架×推理呈现 | 4.858 | 0.029* | 0.029 |
| 学习阶段×提问支架×推理呈现 | 0.069 | 0.794 | 0.000 |

注: * $p<0.05$, ** $p<0.01$, *** $p<0.001$ 。

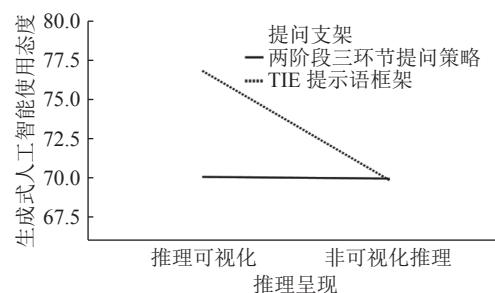


图6 不同提问支架和推理呈现方式的生成式人工智能使用态度

五、研究发现与讨论

(一) 提问支架、推理呈现及学习阶段的交互效应能促进学生创新能力培养

实验结果显示,提问支架、推理呈现及学习阶段的交互效应存在显著差异。在要素增减阶段与非可视化推理下,使用两阶段三环节提问策略班级的创新能力显著高于使用 TIE 提示语框架的班级;在结构创新阶段与非可视化推理下,使用两阶段三环节提问策略班级的创新能力与使用 TIE 提示语框架的班级不存在显著差异。原因可能是:一方面, TIE 提示语框架虽然整合了 TRIZ 理论,但推理大模型在非可视化推理下提供的理论解释较简略,因此其庞杂的体系对学生初期学习而言存在较高认知门槛(闫妮等, 2018)。随着学习的深入,该认知难度才逐渐降低。学生也提到:“使用 TIE 提示语框架向推理大模型提问,虽然获得很多创新方案的启发,但初期对 TRIZ 理论了解不多,部分内容理解困难。随着学习和实践的深入,我才逐步掌握 TRIZ 理论的原理方法,能比较好地运用这些原理

设计方案。”另一方面,两阶段三环节提问策略学习难度低、限制少,更有利于提出多样化问题,而这对创造性思维和解决问题能力的培养至关重要(Raz et al., 2024)。学生也提到:“两阶段三环节提问策略给予我更多提问上的自由,所以我的主动性和积极性较高,也能发散出很多想法。”

(二) 提问支架和学习阶段的交互效应能提高学生的工程知识水平

结果显示,推理呈现的主效应对工程知识水平产生显著差异,且推理可视化优于非可视化推理。已有研究表明,推理可视化可帮助学习者反思知识整合过程,发展元认知能力(郭蕾蕾,2025)。此外,在两个学习阶段中,使用两阶段三环节提问策略班级的工程知识水平显著好于使用TIE提示语框架的班级。原因可能是:认知负荷超载使得学习者有限的认知资源无法深度加工学习材料,进而影响学习者知识建构(高楠等,2023)。课堂观察也表明,使用TIE提示语框架的学生需花更多时间理解该框架的内涵、机理,认知负荷较重,影响知识建构和学习。

(三) 提问支架和推理呈现的交互效应能提升学生的工作纸填写质量与近迁移能力

结果显示,学习阶段的主效应对工作纸填写质量产生显著差异,且结构创新阶段高于要素增减阶段。学生表示:“经过要素增减阶段的学习,我对方案的设计、回答工作纸上的问题越来越得心应手。”换言之,阶段性的学习和反复练习,会对学生工作纸填写产生影响。需深入探究的是,在推理可视化情况下,使用TIE提示语框架班级的工作纸填写质量、近迁移能力显著高于使用两阶段三环节提问策略的班级;在非可视化推理情况下结果则相反。究其原因,利用推理大模型的分析和生成能力来解释和增强TRIZ,并将二者应用到方案设计中,能推动设计构思的创新(Lee et al., 2024),且可视化推理过程能降低学习的认知负荷,促进知识迁移(骆飞等,2025)。因此,结合TIE提示语框架与推理可视化能更好地帮助学生进行方案设计和知识迁移。此外,作品在学习阶段的主效应上存在显著差异,且要素增减阶段显著高于结构创新阶段。原因可能是作品制作更依靠学生的实践操作,而要素增减阶段的作品仅是微创新,要求较低;结构创

