

生成式 AI 融入课堂的过程证据框架构建与实践改革*

马瑞强* *

内蒙古工业大学智能科学与技术学院,
呼和浩特 010051

明治大学研究战略机构, 日本东京 164-8525

王迪 蒋永生 徐荣

内蒙古工业大学智能科学与技术学院,
呼和浩特 010051

摘要 针对生成式 AI 导致本科生作业结果与真实理解不符的现状, 本研究以《计算机网络》课程的 70 名本科生为研究对象, 构建并实施课堂作业过程证据框架。在研究思路, 不以禁用人工智能 (AI) 为目标, 而以 AI 可用、学习、可证为原则, 把可观察、可核验、可追溯的学习过程信息作为评价核心。框架以任务链和证据链为主线, 贯穿课前准备、课中产出和课后分层的三级目标任务。通过思路形成、AI 交互修正、课堂参与与理解表达四类证据实现学习过程显性化。本研究的创新点在于设计最小充分证据包和四维度量规范, 配合抽样口头核验与随堂小测试的一致性检验, 实现了在大班授课条件下的兼顾真实性评价与教师工作量可控。改革成效表明, 学生参与度与学习真实性提升明显, AI 直接生成比例大幅下降, 平均成绩从 70.1 分提升至 85.5 分, 高阶思维能力同步改善。本研究为生成式 AI 时代本科生课程的作业与评价改革提供了可复制的实施路径。

关键字 生成式 AI; 过程证据; 课堂作业重构; 形成性评价; 真实参与

*

Building a Process Evidence Framework for Generative AI-Integrated Classrooms

Ruiqiang Ma* *

College of Intelligent Science and Technology
Inner Mongolia University of Technology,
Hohhot 010051, China;
Research on Financial Strategic Institutions
Meiji University,
Tokyo 164-8525, Japan.

Di Wang Yongsheng Jiang Rong Xu

College of Intelligent Science and Technology
Inner Mongolia University of Technology,
Hohhot 010051, China;

Abstract—Addressing the issue that generative AI causes undergraduate homework outcomes to diverge from genuine understanding, this study takes 70 undergraduates in the Computer Networks course as the research subjects, constructing and implementing a classroom assignment process evidence framework. In terms of research approach, the goal is not to ban AI but to adhere to the principle of "AI usable, learning provable," making observable, verifiable, and traceable learning process information the core of evaluation. The framework takes the "task chain and evidence chain" as its main line, running through three levels of hierarchical tasks: pre-class preparation, in-class production, and post-class assignments. It makes the learning process explicit through four types of evidence: idea formation, AI interaction and revision, classroom participation, and comprehension expression. The innovation of this study lies in designing a minimum sufficient evidence package and a four-dimensional rubric, combined with consistency checks through sampling oral verification and in-class mini-tests, achieving both authentic assessment and manageable teacher workload under large-class theoretical course conditions. Reform results show a significant improvement in student participation and learning authenticity, a substantial decrease in the proportion of AI-generated submissions, an increase in average scores from 70.1 to 85.5, and synchronous improvement in higher-order thinking skills. This study provides a replicable implementation path for reforming assignments and assessments in undergraduate courses in the era of generative AI.

Keywords—Generative AI, Process Evidence, Classroom Assignment Redesign, Formative Assessment, Authentic Engagement

1 引言

1.1 研究背景: 生成式 AI 对传统作业与课堂参与的冲击

*基金资助: 本文得到国家人社部外国专家重点支撑计划 (D20250092) 和省部重点研发 (2025YFSH0157) 的资助。

*通讯作者: 马瑞强 marq@imut.edu.cn。

生成式 AI 的普及对传统作业评价方式带来了显著冲击[1][2]。在这种背景下, 学术界开始反思这是否意味着传统在线考试诚信的终结[3], 甚至有学者激进地讨论 ChatGPT 是否仅是“胡言乱语者”或是传统高等教育评价的终点[4][5]。本实践围绕生成式 AI 进入课堂后“作业结果不再等价于真实学习”的核心矛

