智能体评语反馈能提升学习效果吗？

——基于学习成绩、学习动机与自我效能感的准实验研究

陈道杰[1]　王　觅[2]　钟志贤[1]

（1. 江西师范大学 教师教育高等研究院，江西南昌 330022；
2. 海南师范大学 人工智能学院，海南海口 571127）

[摘要] 数智时代，智能体技术正逐步嵌入教学评价场景，为评语反馈的即时性与个性化提供新的可能。然而，智能体评语反馈能否切实提升学习效果，尚缺乏充分的研究证据。本研究以高中编程教学为研究情境，构建“肯定优点—指出问题—改进建议—鼓励支持”四维评语结构，设计并开发基于多模块架构的评语反馈智能体，并开展准实验研究，即以某高一年级 98 名学生为对象开展为期 12 周的教学实验，探究智能体评语反馈对学习成绩、学习动机与自我效能感的影响。研究发现，智能体评语反馈可有效提升学生的学习成绩，组间差异随时间推移呈线性扩大趋势；可增强学生学习动机的整体表现，对激发努力投入的表现最明显；可显著增强学生的自我效能感，一般学习效能感与具体学习任务效能感提升更突出；学生整体满意度较高，师生访谈证实了“智能体提供即时反馈、教师开展深度引导”的人机协同共识。研究结论可为教学反馈的智能化转型提供参考。

[关键词] 智能体；评语反馈；学习效果；准实验研究

[中图分类号] G434　　**[文献标识码]** A　　**[文章编号]** 1007-2179(2026)06-0094-11

一、问题提出

教学反馈是影响教学效果的重要环节，被视为“对学生学习最具影响力的因素之一”（Hattie & Timperley, 2007）。人工智能技术的蓬勃发展，为教育评价改革注入了新动能（吴砥等，2023），国家对此高度重视并作出积极回应。2026 年 1 月，怀进鹏部长在全国教育工作会议上强调，要“持续深化教育综合改革，深化评价改革，扎实推进人工智能赋能教育”（中华人民共和国教育部，2026）。同年，教育部等五部门印发《“人工智能+教育”行动计划》，提出要“利用智能系统参与教学环节”“提升沉浸式体验和个性评价反馈，提升课堂育人质效”（中华人民共和国中央人民政府，2026）。在此背景下，如何借助智能技术实现教学反馈的即时性与精准化，为不同学习者提供差异化支持，成为当前教育研究亟待回应的实践命题。

有学者尝试引入 GPT-4 等通用大语言模型辅

[收稿日期]2026-07-24　　**[修回日期]**2026-11-06　　**[DOI编码]**10.13966/j.cnki.kfjyyj.2026.06.008

[基金项目] 国家社会科学基金教育学一般项目“GenAI 赋能新质人才培养的生成式学习设计研究”(BCA240057)。

[作者简介] 陈道杰，博士研究生，江西师范大学教师教育高等研究院，研究方向：人工智能教育、教育数字化转型（370069191@qq.com）；王觅（通讯作者），教授，硕士生导师，海南师范大学人工智能学院，研究方向：人工智能教育、数字化资源设计与开发（549857900@qq.com）；钟志贤，教授，博士生导师，江西师范大学教师教育高等研究院，研究方向：教育数字化转型、教师教育、数字人文主义教育。

[引用信息] 陈道杰，王觅，钟志贤(2026). 智能体评语反馈能提升学习效果吗？——基于学习成绩、学习动机与自我效能感的准实验研究[J]. 开放教育研究，32(6)：94-104.

助中小学编程教学(褚乐阳等, 2024)，但此类工具并非为教育场景专门设计，存在直接生成完整代码、削弱学生独立思考、反馈内容缺乏教育温度等问题(傅骞等, 2026)。智能体技术凭借即时生成、结构化诊断与个性化适配的优势，为破解上述难题提供了可行路径。它能对学生的课程作业开展自动分析与诊断，并以自然语言形式生成针对性评语，在一定程度上弥补了传统人工反馈在即时性与个性化方面的不足(姚嘉轩等, 2026)。尽管现有研究在智能体的功能设计、技术实现及教学应用等方面讨论广泛，但关于智能体评语反馈对学生多维学习效果影响的实证研究仍较匮乏。为此，本研究以高中编程教学为情境，探讨智能体评语反馈对学习效果的影响。研究着重探究三个问题：与传统教师反馈相比，智能体评语反馈能否提升学生的学习成绩？能否增强学生的学习动机？能否增强学生的自我效能感？

二、文献综述

(一)教学反馈

反馈是有效教学不可或缺之要素(王靖等, 2023)。围绕教学反馈，既有研究主要从反馈类型、反馈内容与反馈策略三个维度展开。

在反馈类型方面，已有研究提出了多种划分标准：依据功能，分为认知型反馈、元认知型反馈和动机型反馈(Panadero & Lipnevich, 2022)；依据主体，分为人工反馈与技术反馈(Kotlyar & Krasman, 2025)；依据形式，分为口头反馈与书面反馈(Tärning, 2018; 王明娣等, 2025)。其中，评语作为一种以自然语言为媒介的书面反馈形式，因其能够传递丰富的诊断信息与情感信号而备受关注。有研究指出，评语不仅是认知反馈的载体，更是师生情感互动的重要渠道(郁晓华等, 2022)。相关研究还发现，与单纯的分数反馈相比，评语在激发学习动机与提升学业成就方面效果更好(Koenka et al., 2021)。另外，模糊、简化的评语难以促进学习，认知导向与情感激励相结合的评语更有利于学生深度参与(任强等, 2025)。生成式人工智能辅助生成的教师评语亦被证实能有效提升学生的自我调节学习能力和学习动机(罗恒等, 2024)。

在反馈内容方面，有效反馈通常具有具体性、解释性与发展性特征。具体性要求反馈信息清晰指向学生的具体表现，而非笼统的优劣判断(Mandouit & Hattie, 2023)；解释性要求反馈信息不仅告知对错，更应阐释背后的原理与逻辑，帮助学生理解“为何如此”(Förster, 2025)；发展性要求反馈信息为学生提供可操作的改进建议，指向后续行动(魏非等, 2026)。相关研究证实，具备可操作步骤的反馈能有效提升学习者的元认知监控水平和学习效率(Hao et al., 2022)。