新阶段要求制作新作品,要求较高。学生表示:“推理大模型虽然给我们提供了很多创新方案,但真正作品制作得靠自己,且结构创新阶段的作品制作要求显著高于要素增减阶段。”

(四) 提问支架和推理呈现的交互效应能提升学生成式人工智能使用态度

研究结果表明,在推理可视化下,使用TIE提示语框架的学生成式人工智能使用态度显著优于使用两阶段三环节提问策略的学生;在非可视化推理下,两种提问支架下的生成式人工智能使用态度不存在显著差异。究其原因,推理大模型具备可视化推理能力,学生可直观理解大模型的思考过程,提升易用性感知(骆飞等,2025)。同时,TIE提示语框架内含复杂的TRIZ理论,在推理可视化辅助下能为学生提供清晰的逻辑引导,促进深度认知加工。使用TIE提示语框架的学生提到:“推理可视化功能对TRIZ理论进行了较全面解析,使我们能更好地理解其生成的解决方案,显著提升了使用体验。”相比之下,两阶段三环节提问策略可能因开放性高,削弱了部分学生对生成式人工智能的使用态度。

六、结语

本研究结果对于推理大模型教学应用有以下启示:首先,有效的提示设计常需在规范性和开放性之间取得平衡。过于具体的提示可能限制生成内容的多样性和创造性,而过于开放的提示可能导致偏离主题或输出缺乏针对性;其次,TIE提示语框架对中学生而言有一定难度,将其与推理可视化结合或考虑开发更具专业性的教育智能体可在一定程度上解决该问题;最后,推理大模型在促进人类思维发展等方面表现出巨大潜力,但人类仍是决策的主体,利用生成式人工智能辅助和启迪人类思维与想法的生成,实现人机共智才是可行之道。

本研究不足之处在于:其一,学习者虽采用自由配对的方式完成任务,但实际配对以同性别为主,异性组合较少。后续研究可将不同性别组合作为变量,探究其对学习效果的影响。其二,实验对象主要为初一学生,学段单一。后续研究可纳入不同学段的学习者,验证结论的普适性。

[参考文献]