盾展开,聚焦《计算机网络》这类理论性强、概念抽象、推理链长的课程情境:学生可快速获得高质量作业成品,导致课堂参与被弱化、机制性理解难以外显、教师反馈依据不足。为此,论文提出课堂作业“过程证据”框架,不以禁用 AI 为目标,而以“AI 可用但学习必须可证”为原则,把学习过程中的可观察、可核验、可追溯信息纳入作业设计与评价核心。框架采用“作业链+证据链”的耦合思路,以课前微准备、课中即时产出、课后分层任务组织学习,并用思路形成、AI 交互修正、课堂参与、理解表达四类证据贯穿全过程,配套最小充分证据包、过程性量规评价与抽样口头核验和随堂微测的一致性检验,在大班条件下兼顾真实性约束与工作量可控。

论文结合 TCP 连接建立、DNS 与 HTTP、路由与子网划分等主题给出可复制的任务样例与课堂组织流程,说明如何通过随机展示、同伴互评与教师抽检形成快速反馈闭环,促进学生从“生成答案”转向“验证与解释”。最后提出面向 70 人规模理论课的评估方案,从参与行为、作业真实性与一致性、学习成效与学习体验等维度呈现改革效果,归纳常见问题与负担控制策略,并讨论可推广性、隐私与规范边界,给出生成式 AI 时代网络类基础课程作业设计与评价改革的落地路径。

1.2 问题提出:结果导向评价的失真与参与不可见。

当前学生完成作业的路径已发生质变,这对课程教学设计提出了新的分析需求[1]。同时由于 AI 可能存在的算法偏见[6],单纯依赖 AI 进行评价仍存在风险,因此有必要探讨其对教育政策和实践的深远影响[7]。生成式 AI 将信息检索、文本生成以及代码与推理输出集成到低门槛的对话式工具之中,使学生完成作业的首要策略逐渐由“阅读教材、推导思考、独立书写”转向“提出问题、获取生成、再做润色”。这一变化的影响并不限于学术诚信,更深层的挑战在于它动摇了传统作业评价的基本假设。长期以来,作业评价往往以最终答案或报告作为主要依据,默认结果质量能够代表学习质量。当 AI 能够稳定产出高质量文本时,结果导向评价就可能发生失真:表面同样优秀的作业成品背后,学生的学习投入、理解深度与能力发展轨迹可能差异显著,教师也难以仅凭成品判断其真实掌握程度,从而使“参与”和“理解”的关键过程在评价体系变得不可见。

这一矛盾在《计算机网络》课程中尤为突出。作为计算机类核心基础课,本课程知识体系既包含抽象概念,如分层思想与协议栈服务关系,也包含强调机理与约束条件的机制性知识,如拥塞控制、可靠传输与路由选择,同时还要求学生在不同层之间建立因果解释并完成综合推理,例如依据抓包现象解释 TCP 状态变化与重传行为。课程学习天然依赖过程性理解,学

生不仅要知道“是什么”,更要能够解释“为什么如此”以及“在什么条件下会发生变化”。在生成式 AI 广泛使用的背景下,如果作业仍主要考察定义描述、流程复述与标准答案型计算,就更容易被 AI 替代,课堂讨论与即时思考被弱化,推理表达与机制解释难以充分展开,进而削弱学生的真实参与与深度学习。

1.3 研究目标与创新点

本文旨在构建并在《计算机网络》大班理论课情境中验证一套课堂作业“过程证据”框架,使生成式 AI 可被合理使用,同时确保学习过程可见、可核验、可追溯。具体而言,框架要求学生在借助 AI 完成任务时,不仅提交最终答案,更必须呈现思路形成的关键节点,展示对 AI 输出的证据对照与有效验证,并记录必要的修正迭代,从而将“生成”转化为“理解与解释”[17]。在教师侧,框架通过最小充分证据包、量规化评价与抽样核验等机制,帮助教师在工作量可控的前提下获得高质量过程数据,用于开展形成性评价与精准反馈,及时识别概念混淆与推理断裂等学习困难。最终,框架将课前准备、课中即时产出与课后分层作业贯通起来,使课堂参与与学习产出形成闭环,以过程证据持续驱动教学改进,实现从结果导向到过程导向的作业与评价重构。

