

# 大语言模型时代背景下数据库课程实验考核改革研究

李南沙, 霍纬纲\*

中国民航大学计算机与人工智能学院, 天津

收稿日期: 2026年9月9日; 录用日期: 2026年10月10日; 发布日期: 2026年10月16日

## 摘 要

针对大语言模型背景下数据库课程实验考核以结果正确为唯一依据、难以甄别学生真实编程能力的问题, 文章剖析了传统实验考核模式失效的表现与成因, 提出由过程证据留痕、现场代码答辩、抗模型化命题与达成度数据反哺构成的实验考核改革方案, 阐述了如何将评价重心由结果验收转向过程可信与能力可证, 并对改革后的考核环节及其权重构成进行系统设计, 最后通过现场考察、课业成绩与问卷调查三个方面说明教学实践效果。

## 关键词

大语言模型, 数据库课程, 编程能力, 评价改革, 实验考核

# Research on the Reform of Laboratory Assessment in Database Courses in the Era of Large Language Models

Nansha Li, Weigang Huo\*

School of Computer and Artificial Intelligence, Civil Aviation University of China, Tianjin

Received: September 9, 2026; accepted: October 10, 2026; published: October 16, 2026

## Abstract

To address the problem that laboratory assessment in database courses relies solely on the correctness of results and can hardly discriminate students' authentic programming ability in the era of

\*通讯作者。

文章引用: 李南沙, 霍纬纲. 大语言模型时代背景下数据库课程实验考核改革研究[J]. 教育进展, 2026, 16(10): 838-845. DOI: 10.12677/ae.2026.16102176

large language models, this paper analyzes the manifestations and causes of the failure of the traditional laboratory assessment mode, and proposes a reform scheme composed of the retention of process evidence, on-site oral defense of code, anti-model question design, and feedback from achievement data. It expounds how to shift the focus of assessment from result acceptance to credible process and demonstrable competence, systematically designs the reformed assessment components and their weightings, and finally illustrates the effect of the teaching practice through on-site examination, course performance, and questionnaire results.

## Keywords

Large Language Models, Database Courses, Programming Ability, Assessment Reform, Experimental Assessment

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

数据库课程是计算机类专业的核心基础课程,兼具理论性与实践性。其中,基于视图、存储过程、触发器等服务器端编程技术进行数据库应用实现,是课程要求学生掌握的一项关键实践能力。这类能力的培养难以仅靠课堂讲授完成,主要依托实验环节,通过反复的编写、调试与优化训练逐步形成,实验成绩也因此成为课程过程性考核与课程目标达成度评价的重要依据[1]。近年来,以 ChatGPT、DeepSeek 为代表的大语言模型(Large Language Model, LLM)快速普及,其强大的自然语言理解与代码生成能力,正在深刻改变学生完成实验任务的方式。学生只需用自然语言描述题目要求,模型即可直接输出语法正确、可运行的 SQL 语句与过程化代码[2]。这一能力在为学习提供便利的同时,也对以“提交正确代码即通过”为标准的传统实验考核模式构成了直接冲击。当代码能否正确运行不再能够区分“学生自己写的”与“模型代写的”时,实验成绩便可能与学生的真实编程能力相背离,进而使建立在成绩之上的课程目标达成度评价失去可信度[3]。

围绕大模型与课程教学的融合,已有学者开展了有益探索。文献[4]将大语言模型技术引入数据库课程教学,以提升学生学习效率并培养创新思维;文献[5]提出基于大模型智能体的数据库课程过程性评价方法,动态生成适配性评测任务并结合学生画像实时反馈;文献[6]则在程序设计类课程中设计了多点累加式的编程能力评价体系。上述研究多聚焦于如何利用大模型优化教学与评价手段,而对大模型引发的实验考核失真问题本身,特别是如何重构实验考核以恢复其对真实编程能力的甄别力,仍缺乏深入探讨。

