Creative Education Studies 创新教育研究, 2026, 14(11), 117-131
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ces
https://doi.org/10.12677/ces.2026.1411869

Hans汉斯

人工智能驱动计算机类专业过程性考核的实践与反思

蔡维维，耿雅婷，雷玉霞

曲阜师范大学计算机学院，山东 日照

收稿日期：2026年9月18日；录用日期：2026年10月31日；发布日期：2026年11月12日

摘　要

计算机类专业课程的过程性考核仍存在评价维度单一、反馈滞后和个性化支持不足等问题。基于教学改革实践，本文将大语言模型辅助评阅、课程知识图谱和认知诊断引入过程性考核，构建智能考核系统。系统对主观题和代码作答进行分维度辅助评阅，并结合Q矩阵与G-DINA模型形成知识点掌握诊断，据此提供反馈与学习支持。研究以一门专业核心课程开展准实验，比较实验组与对照组在评阅效率、学习表现和反馈利用等方面的变化，并结合访谈与课堂观察进行分析。结果显示，该机制有助于缩短反馈周期，促进学生依据反馈改进任务成果，也为教师识别共性问题和调整教学提供依据。大语言模型仅承担辅助初评和诊断支持，关键评价由教师复核确认。

关键词

人工智能，过程性考核，计算机教育，学情诊断，教学反思

Practice and Reflection on AI-Driven Process-Oriented Assessment for Computer Science Majors

Weiwei Cai, Yating Geng, Yuxia Lei

School of Computer Science, Qufu Normal University, Rizhao Shandong

Received: September 18, 2026; accepted: October 31, 2026; published: November 12, 2026

Abstract

Formative assessment in computer science courses still faces several challenges, including limited evaluation dimensions, delayed feedback, and insufficient personalized support. Drawing on

文章引用: 蔡维维, 耿雅婷, 雷玉霞. 人工智能驱动计算机类专业过程性考核的实践与反思[J]. 创新教育研究, 2026, 14(11): 117-131. DOI: 10.12677/ces.2026.1411869

teaching reform practices, this study integrates large language model-assisted assessment, a course knowledge graph, and cognitive diagnosis into formative assessment, and develops an intelligent assessment system. The system conducts multidimensional auxiliary assessment of open-ended responses and programming submissions, and combines a Q-matrix with the G-DINA model to generate diagnostic information on students' mastery of knowledge points, which is then used to support feedback and learning. A quasi-experimental study was conducted in a core professional course to compare changes in assessment efficiency, learning performance, and feedback use between an experimental group and a comparison group, supplemented by interviews and classroom observations. The results indicate that the proposed mechanism helps shorten the feedback cycle, encourages students to improve their work based on feedback, and provides evidence for instructors to identify common learning difficulties and adjust teaching. The large language model is used only for auxiliary preliminary assessment and diagnostic support, while key evaluative judgments are reviewed and confirmed by instructors.

Keywords

Artificial Intelligence, Formative Assessment, Computer Science Education, Learning Diagnosis, Teaching Reflection

Copyright © 2026 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

近年来，随着“新工科”建设的推进和工程教育专业认证要求的落实，计算机类专业人才培养更加重视学生工程实践能力、问题解决能力和创新能力的形成，这对课程评价方式提出了新的要求。过程性考核作为教学评价的重要组成部分，通过持续记录学生的学习表现并提供反馈，使教师能够了解学生的知识掌握和能力发展情况，并据此调整教学。许多学者进行了相关研究，例如从工程教育认证视角辨析了过程性评价与形成性评价在评价目的、内容、方法和实施要点等方面的联系与区别，并强调评价应嵌入教学过程，通过持续诊断和反馈促进学习改进[1]。将评价嵌入课堂教学并及时向学生提供有针对性的反馈，有助于发挥形成性评价促进学习的功能[2]。然而，在计算机类专业教学实践中，过程性评价仍常被简化为课堂小测和课后作业，评价重点更多集中于学习结果判定，对学习过程中的问题识别与持续跟踪相对不足，因而难以充分发挥过程性评价在学习进程监测和能力发展支持方面的作用。

当前，计算机类专业过程性考核仍存在一些突出问题。其一，考核内容与能力目标匹配不足。选择题、填空题等客观题型便于规模化实施，却难以充分评价算法设计、程序调试、系统架构等复杂工程能力。已有研究基于 OBE 理念提出课程目标导向的过程化考核设计思路，通过建立量化的考核评价标准模型，实现对学生知识成果、技能成果、态度与价值观的全面评价[3]。其二，反馈时效性不足。作业与实验报告批阅周期较长，反馈信息常滞后于学生学习进程，削弱了评价的形成性功能。其三，主观任务评阅的标准化程度有限。程序设计题、算法设计题等高度依赖教师经验，不同教师与班级之间易出现评分尺度差异，影响评价的公平性与一致性。

人工智能技术，尤其是大语言模型的发展，为上述问题的解决提供了新的可能。在国际上，认知诊断与知识追踪研究已积累了较为深厚的基础。贝叶斯知识追踪模型(Bayesian Knowledge Tracing, BKT)以四参数隐马尔可夫过程估计学生对特定技能的掌握概率，成为智能辅导系统中学习者建模的奠基性方法[4]；深度知识追踪(Deep Knowledge Tracing, DKT)进一步将循环神经网络引入知识追踪，在细粒度习题序

列上提升了知识状态的预测精度[5]。认知诊断模型(Cognitive Diagnostic Model, CDM)的理论框架主张将学生作答反应映射到多维属性掌握模式，而非仅给出单一分数[6]。对自适应与智能 Web 教育系统的分类研究指出，学习者模型与个性化内容推送是自适应学习的两大支柱[7]；智能辅导系统研究则概括了任务选择、步骤级反馈、提示以及学习者模型更新等典型运行机制[8]。近年来，大语言模型教育应用研究进一步指出，其在自动评阅、个性化辅导与反馈生成方面具有较大潜力，同时也面临幻觉、公平性与伦理风险[9]。

在国内，相关研究更多从智慧学习环境、学习分析和评价改革等角度讨论技术对教学的支持作用。智慧学习生态系统研究强调学习环境、数据与学习支持之间的协同[10]；人工智能教师角色研究将自动批阅、学习障碍诊断与反馈等列为人工智能可承担的辅助性工作[11]；基于学习行为数据构建综合评价模型的研究，为学习分析支持评价、诊断与干预提供了方法参考[12]。在计算机教育领域，能力导向评价研究提出由基础技能、高阶思维和未来素养构成的评价框架，并引入增值评价与智能反馈[13]；相关教学实践从教学组织与全过程考核等方面探索多元化、全程性评价[14]；还有研究从 OBE-CDIO 理念出发讨论本硕贯通培养及其评价机制[15]，关注人工智能技术在过程性评价中的应用路径[16]，并在人工智能通识教育中探索“学生主导、AI 支持”的人机协作学习及多主体参与的过程性考核方式[17]。