在反馈策略方面，既有研究从时序控制、结构设计与交互模式等维度提出了多种策略。即时反馈适用于基础技能训练与程序性知识习得(陈卫东等, 2025)，延时反馈契合需要深度认知加工的任务场景(Shute, 2008; 蔡旻君等, 2020)，分步式反馈通过任务分解降低学习者的认知负荷(Johnson et al., 2015)，对话式反馈被证实能有效促进学生对复杂概念的理解与高阶思维发展(蒋慧芳等, 2024)。

(二)智能体在编程教育中的应用

智能体指能够感知环境、自主决策并执行行动以实现特定目标的智能系统。基于大语言模型的智能体因其强大的自然语言理解能力，成为该领域发展的重要方向(Xi et al., 2025)。当前，国内外关于智能体在编程教育中的应用研究呈上升趋势。具体而言，国外研究起步较早，2003 年便有研究者开发了基于智能体的编程学习支持系统(Nagao & Ishii, 2003)。2020 年前后，相关研究成果集中涌现，且主要聚焦智能体支持的编程学习与教学应用等主题。例如，有研究者依托智能体的调试与交互功能为学生提供上下文敏感的编程支持，发现学生的自我效能感与编程熟练度均有显著提升(Wang et al., 2025)；还有研究者(Ouyang et al., 2024)比较了 ChatGPT 智能反馈与教师的人工反馈发现，前者在促进学生自我调节方面表现优良，后者更有利于引导学生围绕知识展开讨论。

国内研究虽起步较晚，但发展迅速。2021 年，卢宇等(2021)首次将“智能体”概念与编程教育关联；杨思恩等(2022)提出的编程教育智能体(intelligent programming tutor, IPT)，标志着该技术在教学实践中正式落地。近年来，研究集中于大模

型驱动与人机协作，卢宇等(2024)基于大模型构建了编程作业评估智能体，翟雪松等(2024)以编程教育为例探索智能体支持的人机协作教学模式。在成效检验方面，赵路瑶等(2025)发现生成式人工智能能显著增强学生编程信心，但对编程兴趣的影响不显著；傅骞等(2026)生开发了引导型和追问型对话式教学智能体，发现整合两种策略在提升编程成绩和问题解决能力上显著优于单一策略；冯小燕等(2025)以大学生为对象，发现智能体能有效提升计算思维整体水平及促进各维度的协同发展。

总的来说，既有研究多从教学模式的整体效能或技术功能验证角度切入，对“智能体评语反馈”这一具体干预方式关注不足。基于此，本研究以高中编程教学为情境，设计结构化评语反馈智能体，通过准实验研究检验其对学习成绩、学习动机与自我效能感的影响，以期为智能体评语反馈的教学应用提供实证依据，亦为智能教育工具的设计与优化提供参考。

三、研究设计

(一)智能体设计与开发

本研究以刘明等(2024)的教育智能体技术架构为参照，结合真实教学场景需求，依托扣子(Coze)平台设计并开发面向高中编程教学的评语反馈智能体——“编程学习助手小 D 老师”。该智能体以 DeepSeek-v4-pro 大语言模型为“大脑”，联动规划、记忆、工具与行动等模块，通过可视化工作流编排与知识库调用，实现对作业的自动化分析、诊断与结构化评语的即时生成(见图 1)。

如前文所述，有效的反馈内容常兼具具体性、解释性与发展性，即清晰指向学习表现、阐释背后原理和提供可操作的改进方向。同时，认知引导与情感激励相结合的评语更有利于激发学习动机与维持投入。由此，本研究以哈蒂和廷珀利(Hattie & Timperley, 2007)的反馈层次模型为认知主线，以广泛应用于沟通与评价领域的“三明治法则”(Prochazka et al., 2020)为情感主线，构建“肯定优点—指出问题—改进建议—鼓励支持”四维评语反馈结构(见图 2)。其中，“肯定优点”帮助学习者明确已有进展，“指出问题”揭示目标与当前表现之间的差距，“改进建议”指明后续行动方向，“鼓励支持”维护学习信心与动力。该结构在认知层面遵循从“纠错”到“促学”的递进逻辑，在情感层面以积极开端与鼓励收尾维护学习动机，实现认知引导与情感支持的统一。