2 理论基础与相关研究综述

2.1 形成性评价与学习证据观

形成性评价强调评价应当服务于学习,通过及时反馈帮助学生在过程中持续修正理解与策略,进而实现能力的渐进发展。学习证据观进一步指出,学习成效不应仅依赖一次性考试来体现,而应通过多源证据加以呈现,并与学习目标保持一致。这些证据既包括学习草稿与阶段性产出,也包括思路迭代、课堂讨论、解释表达与迁移应用等能够反映认知变化的过程材料。在生成式 AI 广泛应用的背景下,这一视角的重要性更加凸显。当标准答案更容易被获得时,真正能够区分学习质量的往往不是答案本身,而是学生如何形成答案、如何验证与修正,以及能否将答案转化为可解释、可迁移的理解与能力。本文提出的‘过程证据’框架深受综合素质评价理论的影响,强调评价应服务于学习过程^[8]。这与近期关于 AI 赋能沉浸式学习机理的研究相契合,旨在通过新模式提升学习深度^[9]。

2.2 生成式 AI 时代作业设计的转向

教学设计正从传统范式向 AI 赋能范式转型[10],核心在于如何通过评价机制培养学生的高阶思维能力[11]。此外,不同高校学生在 AI 素养上的差异也提醒我们,评价框架应具备普适性与分层导向[12]。近年来的教学探索大体沿着两条路径推进:一条路径强调制定 AI 使用规范与诚信机制,以减少不当使用带来的

风险；另一条路径则聚焦作业与评价本身的重构，通过引入真实情境任务、反思性任务与过程性评价，把学习的关键过程重新纳入课程要求与评价体系。本文以第二条路径作为主线，不将 AI 简单视为风险源，而把它视为能够放大学习效率与认知资源的工具，同时通过任务设计与评价机制引导学生对 AI 输出进行对照验证与解释复盘，使学生在使用 AI 的过程中仍需完成思考、判断与修正，从而把生成行为转化为学习过程，把外部产出转化为内化理解。

2.3 过程证据的概念界定与原则

本文所指的过程证据，是学生在完成学习任务时产生的、能够反映其认知过程与参与行为的可观察信息集合，其核心价值在于让学习过程可见、可评价、可反馈。过程证据应满足以下原则：第一，可观察，能够通过课堂活动、线上平台或提交材料被教师识别与采集；第二，可核验，能够通过追问、随堂小测或对照检验等方式验证其一致性与真实性；第三，可追溯，能够呈现从初稿到定稿的关键变化以及变化背后的理由与依据；第四，低负担，对学生与教师均具备可规模化实施的可行性，不因证据要求而显著增加无效工作量；第五，对齐目标，证据项与课程核心能力相对应，能够支撑对概念理解、机制解释、问题分析、表达复盘与反思迁移等能力的评价与改进。

3 课程情境与问题诊断

本课程为《计算机网络》，面向本学院大学二年级的两班合班授课，总人数约 70 人。教学以理论讲授为主，辅以少量课堂演示，包括抓包示例、网络拓扑示意和协议流程图，用于帮助学生建立直观理解。学生基础差异明显，有的具备一定编程与实验经验，有的对网络协议缺乏直接体验；同时受课时与条件限制，课堂难以开展大规模上机与持续性实验活动，这使得课程对课堂讲解的组织质量、作业设计的引导性以及反馈的及时性提出更高要求。

在作业方面，传统作业主要由概念问答、流程描述与计算题构成，评价依据多集中于最终文本与标准答案。生成式 AI 普及后，传统作业暴露出更明显的失效风险：一是文本同质化加剧，作业答案结构趋同、表述雷同，难以区分学生个体差异与真实投入；二是错误更容易被隐藏，文本表面专业但关键逻辑可能存在偏差，例如将流量控制与拥塞控制混淆，教师若仅审阅成品不易及时发现；三是学习过程不可见，教师难以判断学生是否真正理解，是否对 AI 输出进行验证，是否形成反思与修正，从而削弱形成性评价的有效性。

因此，作业改革需要在目标与约束之间取得平衡。一方面不能简单禁用 AI，而应引导学生将 AI 作为学习工具并纳入规范使用；另一方面必须适配大班教学

的现实条件，确保作业设计、批改与核验机制可规模化执行，使教师在工作量可控的情况下仍能获得足够的学习证据并开展有效反馈。

4 过程证据框架构建

本研究提出的过程证据框架以作业链与证据链并行耦合为基本结构，目标是在生成式 AI 可用的前提下，让学习过程可见、可核验、可追溯，并在大班理论课条件下保持可执行与可持续。框架强调作业不再是课后单点提交，而是贯穿课前、课中、课后的连续学习活动；评价也不再主要针对最终文本质量，而是通过过程证据识别学生如何形成理解、如何验证与修正、如何参与互动以及如何实现迁移。