- [1] Bozkurt, A., & Sharma, R. C.(2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world[J]. *Asian Journal of Distance Education*, 18(2): i-vii.
- [2] Cain, W.(2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education[J]. *TechTrends*, 68(1): 47-57.
- [3] Chang, H. F., & Li, T.(2025). A framework for collaborating a large language model tool in brainstorming for triggering creative thoughts[J]. *Thinking Skills and Creativity*, 56(2): 101755.
- [4] Chen, B., Zhang, Z., Langrené, N., & Zhu, S.(2025). Unleashing the potential of prompt engineering for large language models[J]. *Patterns*, 6(6): 101260.
- [5] 陈颢鹏,李子菡(2023). ChatGPT进阶:提示工程入门 [M]. 北京:北京大学出版社: 101-102
- [6] 陈小梅,谢姝睿,朱敏捷,谢雅淇(2025). 基于GAI提示工程的学习提示单设计及应用 [J]. 教育信息技术, (Z1): 33-36.
- [7] 陈智(2025). 新一代生成式人工智能的价值定位、演进逻辑及前瞻路径——以DeepSeek为例 [J]. 新疆社会科学, (2): 17-29+170.
- [8] 代晔(2021). 基于化学学科核心素养的问题解决教学理论与实践研究 [D]. 内蒙古师范大学.
- [9] 高楠,刘革平,谢涛,张领(2023). 沉浸式虚拟现实对学习者知识迁移效果及效率的影响研究 [J]. 远程教育杂志, 41(1): 91-101.
- [10] 郁洁(2025). 生成式人工智能的双刃剑效应——DeepSeek在外语教育领域的应用优势与潜在风险探析 [J]. 当代外语研究, (3): 140-151.
- [11] 郭蕾蕾(2025). 生成式人工智能驱动教育变革:机制、风险及应对——以DeepSeek为例 [J]. 重庆高教研究, 13(3): 38-47.
- [12] Hwang, M., Lee, K. H., & Lee, H. K.(2025). A word to the wise: Crafting impactful prompts for ChatGPT[J]. *System*, 133(6) : 103756.
- [13] 何犇,李艺,毛文林(2020). 教育心理学 [M]. 成都: 电子科技大学出版社: 72-73.
- [14] 何珊云,沈演(2025). 学会提问:大学生与生成式人工智能协同学习模式的研究 [J]. 华东师范大学学报(教育科学版), 43(2): 34-48.
- [15] Kang, L. , Shi, X. , & Zhu, K. (2025). Uncovering the mediation of disciplinary literacy in the effect of GAI prompt engineering on pre-service teachers' instructional design[J]. *Education and Information Technologies*. <https://link.springer.com/article/10.1007/s10639-025-1365-7-8>.
- [16] Lee, C. K. M., Liang, J., Yung, K. L., & Keung, K. L.(2024). Generating TRIZ-inspired guidelines for eco-design using Generative Artificial Intelligence[J]. *Advanced Engineering Informatics*, 62(3) : 102846.
- [17] 李丽(2025). 大语言模型视角下DeepSeek赋能高校图书馆学科服务研究 [J]. 图书馆建设, (4): 75-83+170.
- [18] 李希,刘喜平,戴航,姜火文,赵芸,舒晴(2025). 提示工程应用于数据结构课程教学的创新模式设计与实践 [J]. 计算机教育, (4): 131-136.
- [19] 骆飞,马雨璇,焦丽珍(2025). 迭代驱动转型: DeepSeek的技术特质如何赋能教育数字化转型 [J]. 苏州大学学报(教育科学版), 13(3): 28-39.
- [20] Meskó, B.(2023). Prompt engineering as an important emerging skill for medical professionals: tutorial[J]. *Journal of medical Internet research*, 25: e50638.
- [21] 莫雷(2007). 教育心理学 [M]. 北京: 教育科学出版社: 90-92.
- [22] NGO, T. T. A.(2023). The perception by university students of the use of ChatGPT in education[J]. *International journal of emerging technologies in learning*, 18(17): 4-19.
- [23] Raz, T., Reiter-Palmon, R., & Kenett, Y. N.(2024). Open and closed-ended problem solving in humans and AI: the influence of question asking complexity[J]. *Thinking Skills and Creativity*, 53(3): 101598.
- [24] Saliba, T., Ferrari, J., Pozzessere, C., Rotzinger, D., & Fahrni, G.(2025). Can advanced large language models support radiology training? A performance assessment of DeepSeek R1[J]. *European Journal of Radiology Artificial Intelligence*, 3: 100024.
- [25] 沈倩(2025). 生成式人工智能提示语对英语阅读理解能力差异性影响的实证研究 [J]. 现代职业教育, (14): 117-120.
- [26] 沈阳,闫佳琦,邹开元(2025). 基于DeepSeek思维链的人机协同知识生产 [J]. 中国编辑, (4): 95-100.
- [27] Tassotti S.(2024). Assessment of students use of generative artificial intelligence: Prompting strategies and prompt engineering in chemistry education[J]. *Journal of Chemical Education*, 101(6) : 2475-2482.
- [28] Wang, M., Wang, M., Xu, X., Yang, L., Cai, D., & Yin, M.(2023). Unleashing ChatGPT's power: A case study on optimizing information retrieval in flipped classrooms via prompt engineering[J]. *IEEE Transactions on Learning Technologies*, 17: 629-641.
- [29] Xu, Z. , Liu, S. , Huang, W. , Shi, J. , Xu, F. , Tang, X. , & Lu, H. (2025). DeepSeek: Implications for data science and management in the AI era[J]. *Data Science and Management*. <https://www.sciencedirect.com/science/article/pii/S2666764925000451>
- [30] 熊应,罗璇,谢园梅(2019). 教育心理学 [M]. 长沙: 湖南师范大学出版社: 195-197.
- [31] 徐政,邱世琛,葛力铭(2025). DeepSeek赋能拔尖创新人才培养的理论逻辑与实践路径 [J]. 重庆大学学报(社会科学版), 31(3): 118-129.
- [32] 许川根(2025). 基于生成式人工智能的外语学习者人机协同写作投入研究 [J]. 外语与外语教学, (2): 61-73+147.
- [33] 闫妮,钟柏昌(2018). 中小学机器人教育的核心理论研究——论发明创造型教学模式 [J]. 电化教育研究, 39(4): 66-72.
- [34] 燕良轼(2015). 高等教育心理学 [M]. 长沙: 湖南师范大学出版社: 135, 138.
- [35] 赵培培(2025). 生成式人工智能基于BROKE框架的提示语在初中英语阅读教学中的应用研究 [J]. 新课程导学, (21): 33-36.
- [36] 赵晓伟,沈书生,祝智庭(2025). 与AI大模型对话促进深度思维: 提问设计的科学与艺术 [J]. 中国教育学刊, (4): 23-29.
- [37] 钟柏昌,龚佳欣(2022). 跨学科创新能力评价指标体系的构