在功能设计上，该智能体能依据四维结构生成

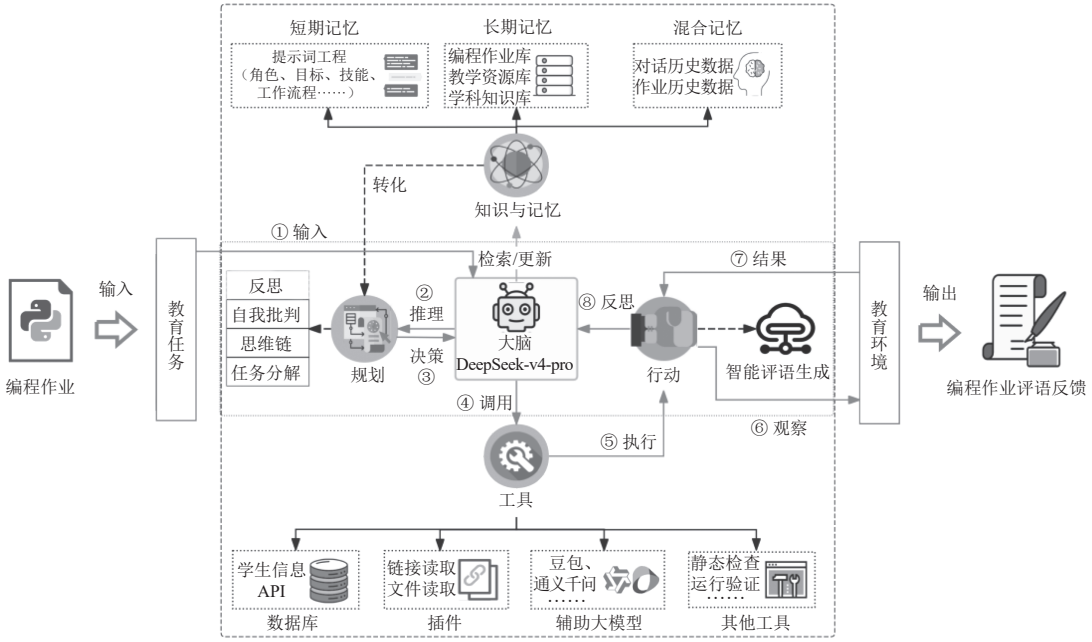


图 1　编程评语反馈智能体技术架构

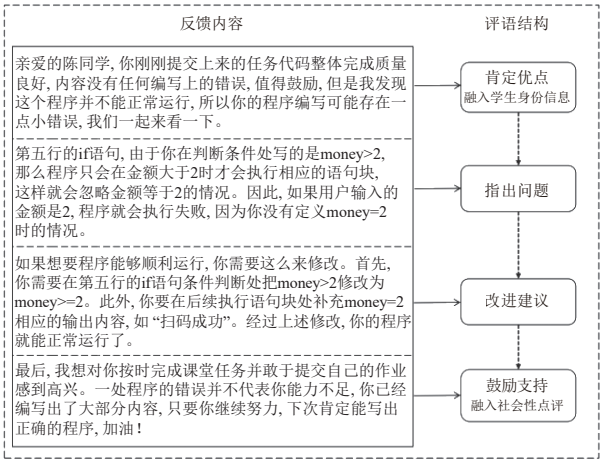


图 2　四维评语反馈结构及输出示例

具有针对性与可操作性的自然语言评语，并结合代码上下文进行逻辑诊断与原理阐释；坚持“引导而非代劳”的原则，通过流程与规则设计杜绝直接提供答案的行为；为教师提供专属后台观测界面，支持查看学生作业提交进度、下载源代码文件与回溯对话内容（见图 3）。

（二）研究对象

本研究选取海南省一所教育信息化示范校的高一年级两个班级学生作为研究对象，每班 49 人，两班学生的学业水平、师资配置等方面基本一致。独立样本 t 检验结果显示，两班第一次测试成绩（$t = 0.387$, $p = 0.700$）、学习动机前测（$t = 1.480$, $p = 0.142$）及自我效能感前测（$t = 0.881$, $p = 0.381$）均无显著差异，同质性良好。本研究将其中一个班级设为实验组（接受智能体评语反馈），另一个设为对照组（接受传统教师反馈）。研究征得全体学生及其



图 3　编程评语反馈智能体功能

监护人的知情同意，研究数据严格保密。

（三）研究工具

本研究开展了四次阶段性测试，测试内容为“计算机解决问题的过程”“作出判断的分支”“周而复始的循环”“可以复用的代码”四个教学模块，各模块满分 50 分。试题由三位一线信息技术教师组成命题小组编制，覆盖各阶段核心知识点与能力目标。题型分别为选择题、填空题和操作题。命题严格遵循《普通高中信息技术课程标准（2017 年版 2020 年修订）》。

学习动机量表改编自麦考利等（McAuley et al., 1989）编制的内在动机量表。考虑到原量表的“感知能力”维度与自我效能感存在构念重叠，为避免重复测量，本研究统一将其称为“价值”维度。该维度源自自我决定理论中的“认同调节”，即个体认可活动的重要性并将其纳入自我价值体系（Ryan & Deci, 2020）。改编后的量表包含兴趣、努力、压力与价值四个维度，共 16 道题。预测试结果显示，量表内部一致性良好（Cronbach’s α = 0.890）。探索性因子分析结果显示，KMO 值为 0.892，Bartlett 球形检验显著（p>0.001），各题项均落在预期因子上，累积方差解释率为 82.043%，表明量表结构效度良好。

自我效能感量表改编自边玉芳编制的量表（2004），包含一般学习效能感、具体学习任务效能感与控制效能感三个维度（共 13 道题，采用李克特五点计分），Cronbach’s α 系数为 0.955，说明其信度较高。此外，本研究还采用改编自黄和金（Hwang & Kim, 2022）的智能体评语反馈满意度调查问卷，并辅以半结构化访谈收集师生使用体验。本研究运用目的性抽样开展访谈：选取授课教师 1 名、课堂观察教师 2 名；按照学业表现分层，选取高、中、低学业水平学生共 5 名，以期覆盖不同层次学生的学习体验。

（四）研究方法

本研究采用混合研究方法。量化数据涵盖四次阶段测试成绩、学习动机与自我效能感前后测结果及满意度问卷调查结果。统计分析借助 SPSS 27.0 完成，具体流程为：采用混合设计方差分析方法比较学习成绩，辅以简单效应检验与多项式趋势对比；采用以各维度前测为协变量的多元协方差分析方法（MANCOVA）比较学习动机与自我效能感，并运用 Bonferroni 法进行多重比较校正；效应量以偏 Eta 平方（η_p^2）报告（初始显著水平被设定为 α = 0.05）。质性研究资料源于对 3 位教师与 5 名学生的半结构化访谈。本研究将访谈录音转录为文本后，采用 NVivo 20 开展主题编码分析，基于“认知体验—情感反应—行为意向”框架进行演绎编码与子主题提炼，以对量化结果进行三角互证与补充解释。

（五）实验流程

研究采用“实验组—对照组”前后测准实验设计，历时 12 周（见图 4）。两组学生在统一教学内容、进度、师资与环境基础上，遵循相同教学流程，即“回顾知识—提出任务—分析任务—讲解编程—学生实操—提交作业—智能体/教师反馈—课堂小结”。差异主要体现在反馈环节：实验组学生提交作业后，当堂获得“编程学习助手小 D 老师”生成的结构化评语；对照组学生则沿用传统方式，由教师根据课堂任务完成情况给予口头点评，并于课后批改作业，下节课集中讲解。