上述研究为人工智能参与教学评价奠定了较为充分的理论与实践基础，但在程序设计等开放性任务中，现有研究仍有进一步深化的空间。知识追踪与认知诊断多建立在结构化作答数据之上，对代码、设计文档等开放成果中所包含的语义信息，其利用仍有拓展空间；部分教学改革研究侧重于理念设计、应用场景与评价框架，对评阅结果如何进入知识诊断、诊断结果又如何进一步转化为反馈和教学调整，其实现过程尚可作更完整的呈现。从已有实践来看，效果评价多以经验总结或单一结果指标为主，对不同评价方式之间的比较，以及定量资料与访谈、观察等质性材料之间的相互印证，值得进一步补充。

Table 1. Representative studies and the positioning of this paper
表 1. 国内外代表性研究与本文定位

研究方向	代表性工作	主要关注	对本文的启发
AI 辅助过程性评价	形成性评价及国内过程性评价研究[2] [3] [13] [16] [17]	评价效率、过程数据、能力导向	为本文将评价嵌入教学过程，并进一步形成“评价–诊断–反馈–改进”闭环提供思路
智能辅导与自动反馈	自适应学习与智能辅导系统研究[7] [8]	自动反馈、学习者建模、个性化支持	启发本文将反馈、学习者状态分析与个性化学习支持纳入系统流程
认知诊断模型	认知诊断研究[6] [18]	从作答反应推断知识属性掌握状态	为本文估计知识属性掌握概率提供了模型基础
知识追踪	知识追踪研究[4] [5]	技能掌握状态随学习过程动态更新	启发本文利用多次任务证据持续更新知识掌握判断
LLM 教育应用	大语言模型教育应用研究[9]	自动评阅、个性化反馈及风险	提示需通过人工复核等机制控制模型幻觉与评价偏差

基于这一认识，本文并不将重点放在单一算法或模型性能本身，而是关注开放性学习任务在实际教学中的评价过程，即学生提交的代码和文本如何得到辅助评阅，评阅结果如何转化为较为具体的知识诊断，以及这些诊断信息如何服务于学生改进和教师后续教学。研究将大语言模型对开放文本和代码的语义分析与课程知识图谱、Q 矩阵及 G-DINA 认知诊断模型结合，在呈现知识点掌握情况的同时，保留对

能力层级和具体作答表现的分析，使评价结果能够为后续反馈提供较明确的依据。在课程实施中，系统按照任务提交、AI 辅助初评、知识诊断、反馈生成、教师复核和教学调整的基本流程运行，强调智能工具与教师专业判断之间的配合。为考察该机制在真实教学情境中的应用情况，本文以《数据结构》课程为例设置实验组和对照组，结合评阅效率、学习表现、学习投入、反馈利用等过程性资料，以及访谈、课堂观察和反思材料进行分析，并通过具体课程案例呈现其实际运行过程。表 1 据此对相关研究与本文工作的衔接关系作进一步说明。

依托山东省人工智能特色学院等平台，我院持续开展 AI 赋能教育教学改革，在产教融合和智慧课程建设等方面开展了相关实践。通过校企合作将企业真实项目拆解并融入教学案例库，有效衔接课程目标与行业需求，为 AI 赋能过程性考核的落地实施奠定了基础。本文其余部分组织如下：第 2 节阐述设计理念；第 3 节介绍系统设计与实现；第 4 节说明实践路径；第 5 节报告准实验研究设计与结果；第 6 节给出以《数据结构》为例的应用案例；第 7 节讨论促进深度反思的教学策略；第 8 节总结实践成果并展望后续工作。

2. AI 赋能过程性考核的设计理念

2.1. 从“分数统计”到“认知诊断”

传统过程性考核通常以分数作为主要评价结果，虽然能够较为直观地反映学生的整体完成情况，也便于课程教学管理，但单一的分数信息难以进一步揭示学生在知识掌握过程中存在的理解偏差和能力薄弱环节，因而难以充分回应计算机类专业对工程思维和问题解决能力培养的要求。AI 赋能过程性考核，旨在推动评价方式由单纯的“分数统计”向“认知诊断”转变，通过分析学生学习过程中暴露出的具体问题，为教师开展针对性教学干预和差异化教学提供依据[13]。

依托大语言模型的代码语义理解与错误归因能力，结合课程知识点的层级关系，构建“知识点–能力层”双维认知诊断框架。其中，知识维度定位学生作答所涉及的具体知识点，能力维度从识记、理解、应用和创新四个层级进行分析，并进一步识别知识遗漏、概念混淆和逻辑偏差等典型问题。对于相同的错误结果，还可进一步判断其主要源于知识理解偏差还是具体问题处理不当。通过将学生学习过程中出现的问题定位到具体知识点及相应能力层级，可提高后续反馈和教学干预的针对性[16]。

2.2. 从“滞后反馈”到“即时闭环”

及时反馈是过程性考核发挥形成性功能的重要条件。在传统教学中，“布置作业–学生完成–教师批阅–课堂讲评”通常需要经历较长的反馈周期，评价反馈容易与学生实际学习过程产生脱节。尤其在编程类课程中，当学生收到反馈时，往往已经难以完整回溯此前的调试过程，即使评价结果本身较为准确，其纠错和改进作用也可能受到影响[16]。

AI 赋能过程性考核可借助自动化分析与即时反馈缩短评价反馈周期，使学生在完成学习任务后及时发现问题，并依据反馈对原有方案进行检查、修改和验证，形成“任务–诊断–改进”的学习闭环。由此，评价不再局限于对学习结果的事后判定，而是进一步嵌入学习过程，为学生持续调整学习策略和改进任务方案提供支持[16]。

2.3. 从“统一标准”到“个性适配”

计算机类专业学生在编程基础、逻辑思维和学习方式等方面存在较为明显的个体差异。采用相对统一的考核内容和反馈方式，往往难以同时兼顾不同学习基础和能力水平学生的实际需求。基础相对薄弱的学生可能面临较大的学习压力，而学有余力的学生则缺少进一步拓展和提升的空间，这与因材施教的

人才培养要求并不完全契合。

AI 赋能过程性考核可以通过整合和分析学生学习过程中的学情数据，形成动态学习画像，并依据学生在知识掌握、能力发展和学习行为等方面的实际情况，对学习任务、反馈方式和学习资源进行差异化调整。对于基础相对薄弱的学生，可侧重基础知识和基本技能的巩固；对于学有余力的学生，则可适当增加具有一定挑战性的拓展任务，从而使过程性评价由统一要求逐步转向更加符合学生实际学习需求的个性化支持[13]。