建与实证研究 [J]. 中国电化教育, (12): 27-34.

设计与应用研究 [J]. 电化教育研究, 46(4): 40-48.

[38] 钟柏昌,裴永漫(2025). 面向工程教育的TIE提示语框架的

(编辑:赵晓丽)

Research on the Impact of Questioning Strategies and Prompt Frameworks for Large Inference Models on Learning Outcomes

ZHONG Baichang & PEI Yongman

(School of Information Technology in Education, South China Normal University, Guangzhou 510631, China)

Abstract: The large reasoning models represented by DeepSeek have revolutionized the mode of human-machine dialogue, reducing the significance of prompt frameworks. Consequentially, the question-prompt strategies based on large reasoning models have abandoned the procedural prompt framework and adopted an open human natural language communication style, which is to some extent more conducive to promoting human-machine interaction. However, the question-prompt strategy, the prompt framework, as well as whether to display the reasoning steps, may have different effects on different groups of people and learning stages. Thus, this study takes the reverse engineering teaching method as an example and designs a 2(two-stage three-step question-asking strategy VS TIE prompt framework) \times 2(visualized reasoning VS non-visualized reasoning) \times 2(element addition and subtraction stage VS structural innovation stage) mixed experimental design to examine the impact of different question-prompt scaffolds, reasoning presentation, and learning stages on students' learning outcomes. With 168 middle school students' participation, the study found that 1) In the element addition and subtraction stage, the combination of the two-stage three-step question-prompt strategy and non-visualized reasoning is more conducive to enhancing students' innovation ability; 2) In both learning stages, the use of the two-stage three-step question-prompt strategy is more conducive to improving students' engineering knowledge level; 3) The combination of the two-stage three-step question-prompt strategy and non-visualized reasoning, as well as the combination of the TIE prompt framework and visualized reasoning, is more conducive to improving students' worksheet scores and near transfer ability; 4) The combination of the TIE prompt framework and visualized reasoning is more conducive to enhancing students' attitude towards the use of GenAI. The study provides effective practical guidance for students to use large reasoning models reasonably and cultivate relevant abilities.

Key words: two-stage three-step questioning strategy; TIE prompt framework; generative artificial intelligence; reasoning visualization; non-visual reasoning