实验分三个阶段：第 1 周为前测与准备阶段，研究者组织两班学生完成第一次阶段测试及动机、效能感量表前测，并向实验组学生介绍智能体操作流程；第 2—11 周为教学干预与反馈实施阶段，两组学生同步学习“编程计算”单元第 2—4 节内容，

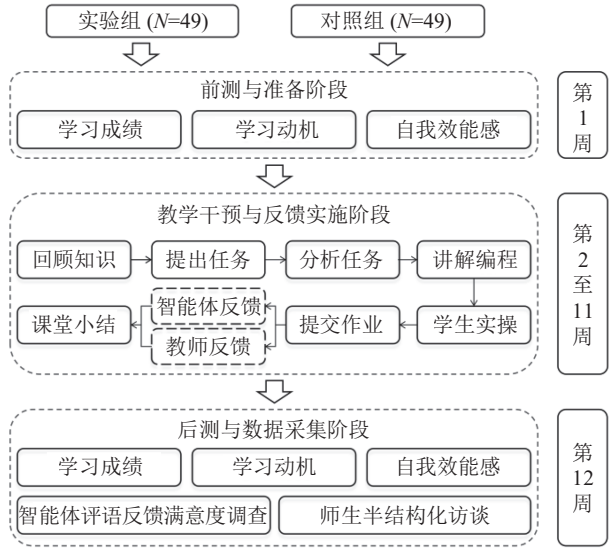


图 4　实验流程

接受差异化反馈干预，并于第 4、7 周参与第二、三次阶段测试；第 12 周为后测与数据采集阶段，即学生接受第四次阶段测试、学习动机与自我效能感后测、满意度调查及师生接受半结构化访谈。

四、研究结果

（一）智能体评语反馈对学习成绩的影响

为检验智能体评语反馈对学生学习成绩的影响，本研究以四次阶段测试成绩为观测指标（见图 5）。在测试 1 阶段，两组得分基本持平，实验组（40.45 分）与对照组（40.12 分）无显著差异，两组学生起点同质。自测试 2 起，随着教学内容由概念理解转向代码编写，两组成绩均有所回落，但实验组（37.04 分）降幅小于对照组（34.49 分）。此后，实验组成绩逐次回升，至测试 4 达 39.71 分；对照组回升幅度有限，后测为 35.18 分，两组分差由测试 2 的 2.55 分扩大至测试 4 的 4.53 分。

以测试时间为组内变量、组别为组间变量进行混合设计方差分析，经 Mauchly 球形检验，数据满足球形假设（$p = 0.461$）。方差分析结果显示，时间主效应、组别主效应及时间与组别的交互效应均显著（见表 1）。这表明，四次测试成绩存在显著差异；实验组整体成绩显著优于对照组，两组成绩变化轨迹存在显著差异。简单效应分析结果显示，两组在测试 1 无显著差异（$p = 0.700$），自测试 2 起实验组显著优于对照组。此外，本研究通过多项式对比检验交互效应的趋势发现，时间与组别交互效应的线性趋势显著，$F(1, 96) = 15.542$，$p < 0.001$，$\eta_p^2 = 0.139$，二次及三次趋势均不显著，这说明两组成绩差距呈显著线性扩大趋势。

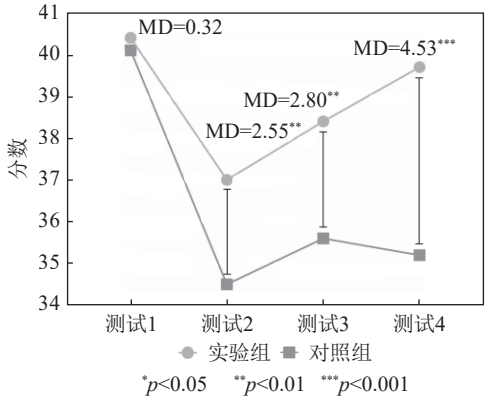


图 5　四次阶段测试的成绩比较

表 1　学习成绩的混合设计方差分析结果

变异来源	自由度（*df*）	*F* 值	*p* 值	η_p^2
时间	3, 288	28.303	< 0.001***	0.228
组别	1, 96	18.909	< 0.001***	0.165
时间×组别	3, 288	5.783	< 0.001***	0.057

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

综合来看，相较于教师反馈，智能体评语反馈能够有效提升学生学习成绩。在编程学习由概念转向实操的难度跃升后，实验组展现出更强的学习韧性，成绩持续回升。两组分差随时间推移呈显著线性扩大趋势，说明智能体即时、个性化的评语反馈对学生学习成绩产生积极影响。

（二）智能体评语反馈对学习动机的影响

为控制前测差异并了解多维度变量的整体效应，本研究以学习动机前测得分为协变量，组别为固定因子，后测得分为因变量，进行多元协方差分析（见表 2）。Box’s M 检验表明，协方差矩阵齐性假设成立（$p = 0.024 > 0.001$）。多变量检验结果显示，在控制前测分数后，组别主效应显著，Pillai’s Trace = 0.164，$F(4, 89) = 4.378$，$p = 0.003$，$\eta_p^2 = 0.164$，说明智能体评语反馈对学习动机整体有显著正向影响。

本研究采用 Bonferroni 法进行多重比较校正（$\alpha = 0.0125$）发现，实验组在“努力”维度上显著优于

表 2　学习动机的协方差分析结果

维度	组别	前测（M ± SD）	后测（M ± SD）	*F* 值	*p* 值	η_p^2
兴趣	实验组	3.83 ± 0.87	4.01 ± 0.86	5.698	0.019*	0.058
	对照组	3.51 ± 1.01	3.56 ± 0.93			
努力	实验组	3.58 ± 0.84	3.78 ± 0.83	13.122	< 0.001***	0.125
	对照组	3.12 ± 1.00	3.15 ± 0.83			
压力	实验组	2.84 ± 0.87	3.16 ± 0.84	4.662	0.033*	0.048
	对照组	2.96 ± 0.98	2.73 ± 0.88			
价值	实验组	3.83 ± 0.82	3.82 ± 0.87	0.982	0.324	0.011
	对照组	3.74 ± 0.74	3.62 ± 0.77			
总分	实验组	56.33 ± 9.75	59.08 ± 9.71	10.838	0.001**	0.102
	对照组	53.27 ± 10.70	52.25 ± 9.85			