3. 系统设计与实现

3.1. 总体架构

为支撑上述设计理念，本研究设计并实现了一套面向计算机类程序设计课程的智能过程性考核系统。系统自下而上分为三层：数据采集层、智能处理层和应用层，整体架构如图 1 所示。

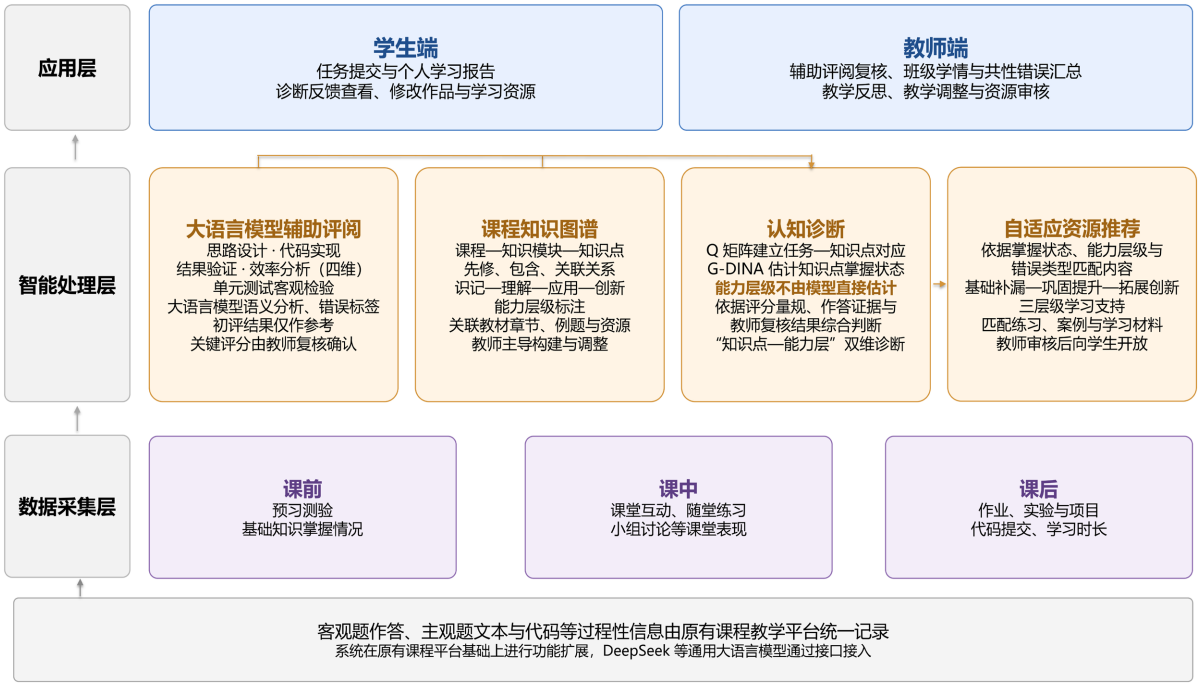


Figure 1. Overall architecture of the system
图 1. 系统总体架构

系统依托现有课程教学平台运行，主要包括数据采集、智能处理和教学应用三个层面。课程任务发布、学生作答及学习过程信息均由平台统一记录，在此基础上接入大语言模型、课程知识图谱、认知诊断和资源推荐等功能。本文关注的重点是这些技术如何参与过程性评价及后续教学改进，因此主要呈现系统的功能关系和运行过程，具体服务器配置、账户管理等部署内容不作展开。

数据采集贯穿课前、课中和课后，主要包括客观题作答、主观题文本与代码、课堂互动以及学习时长等过程性信息。学生提交任务后，系统首先借助大语言模型对代码和文字作答进行辅助分析，并将相关作答证据与课程知识图谱中的知识点建立联系。在此基础上，认知诊断模块形成对知识点掌握情况的辅助判断，能力层级则结合评分量规和学生实际作答加以分析。诊断结果随后用于匹配相应的反馈内容和学习资源。学生端主要呈现个人学习情况和改进建议，教师端则汇总班级整体学情和较为集中的错误

类型，为后续教学判断提供参考。

在具体实现中，系统是在原有课程平台基础上进行功能扩展。DeepSeek 等通用大语言模型通过接口接入，用于主观任务的辅助评阅和问题识别；课程知识图谱承担知识点组织及关系关联；认知诊断模块综合学生在多次任务中的作答证据，资源推荐则依据诊断结果匹配相应的练习、案例和学习材料。这样的设计并不追求技术模块本身的复杂程度，而是尽量使评阅、诊断和反馈能够嵌入原有教学过程，减少教师额外的操作负担。鉴于本文讨论的是过程性评价机制及其教学应用，后文主要说明与评价直接相关的数据流和工作流程。

一次任务从发布到后续教学调整，大致经历任务发布、学生提交、智能评阅、认知诊断、反馈改进和教学反思几个环节，其基本过程如图 2 所示。学生完成并提交任务后，系统先形成辅助评阅结果和诊断信息，教师对其中涉及关键评分、复杂作答和存疑内容进行复核，再将反馈和学习资源提供给学生。学生可以据此检查原有作答并进行修改。与此同时，教师通过汇总后的学情信息了解较为集中的知识难点和错误类型，并据此调整讲解重点、补充案例或安排针对性练习。后续任务中的学习表现又可以作为新的过程性证据进入系统，使评价、学习改进与教学调整保持持续联系。