基于此,本文以数据库课程实验环节为切入点,剖析大模型背景下传统实验考核模式失效的表现与成因,提出面向真实能力甄别的实验考核改革方案,并结合教学实践检验其成效,以期同类实践性课程的考核改革提供参考。

## 2. 数据库课程实验考核的现状与问题分析

### 2.1. 传统实验考核模式及其局限

数据库课程理论与实践并重,其中视图、存储过程、触发器等服务器端编程能力的培养,主要依托实验环节,通过反复的编写、调试与优化训练逐步形成。长期以来,实验考核普遍采用任务驱动、结果验收的组织方式,教师针对各知识单元布置实验任务,学生完成后提交可运行的代码及实验报告,教师

依据代码能否得到正确结果给出成绩，实验成绩进而构成课程过程性考核与课程目标达成度评价的重要数据来源。这一模式建立在两个隐含假设之上，即代码运行正确等价于学生具备相应能力，且所提交的代码确为学生独立完成。在缺乏外部辅助的条件下，上述假设大体成立，学生若能独立提交正确代码，通常意味着其已完成了必要的分析、编码与调试训练。然而，该模式仅关注最终提交的代码结果，既不追踪代码的形成过程，也缺乏对代码归属真实性的验证机制。这一固有缺陷在传统环境下并不突出，但在能够代写代码的外部工具出现后被急剧放大，两个前提假设随之同时失效。

## 2.2. 大语言模型对实验考核效度的影响

大语言模型能够依据自然语言需求直接生成结构完整、可正确运行的 SQL 语句与过程化代码，并可根据报错信息自动修正，几乎能够独立完成传统实验任务。在此条件下，部分学生逐渐将大模型视为答案生成工具而非学习辅助资源，在实验环节直接采用模型输出以应付提交，跳过了自主分析、编码、调试与优化等关键训练环节[5]。这种依赖不仅弱化了学生真实编程能力的养成，也使实验成绩偏离了其应有的评价意义。为客观了解这一现象的普遍程度，面向已修读数据库课程的学生开展了匿名问卷调查，重点考察其在实验环节使用大语言模型的方式及自我能力认知，调查结果如表 1 所示。

**Table 1.** Survey results on students' use of large language models in laboratory sessions  
**表 1.** 学生在实验环节使用大语言模型情况的问卷调查结果

| 调查项目                  | 选择该项的学生占比/% |
|-----------------------|-------------|
| 经常直接使用大语言模型生成实验代码     | 68.5        |
| 偶尔使用大语言模型辅助完成实验       | 24.6        |
| 基本依靠自己编写实验代码          | 6.9         |
| 认为借助大语言模型能够顺利完成实验任务   | 85.2        |
| 承认脱离大语言模型后难以独立完成同类编程题 | 61.3        |
| 认为当前实验成绩能够真实反映自身编程能力  | 27.4        |

由表 1 可见，实验环节的“高通过率”与学生对自身真实能力的低认可度之间存在明显背离。当可借助大模型的实验环节普遍呈现高完成度、而脱离大模型的独立编程能力却明显不足时，过程性实验成绩便与学生的真实编程水平相背离，形成过程性成绩虚高、终结性能力偏弱的倒挂格局。由于实验成绩是课程目标达成度评价的重要支撑，其失真会沿评价链条向上传导，导致以过程性成绩为依据的课程目标达成度被系统性高估，从而掩盖真实存在的教学问题，削弱课程质量评价的可信度。

因此，大模型背景下数据库课程实验考核所面临的核心挑战，并非学生能否使用大模型，而是考核仅以结果正确为评判依据、缺乏对代码生成过程与归属真实性的验证，致使实验成绩效度下降、能力评价失真。破解这一问题的关键，在于将实验考核的重心由验收结果转向甄别真实能力。

## 3. 数据库课程实验考核改革方案设计

### 3.1. 改革总体思路与基本原则

针对上述效度失真问题，本文提出实验考核改革的总体思路：不以禁止学生使用大语言模型为前提，而是通过考核方式的重构，使依赖模型代写难以获得高分、真正理解并能够独立实现方可获得高分，从而