注：表中 *p* 值未经多重比较校正，检验水准为 $\alpha = 0.05$；*$p < 0.05$，**$p < 0.01$，***$p < 0.001$；压力维度为反向计分，得分越高表示压力感越小。

对照组，且经校正后依然稳健；在“兴趣”和“压力”维度上，组间差异在校正前达统计学意义上的显著水平，但经严格校正后未达显著；在“价值”维度上，两组校正前后无显著差异。

上述分析结果表明，智能体评语反馈能增强学习动机，且主要体现在学生努力投入上。“兴趣”与“压力”维度经严格校正后未达统计显著水平，但校正前呈现的显著差异，仍预示其具有一定的积极改善趋势。“价值”维度未表现出显著变化，说明该反馈模式促进价值认同的作用不明显。

（三）智能体评语反馈对自我效能感的影响

本研究同样以自我效能感前测得分为协变量，组别为固定因子，后测得分为因变量进行多元协方差分析（见表 3）。Box’s M 检验显示，协方差矩阵齐性满足（$p = 0.506 > 0.001$），多变量检验显示组别整体效应显著（Pillai’s Trace = 0.143, $F(3, 91) = 5.059$, $p = 0.003$, $\eta_p^2 = 0.143$）。这说明智能体评语反馈对增强自我效能感有显著作用。

经 Bonferroni 校正后（$\alpha = 0.0167$），实验组在“一般学习效能感”和“具体学习任务效能感”维度上显著优于对照组；在“控制效能感”维度上，两组差异未达统计学意义上的显著水平，但实验组后测值略高于对照组。

上述结果表明，智能体评语反馈能有效增强学生的自我效能感，尤其是增强“一般学习效能感”与“具体学习任务效能感”的效果突出。这说明，即时且个性化的评语反馈，能增强学生应对具体编程任务的信心，并由此转化为他们对自身编程学习能力的自信。然而，它对于情绪调节与课堂环境适应的“控制效能感”，其促进作用有限。这说明该反馈模式主要作用于任务层面的能力建构，对自我调控能力的干预效果不明显。

表 3　自我效能感的协方差分析结果

维度	组别	前测（M ± SD）	后测（M ± SD）	F 值	p 值	η_p^2
一般学习效能感	实验组	3.27 ± 0.83	3.65 ± 0.87	15.090	< 0.001***	0.140
	对照组	3.13 ± 0.89	2.95 ± 0.84			
具体学习任务效能感	实验组	3.01 ± 0.80	3.47 ± 0.84	11.577	< 0.001***	0.111
	对照组	2.80 ± 0.88	2.84 ± 0.85			
控制效能感	实验组	3.75 ± 0.72	3.96 ± 0.80	3.907	0.051	0.040
	对照组	3.78 ± 0.65	3.65 ± 0.68			
总分	实验组	42.65 ± 8.94	47.49 ± 9.88	15.751	< 0.001***	0.142
	对照组	40.98 ± 9.85	39.76 ± 8.87			

注：p 值未经多重比较校正，检验水准为 $\alpha = 0.05$；*$p < 0.05$，**$p < 0.01$，***$p < 0.001$。

（四）师生对智能体评语反馈的满意度与使用体验

为全面了解学生对智能体评语反馈的主观评价，本研究采用经改编的满意度量表开展问卷调查，并通过半结构化访谈收集师生使用体验。

1. 满意度问卷调查结果

满意度量表涉及内容满意度、设计满意度与交互满意度三个维度（采用李克特五点计分）。由表 4 可知，学生对智能体评语反馈的总体满意度较高，满意度指数达 87.4%。三个维度的得分均在 4.0 以上，内容满意度和设计满意度并列，交互满意度略低。从题项看，“评语能准确指出代码中的错误或可改进之处”（M = 4.57）和“反馈让我感觉像在与老师互动”（M = 4.53）得分最高，这表明学生对智能体评语反馈的准确性和亲和力高度认可。得分低的题项集中在“与同学交流评语内容”（M = 4.18）和“把评语建议运用到后续编程任务”（M = 4.31），这预示未来研究可在促进同伴互评与反馈迁移等方面优化，以提升学生的编程学习效果。

2. 访谈发现

对 3 位教师与 5 名学生的半结构化访谈文本的主题编码分析发现，受访师生大多认可智能体评语反馈，整体评价较积极。在认知体验上，学生高度认可反馈的即时性与纠错精准性，“提交即反馈”有效解决了传统反馈滞后的问题；教师充分肯定其在减轻工作负担、弥补大班教学短板与帮扶基础薄弱学生方面的价值。在情感反应上，学生反映无压力的人机对话环境有助于缓解编程焦虑，持续的

表 4　学生对智能体评语反馈的满意度

维度	题项数量	平均值（M）	标准差（SD）
内容满意度	5	4.39	0.72
设计满意度	5	4.39	0.73
交互满意度	5	4.34	0.72
总体满意度	15	4.37	0.69

“修正—成功”体验增强了编程学习信心；教师形成“智能体即时反馈、教师深度引导”的人机协同共识，并强调教师核心角色的不可替代性。在行为意向上，受访学生表示收到评语后会先自主复盘再进行修改，自主学习行为显著增强，多数学生愿意挑战有难度的题目。此外，师生提出精简评语内容、强化思维引导、优化教学场景适配等建议。整体来看，访谈结果与满意度调查结果在核心结论上相互印证，表明师生普遍认可智能体的评语反馈，也为后续迭代改进指明了方向。