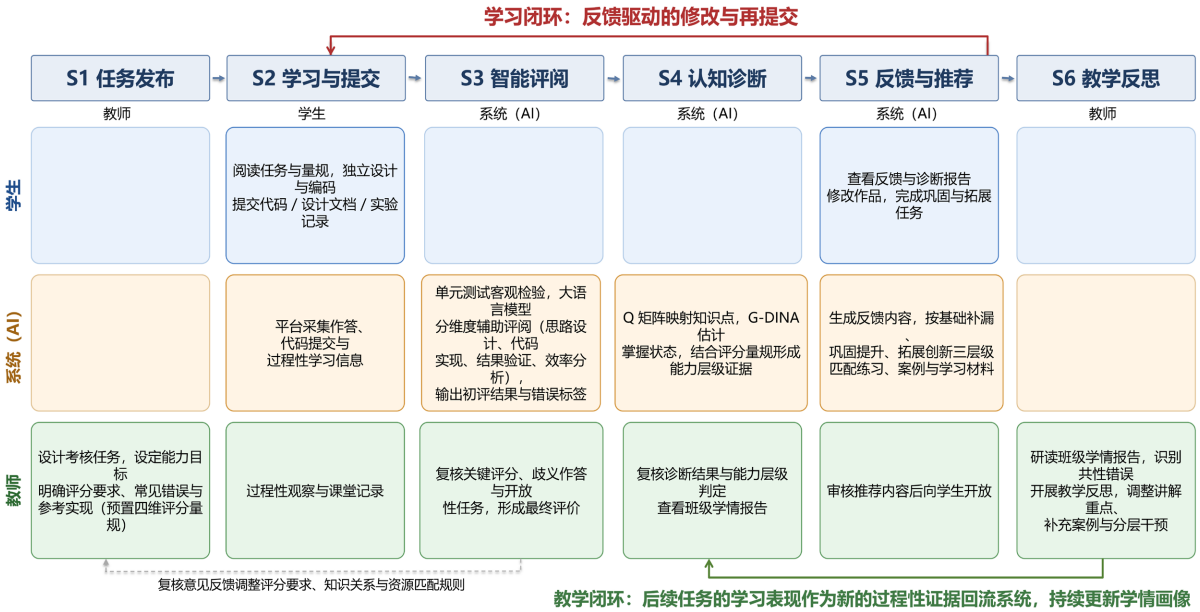


Figure 2. Data flow diagram of formative assessment
图 2. 过程性考核数据流图

图 2 展示任务发布、学生提交、智能评阅、认知诊断、反馈推荐和教师干预之间的数据流向及双向闭环。

3.2. 大语言模型评阅模块

主观题和程序设计题较难采用传统客观题的自动评分方式，这也是本系统重点处理的一类任务。课程平台接入 DeepSeek 等通用大语言模型，用于辅助教师完成此类作答的初步分析。针对程序设计任务，评价主要围绕思路设计、代码实现、结果验证和效率分析几个方面展开。教师在课程建设阶段预先明确评分要求、常见错误和参考实现，学生提交代码后，系统结合题目要求、评分标准、学生作答以及单元测试结果，对作答情况进行初步分析，并给出相应的问题提示和修改建议。

在实际评阅过程中，可客观判断的部分先通过单元测试进行检验，对于程序思路、代码规范和分析说明等较难完全量化的内容，再由大语言模型提供辅助判断。模型给出的结果仅作为教师评阅的参考，不直接用于成绩认定。对于关键评分、存在歧义的作答以及开放性较强的任务，仍由教师结合学生实际提交内容进行复核，并据此形成最终评价。复核中发现的问题还可反过来用于调整评分要求和提示内容。这样的处理既保留了自动化工具在提高初步评阅效率方面的作用，也保证了教师在评价过程中的专业判断和最终责任。

3.3. 课程知识图谱构建

课程知识图谱主要用于组织课程知识内容及其相互关系，并为后续学情诊断提供知识结构基础。图谱构建采用教师主导、系统辅助的方式。首先由任课教师依据课程教学大纲梳理核心知识点，并结合教学内容确定知识点之间的先修(prerequisite)、包含(part-of)与关联(related-to)关系，形成课程知识结构，一般呈现“课程－知识模块－知识点”的层级形态。以《数据结构》课程为例，可围绕线性表、栈与队列、树与二叉树、图、排序和查找等主要内容逐层展开，其中“树与二叉树”又进一步细化为二叉树定义、遍历、线索化、哈夫曼树等具体知识点。

在知识点关系之外，图谱还保留与教学评价直接相关的信息。一方面，根据课程目标和任务要求，对不同知识点所涉及的识记、理解、应用和创新等能力层级进行标注；另一方面，将教材章节、典型例题、代码示例与拓展材料同相应知识点建立联系，为后续反馈与学习资源推荐提供依据。随着课程实施的推进，系统还可以结合学生作答中反复出现的问题发现知识点之间可能存在的关联。例如，当学生在“二叉树递归遍历”和“栈的应用”等内容上经常同时出现错误时，系统可将这一现象作为关联线索提交给教师；教师结合具体作答情况进行判断后，再决定是否对图谱中的知识关系作相应补充或调整。

3.4. “知识点－能力层”双维认知诊断算法

认知诊断模块采用具有较好可解释性的认知诊断模型，并通过 Q 矩阵建立评价任务与课程知识点之间的对应关系。设学生在 K 个知识属性上的掌握状态向量为 $\alpha=(\alpha_1,\cdots,\alpha_K)$，其中 $\alpha_k\in\{0,1\}$ 表示学生是否掌握第 k 个知识属性；第 j 个任务在 Q 矩阵中的行向量记为 $q_j=(q_{j1},\cdots,q_{jK})$，当 $q_{jk}=1$ 时，表示该任务涉及第 k 个知识点。在 DINA 模型中，学生完成任务 j 的理想反应 η_j 定义为 $\eta_j=\prod_k \alpha_k^{q_{jk}}$，即只有当任务所涉及的知识属性均被掌握时，理想反应才取 1 [19]。进一步考虑学生实际作答中可能出现的失误和猜测，引入失误参数 s_j 和猜测参数 g_j，分别表示已掌握相关知识但作答错误，以及尚未掌握相关知识但作答正确的可能性。相应的观测作答概率表示为：

$$P(X_j=1\mid\alpha)=(1-s_j)^{\eta_j}\cdot g_j^{1-\eta_j} \tag{1}$$

考虑到程序设计任务通常同时涉及多个知识点，仅依据单一属性难以充分反映学生的实际作答情况，本文进一步采用 G-DINA 模型描述不同知识属性之间可能存在的交互作用，其线性形式为：

$$P_j=\delta_{j0}+\sum_k \delta_{jk}\alpha_k+\sum_{k<l}\delta_{jkl}\alpha_k\alpha_l+\cdots \tag{2}$$

其中，δ_{j0} 为截距项，δ_{jk} 表示第 k 个知识属性的主效应，δ_{jkl} 表示知识属性 k 与 l 之间的交互效应[18]。与主要以客观题对错结果作为诊断依据的常见应用不同，本系统将单元测试结果与经教师复核的大语言模型分维度初评结果共同作为任务层面的观测证据，再依据 Q 矩阵映射至相应知识属性。具体判定标准由教师结合课程评分量规预先确定，同一知识点在不同任务中的多次表现共同参与知识属性掌握状态的判断，以减少单次作答偶然因素带来的影响。

需要说明的是，G-DINA 模型主要用于分析学生对知识属性的掌握情况，而“识记－理解－应用－

创新”等能力层级并不由该模型直接估计，而是依据评分量规、学生实际作答证据以及教师复核结果综合判断。因此，本文所提出的“知识点–能力层”双维诊断，是在诊断报告中同时呈现知识点掌握信息与能力层级证据，以便教师从不同角度理解学生的学习状况，而不是将两类信息作为性质相同的潜变量进行联合估计。