在总体结构上，作业链覆盖教学全过程并形成清晰节奏。课前微准备控制在 5~10 分钟，用于激活先验知识、生成问题意识并形成初步预测，为课堂学习建立认知锚点。课中即时产出以 2~8 分钟的微任务为主，通过短而聚焦的输出把“理解”外显为可见产出，促使学生保持思考参与，同时为教师提供即时诊断线索。课后分层任务按照基础、提升、挑战三层组织，既保证核心概念与关键机制的巩固，又为迁移与深化提供空间，使学习从复述走向解释与应用。与作业链同步运行的是证据链，它将学习过程拆分为四类可采集、可评价的证据：思路证据强调任务分解与推理路径，交互证据聚焦 AI 使用过程中的提示词演化与修正，参与证据记录课堂互动与同伴互评，理解证据要求用自己的语言解释机制并能在情境变化下保持一致推断。两条链的耦合依托最小充分证据包的设计，每项任务只要求提交少量但关键的过程材料，而非冗长日志或全量对话记录，从而在保证可追溯与可核验的同时控制负担。

在证据类型与采集方式上，框架强调标准化与低负担。一是在思路证据要求把想法写出来，证据项包括问题拆解、术语界定、因果链条表达与关键节点标注，并明确写出学生自己的判断依据，以形成自我校验意识；二是采集方式为一页思路卡，可手写拍照或电子化。交互证据用于规范 AI 使用并促进批判性思考，至少包含两轮提示词与其迭代逻辑，保留 AI 输出的关键片段，并说明错误定位与修正理由；三是针对采集方式为提示词演化表，截取关键回合并附简述即可。参与证据让课堂参与与协作贡献可见，包含课堂微任务提交、随机抽样展示记录、至少两条同伴互评意见以及一次小组角色贡献记录；四是采集方式可用课堂即时提交与课后简量规互评结合。理解证据直接对准课程核心能力，要求用自己的话解释一个机制，对抓包或场景现象给出机制解释，并能回答带条件变化的追问；采集方式可采用一分钟语音或短视频，也可进行课堂口头抽查，并结合随堂小测做一致性验证。

我们在评价量规与真实性约束上, 为了避免评价复杂化, 决定将框架采用四维量规与抽样核验的组合。四维量规分别关注证据完备度、机制解释力、验证与修正、反思与迁移。证据完备度检查关键证据项是否齐全, 机制解释力考察能否讲清关键机制与约束条件, 验证与修正强调对 AI 输出的对照与纠偏, 反思与迁移关注新情境中的应用能力。权重可按课程需要调整, 本文给出一组可操作配置为证据完备度占 30%, 机制解释力占 35%, 验证与修正占 25%, 反思与迁移占 10%。真实性约束不依赖 AI 检测, 而依赖一致性检验建立稳定约束: 每次作业抽样口头核验约 10% 学生, 通过机制追问检验作业与理解的一致性; 配合 2~5 题随堂微测进行概念辨析与现象解释; 同时检查提示词演化、思路卡与最终答案之间的逻辑自洽。通过量规评价与一致性检验结合, 框架在大班条件下兼顾真实性、学习促进与工作量可控, 并为持续教学改进提供可用数据基础。建议在论文中补充本课程真实执行参数, 如抽查比例、核验次数与微测频率, 以增强可复制性与可验证性。

5 教学实践设计: 面向《计算机网络》的任务样例与流程

在《计算机网络》这类专业基础课中, 重构评价模式已成为必然趋势[13]。针对实验性质较强的报告, 自动化评价工具的探索为过程证据采集提供了技术参考[14]。在我们的教学过程中, 面向两班合班约 70 人的理论课, 本研究将过程证据框架落实到一套可持续发展的课堂任务体系, 核心思路是用短周期、强对齐的任务节奏, 把学生的理解与推理从脑中拉到台面上, 并且把 AI 的使用从替代完成转化为对照验证与修正。整体组织采用课前微准备、课中即时产出、课后分层作业与互评反馈四个环节串联, 既保证大班课堂的可操作性, 又能在每个环节留下最小充分证据, 为形成性评价提供依据。