五、结论与讨论

本研究从学习成绩、学习动机与自我效能感三个维度，检验智能体评语反馈对学生学习的作用效果。结果表明，智能体评语反馈在三个维度均发挥了促进作用，但影响效应呈现差异化特征。

第一，智能体评语反馈对学习成绩的提升作用优于教师反馈，且随干预的推进持续增强。从反馈干预理论的视角看，反馈通过引导学习者觉察自身表现与学习目标之间的差距，激发其调整认知与行为（Kluger & DeNisi, 1996）。反馈能否有效触发这一过程，关键在于能否回答学习者关心的三个核心问题：学习目标是什么、当前进展如何、下一步该往哪里去（Hattie & Timperley, 2007）。智能体评语反馈的四维结构恰好回应了上述三个问题。与等待教师集中反馈不同，智能体在学生编程思维最活跃时提供反馈干预，能有效降低学生的认知负荷，促进其深度建构与内化程序性知识。当前，高中编程教学的普遍现象在于，学生提交代码后，往往需等到下节课才能获得教师的批改意见。这种反馈信息与学习情境之间的断裂，使反馈效果大打折扣。智能体即时反馈使“提交、诊断、修正”的反馈闭环能在同一课时内完成，有效化解“反馈滞后于学习进程”的现实困境。教师可依托这一特性重构课堂节奏，从繁重的批改任务中抽身，将更多精力投入课堂，用于观察学生的学习改进过程、捕捉共性错误等。由此，技术工具与教师专业判断互补协同，实现即时反馈与深度引导的有机结合。

第二，智能体评语反馈对学习动机整体具有促进作用，但在具体维度上存在差异。这一结果可借助自我决定理论予以阐释。该理论指出，当外部环境支持自主需求、能力需求和归属需求时，个体的外在动机将更趋于内化与整合（Ryan & Deci, 2020）。智能体通过及时纠错与提供建议满足学生能力发展需求，让学生在获得可操作的修改指引后及时看到改进效果；通过仿人化语气和激励性评价营造归属感，让学生感受到被关注与得到支持。这解释了努力维度的积极变化。值得注意的是，初步检验结果显示兴趣与压力得到显著改善，但经过严格的多重比较校正后，其显著性未能保持。这可能源于智能体评语反馈主要作用于认知层面的代码诊断与纠错，其鼓励多为程序化的言语说服，在激发深层探索欲与提供深度情绪抚慰方面的作用相对有限。另外，价值维度未能显著转变，这一方面可能源于价值认同的变化常需长时间的干预；另一方面，智能体评语反馈对自主需求的支持相对有限，学生缺乏对反馈方式、深度和频率的选择权。从教学实际看，高中生学习编程时往往持“完成任务即止”的态度，较少主动追问代码背后的设计逻辑与优化空间，这种浅表化的学习状态反映出其价值认同的薄弱。智能体虽可在认知层面提供即时支持，但学习意义感的建立仍需教师的持续引导。从教学实践的角度看，本研究认为教师运用智能体反馈应为学生保留一定的自主空间，如允许学生选择反馈的详细程度或自主设定学习目标。同时，教师应引导学生从“关注结果对错”转向“关注思维过程”，通过课堂对话帮助学生建立对编程价值的深层认同，这也是教师难以被智能体取代的价值所在。

第三，智能体评语反馈通过多重路径增强学生的自我效能感，并能有效防止效能感在困难任务中的衰减。从实验结果看，学生根据反馈成功修正代码的体验构成了“亲历成功经验”；评语中嵌入的解题思路提供了“替代性经验”；鼓励性话语构成持续的“言语说服”；无压力的人机对话改善了“情绪状态”（Bandura, 1997）。上述四个方面协同作用，共同提升了学生的整体效能感。然而，学生“控制效能感”维度的改善较小，究其原因，可能在于控制效能感涉及学生对学习过程与结果的主动规划与监控，其改善依赖长期自我调节经验的积累。值得关注的是，实验组效能感有所提升，对

照组略有下降。这或许说明，在传统的教师反馈模式下，随着编程任务难度的逐级递升，学生可能因缺乏即时支持而产生累积性挫败。当前，高中编程教学存在的不容忽视问题是：教师在评价学生编程作业时往往聚焦代码能否运行、结果是否正确，较少关注学生编程的思维路径与调试策略。这种结果导向的评价方式容易让学生形成“对错二元”的能力认知，一旦频繁出错便归因于自身能力不足。智能体评语反馈不仅能发挥“提升”作用，也能起到“保护”作用，有效防止学生面临困难任务时自我效能感的自然衰减。这启发我们，智能体反馈应关注基础薄弱学生的“小步成功”体验，通过细粒度的任务分解与即时正向反馈，帮助学生逐步建立信心，避免因早期挫败而彻底放弃。同时，教师可借助智能体后台数据识别效能感偏低的学生，及时给予针对性鼓励与策略指导。

第四，人机协同是智能体评语反馈有效发挥作用的保障，智能体宜被定位为“第一响应者”，教师则承担深度引导与价值引领的职责。访谈揭示的“智能体即时反馈、教师深度引导”共识具有重要的教学意义。这一模式并非简单地将反馈任务完全交由智能体，而是在明确各自优势边界基础上的人机有机协同。智能体的优势在于即时性、规模化与一致性，它能在学生认知活跃期提供反馈，能同时处理全班数十份作业并保持评价标准统一。教师的不可替代性体现在深度引导、情感联结与价值引领：教师能基于对学生个体背景的了解进行差异化引导，能通过共情回应学生的情绪波动，能在技术工具外传递学科价值与学习意义。这种协同模式本质上是教学流程的重构，教师从繁重、机械性的作业批改工作中解放出来，聚焦高层次的教学设计、针对性的个别化辅导与高质量的课堂互动。当然，智能体存在的评语冗长、精准思路引导不足、复杂逻辑识别能力弱等问题表明，它更适合被定位为学习的“第一响应者”，即在学生遇到问题时提供初步诊断与引导，而非承担最终教学解释的“终极解答者”。教师教学应明确人机分工，智能体负责语法纠错与基础逻辑诊断，教师负责共性问题讲解、高阶思维引导与个别化激励。只有学习者从“被评价者”变为“自我调节的学习者”，智能体评语反馈才能真正超越纠错工具的定位，成为促进深度学习与自主发展的教学伙伴。