3.5. 自适应推荐引擎

诊断完成后，自适应推荐模块结合学生的知识点掌握情况、能力层级表现和主要错误类型，为其匹配相应的学习内容。根据课程教学要求，学习支持大致分为“基础补漏–巩固提升–拓展创新”三个层级。对于存在先修关系且掌握相对薄弱的知识点，系统优先提供基础讲解和针对性练习；在此基础上，再结合学生的能力层级和错误表现，匹配不同难度的案例、代码改写任务及拓展资源。随着学生后续作答情况和教师复核结果的更新，推荐内容也相应调整，使诊断结果能够持续用于后续学习支持，并形成“诊断–推荐–再诊断”的动态改进过程。

3.6. 系统工作流程

系统完成一次评阅与诊断时，首先由学生在课程平台提交作业或代码，相关作答记录随即进入后续分析环节。大语言模型依据既定评分量规对作答进行辅助评阅，形成初步评分、能力表现线索和错误标签；认知诊断模块再通过 Q 矩阵将任务表现与相应知识点建立联系，并据此更新知识掌握情况。随后，推荐模块结合知识点掌握状态、能力层级和主要问题，为学生生成相应的反馈内容和学习资源。相关结果分别呈现在学生学习报告和教师端学情信息中，供学生改进学习、教师开展后续教学判断。教师可对各环节结果进行复核，复核意见也可用于调整评分要求、知识关系和资源匹配规则，使系统在实际教学中不断修正和完善。

4. AI 赋能过程性考核的实践路径

4.1. 智能评阅与反馈机制

智能评阅是 AI 赋能过程性考核的重要基础。相较于依据标准答案比对的客观题，程序设计题、算法设计题等主观任务解题路径多样、过程复杂，对评价的准确性与一致性要求更高。本系统在第 3.2 节所述四维评分量规与错误标签体系下，借助大语言模型对学生作答进行语义分析与逻辑判断，形成分维度评价结果。例如，在《数据结构》课程的二叉树遍历设计任务中，系统分别评价递归与非递归实现的正确性、边界条件处理与复杂度分析，避免仅依据最终输出是否正确而遗漏过程性信息。在此基础上，系统依据错误类型关联相应知识点的讲解、典型例题与学习资源，并将评价反馈与后续修改任务相衔接，形成“评阅–反馈–改进”的连续过程[16]。

在产教融合实践中，学院将部分企业真实项目进行教学化拆解并纳入课程案例库，为评价反馈和学习资源推送提供了相应素材。与单纯提供教材参考答案相比，工程案例能够为学生提供更加贴近实际任务情境的参照，使其进一步认识自身方案与实际工程要求之间的差距，也有助于教师将评价结果转化为更加具体的改进要求。

4.2. 全过程数据采集与分析

过程性考核贯穿课前、课中和课后各教学环节，因此，AI 赋能过程性考核需要对学生学习全过程中的学情数据进行连续采集与分析，为学情诊断和后续教学调整提供数据支持。在课程教学改革实践中，利用大数据、大模型等技术对学生学习行为与学习表现进行全过程分析，已成为支持教学管理与教学决

策的常见做法[16]。

在课程实施过程中，可通过课前预习测验了解学生对基础知识的掌握情况，课中记录课堂互动、随堂练习和小组讨论等环节的学习表现，课后持续跟踪作业、实验及项目完成情况。通过对不同教学环节学情数据的连续采集，可以较为完整地呈现与课程目标和能力达成相关的学习信息。例如，在《机器学习》课程中，课前可采集学生对 Python 基础和数学建模知识的预习情况，课中记录随堂编程练习，课后跟踪课程设计项目的完成进度及代码提交情况。

多源学情数据汇聚后，系统既可形成班级层面的知识点掌握情况和典型错误分布，也可生成学生个体学习报告。教师可以据此识别学习过程中存在的共性问题，并关注需要进行个别指导的学生；学生则可以通过学习报告了解自身在知识掌握和能力发展方面的优势与不足[16]。

在《人工智能导论》《计算机网络》等核心课程的试点中，系统依据诊断结果将学习支持划分为“基础补漏－巩固提升－拓展创新”三个层级，并配套相应习题、案例与拓展资源，结合学生学习过程数据形成动态学习画像，其分层逻辑见第 3.5 节。相关实践为计算机类课程开展差异化教学和过程性评价提供了具体参考，也为后续基于学情数据开展学习支持和教学反思提供了基础。

4.3. 多样化学习活动的 AI 支撑

过程性考核不仅需要关注学生最终完成的学习结果，还应尽可能呈现其分析问题和解决问题的过程。因此，AI 支持下的学习活动设计不应停留于形式上的丰富，而应关注不同活动对学生思维过程、知识运用和实践能力的呈现作用。情景对话、思维阶梯和代码共创等活动，可以在一定程度上记录学生解决问题的过程，为过程性评价提供更加丰富的信息[17]。

在“情景对话”中，教师设置具体问题情境，学生与 AI 进行多轮交互，系统据此记录其分析和解决问题的过程；“思维阶梯”则围绕核心问题设置不同认知层次，用以观察学生能否逐步推进问题解决。例如，在《计算机组成原理》课程的“CPU 性能优化”任务中，评价重点不再局限于某一孤立知识点，而是关注学生能否将 CPU 架构、指令集等相关知识置于同一问题情境中进行综合运用。

学院在《人工智能导论》《数据挖掘》等课程中设计了“代码共创区”和“先编后比”等学习活动。“代码共创区”主要由学生与 AI 协作完成编程任务，并在教师审核后将相关成果纳入课程案例；“先编后比”则引导学生先独立完成代码设计，再与 AI 生成的方案进行比较，在此基础上进行修改和完善。此类活动既能够保留学生独立思考的环节，也能够利用 AI 提供比较和反馈，为学生发现方案差异、反思自身思路提供支持[17]。

此外，在代码评审、项目答辩和系统演示等实践环节中，AI 还可辅助教师开展代码规范性检查、功能完整性验证和文档质量评估等工作，以进一步了解学生实际的程序设计能力。学院与华为、浪潮、数字日照等企业共建校企合作平台，并在相关实践中引入企业真实项目和产业导师，有助于增强考核内容的实践性与针对性，促进课程教学与行业实际需求之间的衔接。

5. 准实验研究的组织实施与效果

5.1. 研究设计与实施

为检验上述机制的实际效果，本研究以《数据结构》课程为载体开展准实验。本研究在同一教学周期内选取学习基础相近的教学班作为研究对象，分为实验组与对照组；两组均以组别指代，文中所引过程性资料不包含班级名称与学生个人身份信息。两组在课程目标、教学内容、教学进度、考核要求与教师组织等方面保持一致，差异仅体现在过程性考核的实施方式上：实验组引入“AI 辅助初评－诊断反馈－教师复核－学生修改”的评价机制，对照组沿用教师批阅与课堂讲评的常规方式。实施前，对两组学