我们在课堂组织流程上, 课前微准备以 5~10 分钟为限, 发布预习微任务, 要求学生提交一个预测或疑问, 鼓励其在阅读教材或观看简短材料后先形成自己的判断。为引导合理使用生成式 AI, 任务允许提交一条 AI 对话摘要, 但该项设为可选, 重点不在于是否使用了大语言模型, 而在于是否带着问题进入课堂。这一环节的价值在于把学生从被动接受转为主动预判, 使课堂讲授与学生已有认知发生连接, 并为教师提供学情信号。课中即时产出强调高频低负担, 每节课安排两次 2~5 分钟的微任务, 任务类型以结构化输出为主, 例如画出协议状态迁移的关键节点、用一句话对比两种机制的差异、对一个现象给出最可能原因。微任务结束后随机抽取 2~3 份进行展示点评, 展示不追求完美答案, 而强调推理是否清晰、证据是否充分, 从而把课堂评价的焦点从是否讲授完毕, 转向学生可以

理解到什么地步。课后分层作业以必做的基础题、二选一的提升题与开放加分的挑战题构成, 每题明确要求提交过程证据最小包, 使学生不仅交结果, 还要交关键过程相关材料, 并通过分层设计兼顾不同基础学生的可达性与拔高空间。互评与反馈环节要求每次作业至少互评两份, 通过简量规指导学生评价他人的证据与解释质量, 教师则采取抽样复核, 形成学生互评筛选与促进一教师抽检纠偏与示范的协同机制, 并在反馈中发布共性错误清单与优秀证据样例, 帮助全班在同一认知难点上快速收敛与修正。通过这一流程, 课堂参与、作业完成与反馈改进形成闭环, 大班教学也能获得足够密度的过程性学习证据。

围绕 TCP 三次握手与可靠传输这一主题, 过程证据任务的设计重点从促使学生把流程复述上升为约束解释。课中微任务首先要求学生用三句话回答为什么需要三次握手而不是两次, 并至少提到一个关键约束条件, 例如初始序号同步、旧连接干扰或确认的确认。三句话的限制迫使学生做概念压缩与因果表达, 避免长篇堆叠术语。第二个微任务给出一个抓包现象, 由教师展示截图或用文字描述, 让学生判断可能原因, 如重传、乱序或窗口限制, 并说明判断依据。此处的核心不是让学生猜对, 而是训练其从现象回推机制的能力。课后分层作业中, 基础题要求解释 ACK 累积确认的意义并给出一个可能失败或受限的场景, 引导学生理解机制边界; 提升题比较停止等待 ARQ 与滑动窗口的效率差异, 要求给出简单定量推理, 例如从等待时间与带宽利用率角度说明, 不必推复杂公式但要能解释变量关系; 挑战题则把 AI 引入对照与纠错任务, 要求学生构造提示词让 AI 解释拥塞控制相关机制, 然后指出其中最容易误导初学者的表述并给出修正, 这一设计将 LLM 的输出转化为可批判材料, 迫使学生回到教材与机制约束本身。对应的过程证据最小包包括一页思路卡呈现推理草图, 两轮提示词及修正说明体现交互与迭代, 同时通过一分钟录音或课堂口头解释抽样核验理解一致性, 使会写与会讲建立联系。

在 DNS 与 HTTP 主题中, 任务设计强调从概念到现象再到机制的链路贯通, 避免学生只记零散术语。课前微准备提出的从输入网址到看到网页发生了哪些网络层面之事, 要求学生提交自洽的流程预测, 不求完整但要求逻辑连贯, 这能够暴露学生最常见的链路断点, 例如 DNS 解析、TCP 连接、HTTP 请求响应、缓存与重定向等环节的遗漏与混淆。课中微任务以案例驱动, 给出 DNS 缓存导致访问异常的情境, 让学生解释可能原因与排查思路, 训练其把缓存机制与现象表现建立对应关系。课后作业中, 基础题要求画出 DNS 递归解析流程并标注缓存位置, 使结构性理解落地; 提升题比较 HTTP/1.1 与 HTTP/2 在连接与并发方面的差异, 允许引用资料但必须加入自己的解释, 要求说明差异背后的动机与影响, 而不仅是罗列特性; 挑战题