总之，人工智能及其衍生技术的飞速发展，正逐步推动智能体深度融入教育教学全过程，为评语反馈实现即时化、个性化与规模化提供了新的可能。然而，此类技术能否切实转化为提升学生学习效果，仍是亟待探索的命题。本研究立足当前教学反馈的现实困境，以高中编程教学为情境，设计并开发适配学生认知特点的评语反馈智能体，通过准实验研究验证其在提升学习成绩、激发学习动机与增强自我效能感方面的积极作用。研究发现，智能体评语反馈能够在一定程度上弥补传统人工反馈在即时性与个性化方面的短板，成为拓展教师教学能力、重构教学流程的伙伴。由此形成的“智能体提供即时反馈、教师开展深度引导”的人机协同模式，能为人工智能赋能教育教学提供路径。

本研究局限在于：样本仅来自一所学校的两个班级，其规模与师生来源有限，研究结论能否推广至不同地区、不同技术条件的学校，有待验证；干预以班级为单位实施，干预效应与班级效应未能完全分离，且两种反馈模式在时机、形式、个别化程度与技术新颖性上存在固有差异，研究效应的精确归因有待检验；12 周的研究周期虽足以观察到学习成绩、努力投入及核心学习效能感的变化，但对学习的深层兴趣与价值认同，以及情绪与环境适应的控制效能感等复杂心理特质，可能需要更长时间的持续干预才能显现。未来研究可扩展至多学校、多学段及多学科样本，采用多班级重复实验或个体随机分配设计，并延长追踪周期，以检验智能体评语反馈的作用边界与优化方向，推动智能教育工具从“可用”走向“善用”。

［参考文献］

[1] Bandura, A.（1997）. Self-efficacy: The exercise of control[M]. New York: W. H. Freeman and Company: 80-106.

[2] 边玉芳（2004）. 学习自我效能感量表的编制 [J]. 心理科学，(5)：1218-1222.

[3] 蔡旻君，郭婉瑢，娄颜超（2020）. 在线学习过程中如何实施有效的反馈——基于自我调节学习理论的在线反馈探讨 [J]. 电化教育研究，41(10)：82-88.

[4] 陈卫东，刘晨阳，李晓茹，丁景洋，苏荣荣（2025）. 身智共融：具身智能技术驱动教与学变革的机制、路径与反思 [J]. 远程教育杂志，

43(6): 63-72.

[5] 褚乐阳, 王浩, 陈向东(2024). 面向大语言模型的青少年人工智能教育 [J]. 中国电化教育, (4): 32-44.

[6] 冯小燕, 张语函, 田晟瑶, 胡萍(2025). 编程学习中 GenAI 对计算思维的影响研究 [J]. 现代教育技术, 35(12): 66-76.

[7] Förster, M., Broder, H. R., Fahr, M. C., Klier, M., & Fink, L.(2025). Tell me more, tell me more: The impact of explanations on learning from feedback provided by Artificial Intelligence[J]. European Journal of Information Systems, 34(2): 323-345.

[8] 傅骞, 赵亚宁, 甘甜甜, 郑娅峰(2026). 对话式教学智能体促进中学生编程学习表现的实证研究 [J]. 电化教育研究, 47(3): 37-46.

[9] Hao, Q., Smith, D., Ding, L., Ko, A., Ottaway, C., Wilson, J., Arakawa, K., Turcan, A., Poehlman, T., & Greer, T.(2022). Towards understanding the effective design of automated formative feedback for programming assignments[J]. Computer Science Education, 32(1): 105-127.

[10] Hattie, J., & Timperley, H.(2007). The power of feedback[J]. Review of Educational Research, 77(1): 81-112.

[11] Hwang, S., & Kim, H. K.(2022). Development and validation of the e-learning satisfaction scale (eLSS)[J]. Teaching and Learning in Nursing, 17(4): 403-409.

[12] 蒋慧芳, 曾文婕, 朱琦(2024). 指向深度学习的对话式反馈：特征、方式与策略 [J]. 开放教育研究, 30(2): 80-88.

[13] Johnson, A., Reisslein, J., & Reisslein, M.(2015). Transitional feedback schedules during computer-based problem-solving practice[J]. Computers & Education, 81(feb.): 270-280.

[14] Kluger, A. N., & DeNisi, A.(1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory[J]. Psychological Bulletin, 119(2): 254-284.

[15] Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H.(2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback[J]. Educational Psychology, 41(7): 922-947.

[16] Kotlyar, I., & Krasman, J.(2025). Student reactions to AI versus human feedback in teamwork skills assessment[J]. International Journal of Educational Technology in Higher Education, 22(1): 57.

[17] 刘明, 杨闽, 吴忠明, 廖剑(2024). 教育大模型智能体的开发、应用现状与未来展望 [J]. 现代教育技术, 34(11): 5-14.

[18] 卢宇, 汤筱玮, 宋佳宸, 余胜泉(2021). 智能时代的中小学人工智能教育：总体定位与核心内容领域 [J]. 中国远程教育, (5): 22-31+77.

[19] 卢宇, 余京蕾, 陈鹏鹤(2024). 基于大模型的教学智能体构建与应用研究 [J]. 中国电化教育, (7): 99-108.

[20] 罗恒, 廖小芳, 茹琦琦, 王志锋(2024). 生成式人工智能支持的教师评语研究：基于初中数学课堂的实践探索 [J]. 电化教育研究, 45(5): 58-66.

[21] Mandouit, L., & Hattie, J.(2023). Revisiting "The power of feedback" from the perspective of the learner[J]. Learning and Instruction, 84: 101718.

[22] McAuley, E., Duncan, T., & Tammen, V. V.(1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis[J]. Research Quarterly for Exercise and Sport, 60(1): 48-58.