生已有的学习基础与学习投入水平进行整体比较，以确认其具有基本可比性；实施过程中，持续记录评价效率、学习表现、学习投入与反馈利用等方面的信息，用以观察不同评价方式所产生的变化。

5.2. 数据来源与分析思路

本研究采用定量与定性相结合的混合研究方法。定量资料主要来自四个方面：其一，记录 AI 辅助初评与常规人工批阅的反馈周期，用以观察评价效率的变化；其二，比较过程性任务、实验报告与阶段性学习成果的完成质量，用以分析学生学习表现的差异；其三，借鉴学习投入的相关测量思路[20]，采用课程学习投入量表了解学生在行为、认知与情感投入方面的总体变化；其四，借助课程平台日志，观察学生查看反馈、修改作品与再次提交等学习行为。定量分析以描述性统计和组间总体比较为主，并关注变化的方向及其教学意义；相关数据均在聚合层面使用，不涉及班级与个体身份信息。定性资料包括师生半结构化访谈、课堂观察与学生反思日志，主要用于分析学生如何使用反馈，以及教师如何据此调整教学。资料收集前向参与师生说明了研究目的与数据用途，并取得知情同意；全部资料均经脱敏与匿名编号处理，不呈现可识别个人身份的信息，仅用于教学改进与研究分析。各评价维度的数据来源、比较方式、总体表现及教师作用如表 2 所示。

Table 2. Data collection dimensions and overall performance of the quasi-experiment
表 2. 准实验数据采集维度及总体表现

评价维度	主要数据来源	比较方式	总体表现	教师作用
评价效率	反馈时间、批阅过程记录	实验组与常规方式总体比较	首轮反馈周期缩短，教师重复性批阅减少	复核关键任务与异常结果
学习表现	过程性任务、实验报告、阶段性成果	组间总体比较	任务完成质量、规范性和问题修正完整度整体改善	确认最终成绩与关键诊断
学习投入	学习投入量表、课堂参与记录	关注行为、认知与情感投入变化	行为投入和认知投入呈积极变化	结合课堂表现解释量表结果
反馈利用	平台日志、修改记录、再次提交情况	观察学生是否依据反馈持续改进	查看反馈、修改和再次提交行为更加积极	判断修改是否真正解决问题
人机协同	AI 初评记录、教师复核记录	比较初评与教师最终判断的一致性	常规任务总体一致，复杂开放任务需教师修正	最终评价由教师确认

5.3. 实施效果

评价效率方面，引入 AI 辅助初评后，学生获得首轮反馈的周期有所缩短。教师不再需要对所有作业逐项完成重复性初评，而可以将更多精力用于关键问题复核、典型错误分析和后续教学调整。与常规评价方式相比，评价工作的重心逐步由重复性批阅转向诊断分析与专业判断，AI 在提高反馈时效方面表现出一定的辅助作用。

在学习表现上，实验组在过程性任务完成质量、实验报告规范性和问题修正完整度等方面总体表现较好。学生能够较及时地依据反馈补充和修正任务成果，从过程性任务完成情况看，作品的规范性和完整性呈现改善趋势。

从学习投入和反馈利用情况来看，实验组学生在获得诊断反馈后，更愿意重新检查代码、补充实验

报告并再次提交，对薄弱知识点、错误类型和修改过程的关注也有所增加。与只关注最终得分相比，部分学生开始更多地将评价信息用于后续学习和任务修订，这与形成性评价强调通过反馈促进学习改进的基本取向相一致。

在人机协同评价过程中，系统初评与教师专业判断在多数常规任务上的结论基本一致，但面对开放性表达、复杂算法思路或题意存在歧义的任务时，模型判断仍可能出现偏差。因此，本文始终将 AI 定位为教师评价的辅助工具：系统提供初步评分、错误标签和诊断线索，关键结论及最终成绩由教师复核确认。教师复核不仅是必要的质量控制环节，也体现了教师在教学评价中的专业责任。

5.4. 师生反馈

访谈、课堂观察和学生反思材料为前述结果提供了进一步说明。结合师生反馈可以发现，学生在任务完成后较快获得反馈，更愿意在思路尚清晰时继续检查和修改，而不是把一次提交视为任务的结束。诊断报告对薄弱知识点和常见错误的呈现，也使评价信息更容易转化为后续学习中的具体改进线索。对教师而言，整体学情信息有助于识别较为集中的学习困难，并据此调整讲解重点和练习安排。不过，对于开放性较强、判断依据较复杂的任务，教师的专业复核仍然不可替代。总体来看，定量资料所反映的变化趋势与访谈、观察等定性材料可以相互印证。

结合准实验资料和教学过程记录可以看出，AI 辅助过程性考核的价值并不在于替代教师完成评价，而在于缩短前期分析和反馈生成所需的时间，使教师能够把更多精力放在结果复核、共性问题分析和后续教学调整上。从本课程的实施情况看，学生获得反馈更为及时，对评价信息的利用也更加充分，部分过程性作品的规范性和完整性有所改善。与此同时，教师能够依据更连续的学情信息开展教学判断。上述变化说明，AI 更适合作为过程性评价中的辅助工具，其作用仍需建立在教师专业判断和持续教学观察的基础上。

6. 典型课程案例：以《数据结构》为例

本节以《数据结构》课程中“二叉树遍历”单元的一次程序设计作业为例，完整展示系统从 AI 辅助评阅、诊断报告、个性化反馈到教师教学反思的全过程。

6.1. 考核任务说明

该任务面向“树与二叉树”知识单元，要求学生根据给定的前序遍历序列和中序遍历序列重建二叉树，并输出前序、中序、后序和层序遍历结果；在此基础上，再实现后序遍历的非递归版本，并比较相关实现的时间与空间复杂度。任务涉及“二叉树定义”“遍历”“栈的应用”“复杂度分析”等知识点，能力要求主要涉及理解与应用，并包含一定的综合分析要求。

6.2. AI 辅助评阅与诊断报告

学生在线提交代码后，系统先通过预设单元测试检查可以客观判断的程序运行结果，再依据“思路设计－代码实现－结果验证－效率分析”四个维度，对学生作答进行辅助分析并形成初步评价、错误标签和相关诊断信息。这些信息与课程知识点建立联系后，形成个人学习诊断报告。需要说明的是，该报告主要为教师后续判断提供参考，涉及关键评分、错误归因和能力层级判定的内容，仍需由教师结合学生实际作答进行复核，复核后的结果再用于反馈和后续教学分析。