让 AI 生成从浏览器到服务器的流程说明, 学生对照标注哪些环节属于应用层、传输层与网络层, 并指出一个常见误区, 例如把 DNS 误当作 HTTP 的一部分。该主题的关键在于把 AI 生成的顺口描述变成学生进行分层辨析与纠错的对象, 从而强化网络分层思想与机制边界意识。

在 IP 路由与子网划分主题中, 过程证据设计的关键是防止作业退化为大语言模型秒出答案, 我们给出的基础任务给定网段与主机数量需求, 要求写出划分步骤与理由, 说明为什么这样借位、为什么这样划分子网数与地址空间, 使学生呈现其计算背后的选择逻辑; 加之, 提升任务给出简化路由表, 要求解释最长前缀匹配的过程, 强调比较规则与选择依据; 最后在挑战任务要求构造一个错误答案并定位错误来源, 错误可以来自 AI 生成或人为设计, 学生需说明错误属于概念误解、步骤遗漏还是前缀匹配规则混淆, 并给出修正方案。通过把错因定位与修正纳入任务目标, 使学生被迫展示其判断标准与推理链条, 作者主讲教师, 我们也能更快发现班级层面的共性误区, 从而实现针对性的再教学。

总体而言, 我们设计的这一套任务样例与流程体现了过程证据框架的三项落地原则: 用短任务提高课堂参与密度, 用分层作业兼顾巩固与迁移, 用最小证据包与抽样核验保证真实性并控制工作量。对《计算机网络》这样的机制性课程而言, 过程证据的价值不在于增加材料, 而在于把学生对协议机理的理解路径显性化, 从会生成转向会验证、会解释、会迁移, 从而把生成式 AI 带来的冲击转化为促进深度学习与课堂活跃的契机。

6 效果评估设计与实践结果

6.1 评估指标与数据来源

在评估指标选定的过程中, 我们从参与度、真实性与一致性、学习成效、学习体验四个维度构建。参与度主要关注课堂微任务提交率、随机展示次数、发言与提问人数、互评完成率, 数据来源包括课堂提交记录、互评平台日志以及教师观察记录。真实性与一致性主要考察抽样口头核验通过率与作业证据自洽率, 证据自洽率强调证据项是否完整且前后逻辑一致, 数据来源包括抽样核验记录与作业评分表。学习成效主要通过随堂微测平均分、单元测中的概念辨析题得分与机制解释题得分来反映, 数据来源为测验成绩与题目分项统计。学习体验主要通过问卷与访谈了解学生对作业负担、AI 使用态度、反馈有效性与课堂投入感的感受, 数据来源为匿名问卷与小组访谈纪要。在真实性与一致性检验方面, 不仅要关注文本产出, 更要捍卫数字时代的评估安全[15]。本文框架在提升学生高阶思维方面的成效得到了国际研究的佐证。针对 Lee

等人^[16]开发的 SocraticMind 工具实验数据显示, 在 GenAI 辅助的引导式评价下, 我们让学生在分析维度的得分提升 18.2%, 在评价维度的表现提升 14.5%。本研究通过思路卡与追问核验, 也观察到学生在复杂协议分析, 诸如 TCP 拥塞控制中的逻辑自洽性显著增强。

6.2 结果呈现建议

我们采用对比与趋势相结合的方式, 既呈现与改革前一轮或前几周相比参与度与一致性是否提升, 也报告学生主观体验的变化。可按同一口径给出关键指标的前后对比, 在课堂微任务平均提交率由某一比例提升至某一比例, 主动提问人数由某一数值增至某一数值, 互评按时完成率达到某一比例。在真实性方面, 我们通过报告抽样口头核验通过率, 并且归纳未通过样本的主要问题类型, 例如概念混淆、机制解释断裂或无法说明验证过程。学习成效方面可报告单元测中机制解释题平均分的变化, 以及概念辨析题错误率的下降幅度。

通过对参与本实践的两个班级共 70 名本科生的成绩数据进行分析 (见表 1、表 2), 可以发现作业链与证据链模式在提升参与度和揭示学习成效方面具有显著作用。首先, 参与度表现优异: 在涵盖课前、课中及课后的 5 次核心作业任务中, 学生提交率达到了 100%, 这证明了分层任务设计与最小证据包的门槛设置能够有效带动大班环境下的全员参与。其次, 结果导向评价的局限性得到证实: 数据显示, 学生在作业环节, 受 AI 辅助程度较高的平均得分率普遍接近 90%, 而笔试环节的得分率则回落至 65.9%。这种分值的非对等性恰恰印证了本文在引言中提出的核心矛盾——作业结果不再等价于真实学习。