引导学生将大语言模型作为学习辅助资源而非答案生成工具加以运用。依据这一思路，确立三项基本原则。其一为过程可信原则，即评价不应仅依据最终代码，而应关注代码的形成过程，并保证过程证据的真实可追溯。其二为能力可证原则，即学生须能对其提交的成果作出解释与辩护，以证明成果确为自身能力的体现。其三为数据反哺原则，即将课程目标达成度数据由事后统计结果转化为过程改进信号，使考核改革形成可持续迭代的闭环。基于上述原则，构建由过程证据留痕、现场代码答辩、抗模型化命题与达成度数据反哺四个环节组成的实验考核改革方案，各环节所针对的问题、主要举措与预期作用如表 2 所示。

Table 2. Overall framework of the reformed laboratory assessment for the database course  
表 2. 数据库课程实验考核改革总体框架

| 改革环节    | 所针对的问题           | 主要举措                | 预期作用             |
|---------|------------------|---------------------|------------------|
| 过程证据留痕  | 仅验收结果，无法追溯代码形成过程 | 提交阶段成果、调试记录与使用声明    | 抑制只交结果、跳过训练的代写行为 |
| 现场代码答辩  | 无法验证代码的真实归属      | 就提交代码进行随机追问         | 验证成果确为学生自身能力的体现  |
| 抗模型化命题  | 传统题型可被模型直接解答     | 设计纠错、解读、迁移类题型       | 使评价指向学生真实的认知过程   |
| 达成度数据反哺 | 评价缺乏持续改进机制       | 建立数据诊断 - 调整 - 再诊断闭环 | 保持考核对模型演进的长效适配   |

3.2. 面向过程可信与能力可证的考核环节重构

传统实验考核的效度失真，根源在于其仅以最终代码是否正确为评判依据，而对代码的形成过程与真实归属缺乏约束。为此，改革首先从考核环节入手，通过过程证据留痕、现场代码答辩与抗模型化命题三项举措，将评价重心由结果验收转向过程可信与能力可证。

在过程证据方面，将实验环节的评价依据由提交结果转向留存过程证据。学生在提交代码的同时须提交实验过程记录，包括关键节点的阶段性成果、调试过程中遇到的错误及其修正思路，并附具大语言模型使用声明，如实说明在何处使用了大语言模型、获得了何种反馈以及哪些部分为独立完成。评分时将过程是否可解释、可追溯作为过程性成绩的前置条件，对仅提交正确结果而无法提供合理过程证据者不予高分，从而抑制只交结果、跳过训练的代写行为。

Table 3. Design of anti-model question types for laboratory assessment  
表 3. 面向实验考核的抗模型化题型设计

| 题型      | 示例                            | 设计意图          |
|---------|-------------------------------|---------------|
| 纠错诊断题   | 给出含错误的存储过程，要求定位错误、解释成因并说明修改思路 | 考查调试能力与原理理解   |
| 执行计划解读题 | 提供查询的执行计划，要求解读并提出优化方案         | 考查查询优化的分析判断能力 |
| 变式迁移题   | 在已讲授的业务场景上改变约束规则，要求重新设计代码     | 考查知识迁移而非范式套用  |
| 人机对比批判题 | 提供大语言模型生成的代码方案，要求指出其局限并加以改进   | 考查批判性采信与高阶思维  |

在命题设计方面, 针对大语言模型能够直接产出标准答案而使传统题型失效的问题, 对实验任务与阶段测验的命题方式进行重构, 将考核重心由生成正确代码转向理解、判断与迁移, 设计一批抗模型化题型, 如表 3 所示。此类题型的共同特征在于, 即便学生借助大语言模型获得初步结果, 仍须在理解的基础上完成诊断、解释、辨析或迁移, 无法通过简单复制模型输出而得分, 从而使评价重新指向学生真实的认知过程。

综合上述举措, 改革后数据库课程实验考核评价体系的构成如表 4 所示。各环节在权重分配上向可验证真实能力的部分倾斜, 使实验成绩能够更准确地反映学生的真实编程水平。