以一份经匿名化处理的代表性作答为例(仅用于说明系统的评阅与诊断过程，不涉及具体学生身份，也不作个体成绩比较)，学生能够完成二叉树重建和常规递归遍历，但在非递归后序遍历的实现中，对辅助栈的使用以及节点访问时机的处理还不够准确；复杂度分析也主要集中在时间开销，对空间复杂度的

说明不够充分。结合这些作答表现，系统将相关问题关联到“二叉树遍历”“栈的应用”和“复杂度分析”等知识点，并给出相应的能力表现提示。教师复核后认为，系统对主要问题的判断与实际作答情况基本一致，同时对个别表述性问题进行了修正。由此可以看出，AI 在快速整理作答信息和提示可能问题方面具有一定作用，但最终评价仍需依赖教师的专业判断。

6.3. 个性化反馈与学习资源

针对上述问题，系统根据学生的具体作答情况提供相应反馈。对于非递归遍历中出现的栈操作和访问时机问题，反馈首先引导学生回到相关知识点重新检查实现逻辑，并结合基础练习、相似案例和拓展任务帮助其理解错误原因、完成针对性修改。相关讲解材料也根据诊断结果进行匹配，供学生进一步查阅。所有推荐内容均由教师审核后再向学生开放。学生依据反馈重新检查代码并完成修改，后续提交中原有问题得到一定程度的纠正。从这一过程看，诊断反馈能够为学生后续修订提供较为明确的方向，其作用主要体现在帮助学生把评价信息转化为具体的学习行动。

6.4. 教师端数据与教学反思

教师端主要从整体学情出发呈现知识点掌握情况、能力表现和较为集中的错误类型，仅以聚合后的总体信息呈现，不提供个体排名。以本次任务为例，学生在递归遍历相关内容上的理解总体较为稳定，而在“递归过程向显式栈操作转换”和“空间复杂度分析”等内容上更容易出现共性问题。结合这些信息，教师对后续习题课作了相应调整，增加递归调用栈的可视化演示，并引入实际工程场景中的非递归遍历案例，同时安排针对性练习进行巩固。后续任务和课堂表现中，同类问题出现得相对减少，学生对相关知识的应用也更为完整。从这一过程看，系统汇总的学情信息不仅有助于教师发现学生学习中的共性困难，也为后续教学内容、讲解方式和练习安排的调整提供了较为具体的参考。典型任务各环节的系统与教师处理、案例表现如表 3 所示。

Table 3. The submission-diagnosis-feedback-review-improvement chain of a typical task
表 3. 典型任务“提交–诊断–反馈–复核–改进”链条

环节	系统与教师处理	案例表现
首轮提交	单元测试检查客观结果；大语言模型进行四维辅助初评；教师复核关键评分	递归遍历基本正确，非递归后序遍历和复杂度分析存在问题
知识诊断	将作答证据映射到“二叉树遍历”“栈的应用”“复杂度分析”等知识点	定位到栈操作、访问时机和空间复杂度说明等薄弱环节
个性化反馈	系统生成修改提示并匹配练习、案例和课程资源；教师审核后反馈	学生能够据此重新检查实现逻辑并完成针对性修改
再次提交	系统重新检测，教师查看修改结果	同类错误得到纠正，代码与说明更加完整
教师干预	汇总共性错误，调整习题课与案例讲解	增加递归转栈演示和工程化案例，强化薄弱知识点
后续观察	结合后续任务和课堂表现继续跟踪	相关知识应用更加完整，同类问题出现频率下降

该案例较为完整地呈现了 AI 辅助过程性考核在实际教学中的实施过程，包括学生提交、辅助评阅、知识诊断、反馈修改、教师复核以及后续教学调整。这里关注的并非某一学生或某一教学班的具体成绩

变化，而是通过具体任务说明，在教师主导下，系统如何对学生作答中分散的信息进行整理和分析，并将其转化为可供学生改进学习、教师开展教学判断的诊断依据。由此，任务评价不再止于一次性的结果判定，而是与反馈、修订和后续教学调整联系起来，体现出“任务－诊断－反馈－复核－改进”的基本过程。

7. 促进深度反思的教学策略

7.1. 基于学情报告的学生自评

过程性考核应帮助学生了解自身学习状况并调整学习方式。自我反思是培养学生自主学习能力的重要环节。AI 生成的过程性考核学情报告涵盖知识掌握、能力发展和学习行为等信息，可为学生开展自我评价提供相应的材料依据。在此基础上，教师仍需引导学生定期研读学情报告，将相关信息与课程目标及个人学习目标相联系，开展较为系统的自我评价，使自我评价贯穿整个学习过程[1]。

在具体实践中，学院采用“反思日志”的形式，引导学生依据学情报告开展结构化反思，主要围绕“当前学习优势是什么”“存在的主要问题有哪些”“下一步准备怎样改进”等问题展开。通过这一过程，学生由被动接收评价结果逐步转向主动审视和分析自身的学习过程与学习成果。结构化反思有助于学生逐步形成对自身认知过程进行监控和调节的能力，即元认知能力，并为自主学习与持续改进提供支持。例如，在《程序设计基础》《数据结构》等课程中，通过设置“反思提示”开展交互设计，如“若重新设计这段代码，你会做哪些优化？”“你的算法在效率与可读性上还有哪些提升空间？”等，引导学生说明代码的不足及修改理由。

7.2. 学情数据驱动的教学反思

教学反思是教师专业发展与自我提升的重要途径，也是推动计算机类专业教学改革持续推进和教学质量不断改进的重要保障。借助 AI 考核系统，教师可以获得班级整体及学生个体的学情分析报告，其中包括知识点掌握率、能力维度达成度和典型错误类型分布等信息。基于这些数据，教师能够对教学过程进行较为系统的回顾、分析与审视，及时发现教学环节中需要调整和改进之处，从而进一步优化教学安排，并增强教学反思意识与自我监控能力。

学校近期开展的一系列 AI 赋能教育教学培训中提出，教师应主动运用 AI 相关工具与方法，提升基于数据开展教学反思的能力，并将教学改革建立在更加充分的数据分析和实践依据之上。

例如，在《计算机网络》课程中，当系统数据显示“TCP/IP 协议”相关知识掌握率偏低时，教师可进一步反思讲解方式、例题选择及练习难度是否适当；若仅有少数学生持续出现同类问题，则可结合其个人学习记录进行进一步分析。在此基础上，教师可补充企业真实网络场景案例，并对练习内容进行重新组织，再通过后续数据观察相关指标是否发生变化。通过对多轮教学数据进行纵向比较，教师能够逐步将教学反思由经验判断转向基于数据的分析与改进，在一定程度上减少教学决策对个人经验的单一依赖，更加准确地评估教学改革的实际效果，提高教学改进的针对性，进而形成“设计－实施－评价－反思－改进”的持续改进过程。