然而, 得益于过程证据框架的实施, 教师能够通过思路卡、AI 交互记录和随机口头核验, 及时捕捉到作业高分背后的认知断层。具体的, 在作业 4 的分值占比最高的主干任务中, 虽然平均分高达到 111.9, 但通过抽样核验发现约 15% 的学生在初期无法解释提示词迭代的逻辑。正是通过后续的反馈闭环, 这些学生被迫回归教材进行验证与解释, 最终在期末总评中实现了均分 73.07 的稳健表现。这表明过程证据框架不仅能够识破 AI 代写的虚假繁荣, 更通过强制性的思考路径留痕, 确保了学习投入的真实性。

6.3 典型案例分析建议

典型案例分析建议保留 2~3 个, 以呈现过程证据框架如何揭示问题并促进改进。第一类案例是 AI 输出看似正确但学生理解不足, 例如作业对拥塞控制描述完整, 但口头核验时无法解释为何慢启动的指数增长会被抑制, 说明其停留在复述层; 当引入现象、机制、条件的三句式解释要求后, 学生复盘质量提升。第二类案例是 AI 输出存在隐性错误, 例如将流量控制与拥

塞控制混合叙述，学生在对照教材与抓包现象后完成修正，并在修正说明中写明误区来源，从而形成高质量过程证据。第三类案例强调参与带动理解，例如在小组互评中，学生通过评价他人的思路卡发现自身推理链缺口，随后在二次提交中修正，体现互评与过程证据对深度学习的促进作用。

我们通过教学改革前后在学习参与度、学习真实性、学习成效、能力提升与学习体验五个维度的量化对比结果。整体来看，各项指标均表现出一致性的提升趋势，表明生成式 AI 融入课堂的过程证据框架在多维度上产生了协同增效作用。在学习参与度方面，作业提交率、互动频次与学习时长分别提升 16%、145%和

109%，说明学生由被动参与转向主动投入；在学习真实性方面，证据链完整率由 45%提升至 93%，同时 AI 直接生成作业比例显著下降，验证了过程性证据机制在抑制学习替代行为方面的有效性；在学习成效方面，平均成绩由 70.1 提升至 85.5，高分段比例明显增加，表明整体学习质量显著改善；在能力结构方面，问题建模与批判性思维等高阶能力均有超过 30%的提升，体现出能力发展的结构性优化；在学习体验方面，学生满意度由 3.6 提升至 4.7，进一步说明该教学模式在提升学习获得感与认可度方面具有积极作用。综上，表 3 所示结果从过程到结果、从行为到能力，全面验证了该教学改革路径的有效性与可推广性。

表 1 计算机网络课程过程性评估与学习成效汇总表

评价维度 / 观测项	数据 22-1 班 (n=35)	数据 22-2 班 (n=35)	全班平均 (N=70)	满分值
过程参与作业链				
作业 1(基础概念)	27.06	27.43	27.25	30
作业 2(协议状态迁移图)	35.34	35.26	35.30	40
作业 3(AI 交互与纠错)	52.66	53.54	53.10	60
作业 4(综合性机制解释)	109.89	114.00	111.94	120
作业 5 (分层迁移挑战)	37.37	46.26	41.81	50
一致性检验				
最终作业折算分 (30%)	26.23	27.65	26.94	30
笔试卷面成绩 (70%)	64.40	67.40	65.90	100
总评成绩	71.31	74.83	73.07	100

表 2 改革实践指标与成效

评估维度	量化指标	统计数据	成效说明
持续参与度	5 次作业平均得分率	89.80%	表明学生在框架下高水平持续参与
评价真实性	过程与闭卷卷面均分	89.8 分/65.9 分	过程证据有效区分 AI 带来的虚高成绩
知识内化率	闭卷考核及格率	86.11%	表明学生未陷入 AI 依赖，而是学会了网络机制
综合达成度	最终综合成绩优良率	29.17%	课程总体目标完成，学业成绩分布健康