为保证过程证据评估与现场答辩评分的一致性、透明性与可重复性, 避免评价因教师主观差异而失准, 本文进一步设计了统一的评分量规。过程证据环节与现场答辩环节的评分量规分别如表 5、表 6 所示, 每一环节均从若干维度分级赋分, 教师依据可观测的行为表现对照量规打分, 从而使“过程可信”与“能力可证”的评价要求落实为可操作的评分标准。

**Table 4.** Structure of the reformed laboratory assessment system for the database course  
**表 4.** 改革后数据库课程实验考核评价体系构成

| 考核环节 | 主要考核内容         | 抗模型化设计       | 权重/% |
|------|----------------|--------------|------|
| 实验过程 | 阶段成果、调试记录、使用声明 | 过程留痕与可解释性评分  | 30   |
| 现场答辩 | 就提交代码接受随机追问    | 现场作答, 验证真实归属 | 20   |
| 阶段测验 | 纠错、执行计划解读、变式迁移 | 抗模型化题型       | 20   |
| 综合项目 | 真实场景数据库设计与开发   | 过程留痕与答辩追问    | 30   |

**Table 5.** Scoring rubric for the process-evidence component  
**表 5.** 过程证据环节评分量规

| 评价维度    | 优秀(90~100)           | 合格(60~89)      | 不合格(0~59)     |
|---------|----------------------|----------------|---------------|
| 阶段成果完整性 | 关键节点成果齐全, 能反映完整开发过程  | 成果基本齐全, 个别节点缺失 | 仅有最终代码, 无阶段成果 |
| 调试记录真实性 | 错误与修正过程详实、逻辑连贯, 可追溯  | 有调试记录但较简略      | 无调试记录或明显拼凑    |
| 使用声明规范性 | 如实、具体说明模型使用范围与独立完成部分 | 声明较笼统          | 未声明或声明不实      |
| 独立完成程度  | 核心逻辑由本人独立完成, 模型仅作参考  | 部分依赖模型, 但有自主加工 | 整体依赖模型输出      |

**Table 6.** Scoring rubric for the on-site oral defense component  
**表 6.** 现场代码答辩环节评分量规

| 评价维度   | 优秀(90~100)      | 合格(60~89)     | 不合格(0~59)  |
|--------|-----------------|---------------|------------|
| 解题思路阐述 | 思路清晰完整, 能说明设计依据 | 思路基本清楚, 略有含糊  | 无法说明整体思路   |
| 代码逐句解释 | 准确解释关键语句含义与作用   | 能解释多数语句, 个别不清 | 难以解释代码执行逻辑 |
| 变式追问应对 | 能就约束变化给出合理调整方案  | 经提示后可作答       | 无法应对追问     |
| 表达与逻辑  | 表达流畅、条理清晰       | 表达基本清楚        | 表达混乱、逻辑不清  |

### 3.3. 持续改进机制

考核环节的重构解决了成绩效度问题,但要使改革保持长效,还需建立以数据为驱动的持续改进机制,将课程目标达成度数据由事后统计结果转化为过程改进信号。在诊断出编程能力相关目标存在过程性与终结性得分倒挂后,一方面收紧过程性环节的证据要求与答辩比重,另一方面提高编程类终结性考核的权重与区分度。每一教学周期依据各环节成绩分布、试卷各部分得分率及目标达成度拆解结果,动态调整实验任务设计、题型配比与评价标准,形成数据诊断、方案调整与再诊断的迭代闭环,使实验考核能够持续适配大语言模型不断演进所带来的新变化。

### 3.4. 实施保障:教师培训与技术支持

改革方案涉及过程证据评估、现场答辩与抗模型化命题等新型考核方式,对教师的评价能力与实施成本提出了更高要求,需要相应的实施保障予以支撑。

在教师培训方面,建议以课程组为单位组织统一培训,围绕表5、表6所示评分量规开展集体研讨与试评,通过对同一批学生成果的联合评分校准,统一各教师对量规各等级的把握尺度,减小主观差异;同时建立抗模型化题型的共享题库,由课程组共同命题、共同评议,既降低单个教师的命题负担,又保证题目的质量与难度一致性。