[23] Nagao, K. , & Ishii, N. (2003). Evaluation of learning support system for agent-based C programming[C]// Knowledge-Based Intelligent Information and Engineering Systems (KES 2003). Oxford, UK. Springer: 540-546.

[24] Ouyang, F., Guo, M., Zhang, N., Bai, X., & Jiao, P.(2024). Comparing the effects of instructor manual feedback and ChatGPT intelligent feedback on collaborative programming in China' s higher education[J]. IEEE Transactions on Learning Technologies, 17: 2227-2239.

[25] Panadero, E., & Lipnevich, A.(2022). A review of feedback models and typologies: Towards an integrative model of feedback elements[J]. Educational Research Review, 35: 100416.

[26] Prochazka, J., Ovcari, M., & Durinik, M.(2020). Sandwich feedback: the empirical evidence of its effectiveness[J]. Learning and Motivation, 71: 101649.

[27] 任强, 管珂滢(2025). 对话哲学视域下作业评语的教育检视与重构 [J]. 宁波大学学报(教育科学版), 47(6): 67-77.

[28] Ryan, R., & Deci, E.(2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions[J]. Contemporary Educational Psychology, 61: 101860.

[29] Shute, V. J.(2008). Focus on formative feedback[J]. Review of Educational Research, 78(1): 153-189.

[30] Tärning, B.(2018) . Review of feedback in digital applications–Does the feedback they provide support learning?[J]. Journal of Information Technology Education: Research, 17: 247-283.

[31] Wang, H., Wang, C., Chen, Z., Liu, F., Bao, C., & Xu, X.(2025). Impact of AI-agent-supported collaborative learning on the learning outcomes of University programming courses[J]. Education and Information Technologies, 30(12): 17717-17749.

[32] 王靖, 巴安妮, 吴宝锁(2023). 智能教学反馈的目标、机制与架构 [J]. 现代远程教育研究, 35(3): 102-112.

[33] 王明娣, 程雅雯(2025). 数智赋能素养生成的课堂反馈：结构模型与实现机制 [J]. 中国电化教育, (11): 34-41.

[34] 魏非, 杨可欣, 杨淑婷(2026). 生成式人工智能赋能教师能力场景化评估：框架、路径与实践 [J]. 开放教育研究, 32(2): 66-75.

[35] 吴砥, 郭庆, 吴龙凯, 程浩(2023). 智能技术赋能教育评价改革 [J]. 开放教育研究, 29(4): 4-10.

[36] Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Qin, W., Zheng, Y., Qiu, X., Huang, X., Zhang, Q., & Gui, T.(2025). The rise and potential of large language model based agents: A survey[J].

Science China-Information Sciences, 68(2): 121101.

[37] 杨思恩, 沈国华, 李鑫, 黄志球, 周宇(2022). 基于代码推荐的编程教育在线辅导方法实践探索 [J]. 计算机教育, (3): 153-156.

[38] 姚嘉轩, 夏小俊(2026). 大模型赋能的编程作业评估智能体设计与应用 [J]. 计算机教育, (2): 98-103.

[39] 郁晓华, 战晓瑜(2022). 教师评语知多少?——探析文本后面的情感价值 [J]. 电化教育研究, 43(7): 97-105.

[40] 翟雪松, 张丽洁, 夏亮亮, 徐鑫, 朱强(2024). 基于 GAI 的逆向工程教学思维在人机协作中的应用研究——以编程教育为例 [J]. 电化教育研究, 45(9): 61-68.

[41] 赵路瑶, 杨现民, 兰颖(2025). 生成式人工智能赋能高中编程教学的成效研究 [J]. 现代教育技术, 35(9): 56-65.

[42] 中华人民共和国教育部(2026). 2026 年全国教育工作会议召开 [EB/OL]. [2026-09-10]. http://www.moe.gov.cn/jyb_zzjg/huodong/202601/t20260108_1426054.html.

[43] 中华人民共和国中央人民政府(2026). 教育部等五部门关于印发《“人工智能+教育”行动计划》的通知 [EB/OL]. [2026-09-10]. https://www.gov.cn/zhengce/zhengceku/202604/content_7065138.htm.

（编辑：魏志慧）

Can Agent-Based Comment Feedback Improve Learning Outcomes? —A Quasi-Experimental Study on Academic Achievement, Learning Motivation, and Self-Efficacy

CHEN Daojie[1], WANG Mi[2] & ZHONG Zhixian[1]

（1. *Institute of Teacher Education for Advanced Study*, *Jiangxi Normal University*, *Nanchang* 330022, *China*； 2. *School of Artificial Intelligence*, *Hainan Normal University*, *Haikou* 571127, *China*）

Abstract: *In the era of digital intelligence, agent technology is gradually being infused into teaching evaluations, with new possibilities for the immediacy and personalization of comment feedback. However, whether agent-based comment feedback can effectively improve learning outcomes still lacks sufficient empirical evidence. To address this, using senior high school programming teaching as the research context, this study, with a quasi-experimental design, constructed a four-dimensional comment structure (“affirming strengths, pinpointing problems, proposing improvements, and offering encouragement”), designed and developed a comment feedback agent based on a multi-module architecture. A 12-week experiment, with 98 first-year students from a senior high school, explored the effects of agent-based comment feedback on academic achievement, learning motivation, and self-efficacy. The study found that agent-based comment feedback effectively improved students’ academic achievement, with between-group differences showing a linearly widening trend over time; positively promoted students’ overall learning motivation, particularly in stimulating their effort investment; and significantly enhanced students’ self-efficacy, with more prominent improvements in general learning self-efficacy and specific task self-efficacy. In addition, students showed high overall satisfaction, and teacher-student interviews further revealed a human-AI collaborative consensus of “immediate feedback from the agent and in-depth guidance from the teacher.” The study concluded that agent-based comment feedback had a positive effect on improving students’ learning outcomes, which can provide practical implications and insights for the intelligent transformation of teaching feedback.*

Key words: *agent*; *comment feedback*; *learning outcomes*; *quasi-experimental study*