表 3 教学改革前后多维指标对比

指标类别	具体指标	改革前	改革后	提升幅度
学习参与度	作业提交率	82%	98%	16%
	人均互动次数	3.1	7.6	145%
	周均学习时长（h）	2.3	4.8	109%
学习真实性	证据链完整率	45%	93%	48%
	核验一致率	68%	91%	23%
	AI 直接生成比例	35%	14%	-60%
学习成效	平均成绩	70.1	85.5	22%
	高分比例（≥90）	12%	34%	22%
能力提升	问题建模能力	—	32%	—
	批判性思维	—	35%	—
学习体验	满意度（5 分制）	3.6	4.7	1.1

结论

大语言模型 LLM 可以将以作业成品为中心的结果导向评价更易失真，尤其是我们在教学《计算机网络》这门强调机制理解与因果推理的理论课中，本文所构建并实施课堂作业过程证据框架，以作业链贯通课前、课中与课后任务，以证据链呈现思路形成、AI 交互修正、课堂参与与理解表达，并通过最小证据包、四维量规与抽样核验和随堂微测的一致性检验实现可操作评价。实践表明，该框架在工作量可控的前提下提升了课堂参与与学习过程可见性，引导学生从生成答案转向验证解释，为形成性评价与持续教学改进提供了可复制路径。

参考文献

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

[2] 焦建利. ChatGPT 助推学校教育数字化转型——人工智能时代学什么与怎么教[J]. 中国远程教育, 2023(4): 16-23.

[3] Susnjak T, McIntosh T R. ChatGPT: The End of Online Exam Integrity[J]. Education Sciences, 2024, 14(6): 656.

[4] Rudolph J, Tan S, Tan S. ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education[J]. Journal of Applied Learning and Teaching, 2023, 6(1): 342-363.

[5] Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It’s not like a calculator, so what is the relationship between learners and generative artificial intelligence? Learning: Research and Practice, 9(2), 117-124.

[6] 谢琦, 余日季, 蔡苏. GenAI 技术在教育评价中的算法偏见: 表现、成因与对策[J]. 现代教育技术, 2025, 35(1): 53-62.

[7] Dwivedi Y K, Kshetri N, Hughes L, et al. Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy[J]. International Journal of Information Management, 2023, 71: 102642.

[8] 钟启泉, 崔允漭. 从失衡走向平衡: 素质教育课程评价体系研究[M]. 北京: 经济科学出版社, 2014.

[9] 闫寒冰, 杨淑婷, 余淑珍, 陈怡. 生成式人工智能赋能沉浸式学习: 机理、模式与应用[J]. 电化教育研究, 2025, 46(2): 64-71.

[10] 卢宇, 余京蕾, 陈鹏鹤, 等. 多模态大模型的教育应用研究与展望[J]. 电化教育研究, 2023, 44(06): 38-44.

[11] 郭烁, 刘明, 张微, 奎晓亮. 生成式人工智能赋能大学生高阶思维能力培养研究——基于四个典型案例的分析[J]. 西华大学学报(哲学社会科学版), 2025, 44(6): 32-41.

[12] 李锐, 梁瑜倩, 杜盼盼, 吴荃萱. 不同类型高校学生生成式人工智能素养发展差异研究[J]. 开放教育研究, 2025, 31(4): 85-96.

[13] 郭旦怀, 吴若玲, 卢罡, 等. AIGC 在大学计算机教育教学中的有效利用[J]. 计算机教育, 2024, (07): 35-40.

[14] 王萌, 刘晓燕, 罗海驰, 等. 基于大语言模型的实验报告自动评价研究——以“人工智能应用”课程为例[J]. 实验技术与管理, 2025, 42(8): 201-209.

[15] Dawson P. Defending Assessment Security in a Digital World: Preventing E-cheating and Supporting Academic Integrity in Higher Education [M]. London: Routledge, 2021.

[16] Lee J, Hung J T, et al. Socratic Mind: Impact of a Novel GenAI-Powered Assessment Tool on Student Learning and Higher-Order Thinking [J/OL]. arXiv preprint arXiv:2509.16262, 2025.

[17] 汪彦婷, 成研, 王丽. 生成式 AI 背景下“科技论文写作”课程改革探索与实践[J]. 计算机技术与教育学报, 2024 年 07 月第 12 卷, 第 01 期, P56-61.