在技术支持方面,“过程证据留痕”若完全依赖学生手工提交,将增加师生负担,可借助能够自动记录编程过程的工具加以降本增效。例如,采用支持代码版本管理的在线编程平台或在线判题系统,如希冀实验平台,自动留存学生的代码提交历史与调试轨迹,采用课程管理平台归档阶段成果与大语言模型使用声明,从而使过程证据的采集由“事后补交”转变为“过程自动沉淀”,在减轻实施成本的同时提升证据的真实性与可追溯性。

## 4. 教学实践与成效分析

为初步检验改革方案的可行性,本文采用准实验设计,选取两个自然教学班作为研究对象,共80名学生。两班在生源基础、先修课程成绩及授课教师、教学内容、课时安排上基本一致,具备可比性,其中一个班(40人)作为实验组,按本文改革方案组织实验考核,另一个班(42人)作为对照组,沿用传统的结果验收式考核。研究选取存储过程与触发器两个编程任务较为集中的单元,按照改革方案的思路组织实验考核,即要求学生在提交代码的同时留存调试记录并作出大语言模型使用声明,随后在实验课上对部分学生进行现场考察,请其讲述解题思路、逐句解释所提交代码的含义,并回答约束条件变化时应如何调整代码等追问。同时,在试点单元采用纠错诊断、变式迁移等抗模型化题型替代传统的直接编程题,教师依据前文表5、表6所示评分量规对过程证据与答辩表现进行评分,以保证评价的一致性。试点结束后,通过现场考察记录、课业成绩对比与小范围问卷调查三个方面对成效进行分析。

从现场考察的情况看,改革方案能够较为有效地甄别学生对实验内容的真实理解程度。在传统考核方式下,学生提交的代码大多能够正确运行,彼此差异不大;而在现场考察中,学生的表现则呈现出明显分化。一部分学生能够清晰地阐述解题思路,准确解释游标的作用、事务的边界以及触发器的触发时机,并能针对追问给出合理的调整方案,表明其确实掌握了相应的编程技能;另一部分学生虽然提交了可正确运行的代码,却难以说明代码的执行逻辑,暴露出其代码在很大程度上依赖模型代写、缺乏真实理解。据统计,实验组约三成学生在现场考察中难以完整解释自己所提交的代码,而这一问题在传统的结果验收环节完全无法察觉。这种在结果验收环节难以察觉、却在现场考察中清晰显现的能力差异,说明现场代码答辩确实能够穿透“代码正确”的表象,还原学生的真实编程水平。

从课业成绩看,试点单元的成绩分布较传统考核方式发生了合理的变化。传统方式下,实验环节因

代码普遍能够正确运行, 学生得分集中于高分段, 区分度很低; 引入过程证据与现场考察后, 试点单元的实验成绩不再普遍性地集中于高分, 成绩分布明显拉开, 能够真实反映学生之间编程能力的差异。与仍采用传统考核的对照组相比, 对照组实验成绩仍普遍集中于高分段, 而实验组成绩分布更为分散、区分度明显更高; 在随后不允许使用大模型的闭卷编程环节中, 实验组的平均得分与及格率均高于对照组, 高分段“虚高”现象得到明显缓解。更值得关注的是, 经过试点单元的训练与考核, 学生在后续闭卷编程环节的表现有所改善, 逻辑不清、语法错误频发的现象较此前有所缓解。究其原因, 在于过程证据留痕与现场考察的约束促使学生不再满足于直接复制模型输出, 而是重新投入到自主分析、编写与调试之中, 这一转变反过来促进了其真实编程能力的提升。