7.3. 构建反思共同体

反思不仅是个体层面的学习活动，也需要通过群体交流进一步深化。构建反思共同体，是计算机类专业培养学生团队协作能力和交流表达能力的重要途径。依托 AI 平台搭建线上学习社区，可以为学生之间以及师生之间的反思交流提供支持。学生既可以在社区中分享学习心得和反思体会，也可以通过阅读他人的反思内容获得启发。教师还可定期组织反思交流活动，选取具有代表性的反思案例进行点评和讨

论，引导学生进一步深化对学习问题的认识。例如，在《软件工程》课程中，教师可组织学生交流课程设计项目中的反思体会，并围绕代码优化、项目管理等方面分享经验与不足[17]。

学院建设了跨地域虚拟教研室，组织多所高校计算机类专业教师围绕教学中的重点、难点问题开展共同研讨，并通过教学观摩、互评等方式相互学习和借鉴，逐步形成跨校际教研共同体。依托这一平台，教师能够在持续交流与实践反馈中加强教学反思，进而促进教学质量的改进。在此基础上，该模式也可为学生反思共同体的建设提供参考。借助“云课堂”和资源共享平台，可以进一步加强学生之间以及不同学校学生之间的交流互动，将个体层面的反思拓展为群体性的学习实践，使学生在交流与讨论中深化对学习问题的认识，增强反思的针对性和实效性。

8. 实践效果与展望

8.1. 相关教学改革成果

近五年来，学院持续推进教学改革与人才培养实践，并形成了一定的阶段性成果。相关成果包括省级教学改革项目 5 项、省级教学成果奖 4 项；在学科竞赛方面，蓝桥杯获国家级奖项 120 项，其他 A 类学科竞赛获国家级奖项 50 余项；在创新创业训练方面，累计立项省级大学生创新创业训练计划项目等 30 余项。

8.2. 局限与展望

结合本课程的实施情况，仍有几个方面有待继续深化。其一，认知诊断的准确性仍有提升空间。面对专业性较强、解题路径多样的计算机类问题，大模型的判断结果可能存在偏差，因此，人机协同复核仍具有必要性。其二，数据使用规范仍需进一步完善。学习过程数据涉及个人隐私，在采集范围、保存期限与使用主体等方面，宜进一步明确相应规范，并确保数据仅用于教学改进目的。其三，教师技术素养仍需进一步提升。这里所指的技术素养不仅包括相关工具的操作能力，也包括对模型局限性的认识以及对诊断结果进行合理解释的能力。

随着人工智能技术的持续发展，未来可进一步推动过程性考核与反思向智能化、个性化和融合化方向拓展。一方面，随着 AI 语义理解与推理能力的提升，其在复杂工程问题评价中的应用有望进一步提高评价的准确性与可靠性；另一方面，虚拟现实、增强现实等技术与 AI 的进一步融合，也将为过程性考核与反思提供更加多样的应用场景。

基金项目

山东省高等教育本科教学改革研究项目(Z2024260)。

参考文献

[1] 李志义, 黎青青. 过程性评价与形成性评价辨析——工程教育专业认证视角[J]. 高等工程教育研究, 2022(5): 6-11.

[2] Black, P. and Wiliam, D. (1998) Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, **80**, 139-144, 146-148.

[3] 姚丽丽, 曹阳, 李晓明, 张维冬. 基于 OBE 理念的计算机类专业课程过程化考核评价模型[J]. 计算机教育, 2025(8): 230-235.

[4] Corbett, A.T. and Anderson, J.R. (1995) Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modelling and User-Adapted Interaction*, **4**, 253-278. https://doi.org/10.1007/bf01099821

[5] Piech, C., Bassen, J., Huang, J., *et al*. (2015) Deep Knowledge Tracing. In: Cortes, C., Lawrence, N.D., Lee, D.D., *et al*., Eds., *Advances in Neural Information Processing Systems* 28, Curran Associates, Inc., 505-513.

[6] Leighton, J.P. and Gierl, M.J. (Eds.) (2007) Cognitive Diagnostic Assessment for Education: Theory and Applications.

Cambridge University Press.

[7] Brusilovsky, P. and Peylo, C. (2003) Adaptive and Intelligent Web-Based Educational Systems. *International Journal of Artificial Intelligence in Education*, **13**, 159-172. https://doi.org/10.3233/irg-2003-13(2-4)02

[8] VanLehn, K. (2006) The Behavior of Tutoring Systems. *International Journal of Artificial Intelligence in Education*, **16**, 227-265. https://doi.org/10.3233/irg-2006-16(3)02

[9] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., *et al*. (2023) ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, **103**, Article ID: 102274. https://doi.org/10.1016/j.lindif.2023.102274

[10] 祝智庭, 彭红超. 智慧学习生态系统研究之兴起[J]. 中国电化教育, 2017(6): 1-10+23.

[11] 余胜泉. 人工智能教师的未来角色[J]. 开放教育研究, 2018, 24(1): 16-28.

[12] 郑勤华, 陈耀华, 孙洪涛, 陈丽. 基于学习分析的在线学习测评建模与应用——学习者综合评价参考模型研究[J]. 电化教育研究, 2016, 37(9): 33-40.

[13] 杨昊, 朱烨, 沈艳, 陈敏. 智能时代下计算机专业能力导向教学评价体系研究[J]. 高教学刊, 2025, 11(34): 92-96.

[14] 雷玉霞, 赵景秀. 人工智能“多元化全程性阶梯式”教学研究与实践[J]. 创新教育研究, 2020, 8(3): 298-306.

[15] 王华, 李琪, 刘效武, 雷玉霞. OBE-CDIO 理念下计算机类专业“房子”型本硕贯通创新培养模式探索[J]. 创新教育研究, 2026, 14(2): 559-566.

[16] 唐满, 郭东恩, 詹英. “AI+”背景下计算机类课程过程性评价改革[J]. 计算机教育, 2025(10): 133-138.

[17] 陆丽娜, 谷学强, 刘鸿福, 陈少飞, 谢愈. 人工智能通识教育过程性考核改革创新[J]. 电气电子教学学报, 2025, 47(5): 32-34.

[18] de la Torre, J. (2011) The Generalized DINA Model Framework. *Psychometrika*, **76**, 179-199. https://doi.org/10.1007/s11336-011-9207-7

[19] Junker, B.W. and Sijtsma, K. (2001) Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory. *Applied Psychological Measurement*, **25**, 258-272. https://doi.org/10.1177/01466210122032064

[20] 罗燕, 海蒂·罗斯, 岑逾豪. 国际比较视野中的高等教育测量——NSSE-China 工具的开发: 文化适应与信度、效度报告[J]. 复旦教育论坛, 2009, 7(5): 12-18.