从学习行为看, 改革方案对学生使用大语言模型的方式产生了积极的引导作用。为了解试点单元中学生学习行为与态度的变化, 面向参与试点的学生发放了小范围问卷, 结果如表 7 所示。数据显示, 多数学生认可现场考察与抗模型化命题对甄别真实能力的作用, 认为这种考核方式使自己必须真正理解代码才能获得较好成绩; 超过七成的受访学生表示, 在明确实验考核需要接受现场追问后, 会更主动地弄清每一条语句的含义, 对大语言模型的依赖有所减少。这表明, 将评价重心由结果验收转向过程可信与能力可证, 能够有效引导学生从“直接获取答案”转向“理解并独立实现”。与此同时, 也有部分学生反映现场考察在一定程度上增加了学习压力, 个别学生对答辩形式存在紧张情绪, 这提示在后续推广中需要合理设计考察的方式与节奏, 在保证甄别效度的同时兼顾学生的接受度。

Table 7. Main results of the questionnaire

表 7. 问卷调查主要结果

| 调查项目                     | 持认同态度的学生占比/% |
|--------------------------|--------------|
| 认为现场考察能够真实反映自身对代码的理解     | 84.6         |
| 认为抗模型化题型无法依靠复制模型输出直接完成   | 80.8         |
| 表示会因需接受现场追问而更主动理解每条语句    | 73.1         |
| 认为改革后的考核方式减少了自己对大模型的直接依赖 | 69.2         |

综合现场考察、课业成绩与问卷调查三个方面的结果可以看出, 尽管试点范围有限, 但改革方案在甄别真实编程能力、抑制模型代写、引导学生回归自主训练方面已初步显现成效, 验证了其可行性与有效性。当然, 本次试点样本规模较小、周期较短, 所得结论更多具有探索性质, 方案的系统性成效仍有待在更大范围、更长周期的教学实践中进一步检验; 同时, 现场答辩与抗模型化命题的组织成本较高, 其在大班教学中的适用性也需要进一步优化。

## 5. 结语

大语言模型对课程考核的冲击, 不仅在于学生可能借助模型完成本应独立完成任务, 更深层的影响在于其使以结果正确为依据的实验考核丧失了甄别真实能力的效度, 进而导致课程目标达成度这一质量评价关键指标被系统性高估。本文以数据库课程实验环节为切入点, 分析了传统实验考核模式在大语言模型背景下失效的表现与成因, 提出了由过程证据留痕、现场代码答辩、抗模型化命题与达成度数据反哺构成的改革方案, 将实验考核由结果验收转向过程可信与能力可证。小范围试点表明, 该方案能够有效甄别学生的真实理解程度, 引导学生减少对模型代写的依赖、回归自主编程训练, 并在一定程度上促进了编程能力的提升, 初步验证了其可行性与有效性。

## 基金项目

高校教育教学改革与研究项目“大语言模型赋能数据库系统原理课程实践教学改革研究”，项目编号：CAUC-2026-C2-12。

## 参考文献

- [1] 陈红, 卢卫, 杜小勇. “101 计划”数据库课程的改革探索与实践[J]. 计算机教育, 2023(11): 22-28.
- [2] 兰国帅, 杜水莲, 宋帆, 等. 生成式人工智能教育: 关键争议、促进方法与未来议题——UNESCO《生成式人工智能教育 and 研究应用指南》报告要点与思考[J]. 开放教育研究, 2023, 29(6): 15-26.
- [3] 陈凯泉, 韩小利, 郑湛飞, 等. 人机协同视阈下智能教育的场景建构及应用模式分析——国内外近十年人机协同教育研究综述[J]. 远程教育杂志, 2022, 40(2): 3-14.
- [4] 王晓玲, 岳文静, 王昊奋, 等. 大语言模型技术融入数据库课程的教学探索[J]. 计算机教育, 2024(9): 28-32.
- [5] 李希, 姜火文, 戴航, 等. 基于大模型智能体的数据库课程过程性评价设计与实践[J]. 计算机教育, 2026(6): 57-62.
- [6] 陈娟, 邓春燕, 吕帅, 等. 程序设计基础课程形成性评价的设计与实践[J]. 计算机教育, 2023(3): 61-64.