

高中生物学混合式组卷路径探索

蒯 龙¹, 田慧敏¹, 韩殿国², 徐立新¹, 李长远¹, 郭 成^{1*}

¹赤峰学院化学与生命科学学院, 内蒙古 赤峰

²赤峰学院附属中学, 内蒙古 赤峰

收稿日期: 2025年10月17日; 录用日期: 2025年11月17日; 发布日期: 2025年11月24日

摘 要

随着教育信息化的深入发展, 组卷方式呈现多元化趋势。本文通过对人工组卷、高中生物题库组卷和AIGC智能组卷三种模式的深入分析, 比较了各自的优势与不足, 并探讨了混合组卷模式的优势与发展, 并以生成式人工智能软件Claude组卷为例进行实践。认为, 传统人工组卷在个性化和教学针对性方面不可替代, 题库组卷在标准化和资源共享方面具有优势, AI组卷在效率和客观性方面表现突出。基于IAGC高中生物学混合组卷模式能够有效整合三种方式的优点, 提高组卷效率及质量, 为一线教师进行教育评价提供更高效率的解决方案。

关键词

传统人工组卷, 试题库组卷, AIGC组卷, 混合模式, 高中生物学

Exploration of Hybrid Test Paper Composition Pathways for High School Biology

Long Kuai¹, Huimin Tian¹, Dianguo Han², Lixin Xu¹, Changyuan Li¹, Cheng Guo^{1*}

¹College of Chemistry and Life Sciences, Chifeng University, Chifeng Inner Mongolia

²Affiliated High School of Chifeng University, Chifeng Inner Mongolia

Received: October 17, 2025; accepted: November 17, 2025; published: November 24, 2025

Abstract

With the deepening development of educational informatization, test paper composition approaches exhibit diversified trends. This paper conducts an in-depth analysis of three modalities: manual test paper composition, high school biology question bank-based composition, and AIGC intelligent test paper composition. The study compares their respective advantages and limitations, explores the

*通讯作者。

文章引用: 蒯龙, 田慧敏, 韩殿国, 徐立新, 李长远, 郭成. 高中生物学混合式组卷路径探索[J]. 教育进展, 2025, 15(11): 1431-1440. DOI: 10.12677/ae.2025.15112185

merits and development of hybrid composition models, and provides practical implementation using the generative artificial intelligence software Claude as an exemplar for test paper composition. The findings indicate that traditional manual composition remains irreplaceable in terms of personalization and pedagogical targeting, question bank-based composition demonstrates advantages in standardization and resource sharing, while AI-based composition exhibits outstanding performance in efficiency and objectivity. The AIGC-based hybrid test paper composition model for high school biology can effectively integrate the advantages of all three approaches, enhancing both the efficiency and quality of test paper composition, thereby providing frontline educators with more efficient solutions for educational assessment.

Keywords

Traditional Manual Test Paper Composition, Question Bank-Based Test Paper Composition, AIGC Test Paper Composition, Hybrid Model, High School Biology

Copyright © 2025 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

教育评价是教育活动中的重要环节，是教育活动的“指挥棒”[1]。2019年6月国务院办公厅颁布了《关于新时代推进普通高中育人方式改革的指导意见》，指出要“加强命题能力建设”，提出新时代评价育人的新要求。2020年10月，国务院印发的《深化新时代教育评价改革总体方案》提出要完善教师参与命题的相关机制。考试命题设计能力重要性不言而喻，考试命题设计能力是教师必备的六大能力之一[2]。这一系列政策的提出对中小学教师命题能力提出了新要求，在日常测试方面，多数中小学教师对命题技术的运用程度比较薄弱[3]，命题工具、组卷工具知之甚少，组卷作为教育评价的核心技术，其质量直接影响着教学效果的检验和学生学习成果的评估。

高中生物学作为自然科学的重要学科，其知识体系复杂、概念抽象、实验性强，对组卷的科学性和合理性提出了更高要求。如何在保证试卷质量的前提下，提高组卷效率，减少教师工作负担，已成为当前教育技术研究的热点问题。本文旨在通过对三种主要组卷方式的比较分析，探讨各自的优势与局限性，并根据当前国内主流组卷平台的发展特点提出混合组卷模式的优势、实践策略。

2. 三种组卷方式及特点

试卷组配简称组卷或配题，指根据考试需要，按照命题细目表的设计，从已审试题中选择合适试题，按照题型、内容或难易顺序等试卷编排规范，组配成一套试卷的试题[4]，目前，常见组卷方式为传统人工组卷和生物题库组卷，随着生成式人工智能(AIGC)的发展，出现了AIGC组卷。

2.1. 传统人工组卷

传统的教师出卷方式主要是由出题者根据自己的知识收集、选取试题、编制试题，其特点是教师可根据教学目标、学生特点和自己的专业判断，手工选择、编写和组织题目的方式，有经验的教师能够创造性地设计题目，将最新科学发现、生活实例或时事热点融入其中，增加题目的新颖性和吸引力，教师出卷方式主要是由出题者根据自己的知识收集、选取试题、编制试题[5]。张玉强等(2025)在8403名普通教师调查问卷中随机抽取3165份分析，发现对于“您在日常测验中如何命题？”这一问题的回答中，

67.25%的教师选择教辅资料中的试题, 43.97%的教师选择直接下载套卷, 45.06%教师凭借经验命题[5]。

人工组卷在实际应用中面临诸多制约因素。效率低下构成核心问题, 设计高质量试卷耗时较长, 新教师往往需要投入数倍于教学的时间来准备试卷, 在教师工作负担繁重的现实下格外突出。组卷质量波动明显, 高度依赖教师个人素养与当时状态, 经验丰富的教师能够设计出优质试卷, 而缺乏经验或状态欠佳的教师在试卷构建上常现不足, 质量稳定性难以保障。主观色彩浓厚带来公平性隐患, 教师可能无意中偏向特定知识点或题型, 个人喜好会影响命题选择, 不同教师的评判标准存在差异, 这些因素都可能损害评价的公正性。资源配置效率偏低的问题也十分明显, 各教师重复进行相似的组卷工作, 形成大量重复劳动, 而优秀教师积累的组卷智慧和成果缺乏有效的传播共享机制。科技进步和课程变革要求教师不断更新知识储备, 但人工组卷模式下的知识更新往往呈现分散化、非系统化特征, 更新过程存在明显困难。

2.2. 题库组卷系统

题库组卷主要从已建题库中根据考试内容、章、节、题目类型、难度、时间等选题自动生成满足教学和教师要求的各类试题。具体处理方法是在题库中随机搜索满足条件的试题, 直到试题总数满足要求。题库组卷系统的研究起步较早, 主要集中在题库建设、题目分类和组卷算法等方面。金惠云(1999)利用遗传算法能解决上述第一种方法的盲目随机性, 并能从群体中选择更满足条件的个体, 相对来说具有较强的智能性[6]; 目前智能组卷算法主要有遗传算法、随机抽取法、回溯试探法, 王友仁等(2006)提出了一种基于自适应混合进化机制(Adaptive Hybrid Evolutionary Mechanism, AHM)的智能组卷方法解决这些算法常因约束条件的局部满足而导致组卷失败, 或选取试题缺乏随机性, 组卷过程所需时间也较长, 无法满足实时应用等问题[7]; 国际学生评估项目(Programme for International Students Assessment, 简称 PISA)运用了“平衡不完全矩阵取样设计(Balanced Incomplete Matrix Sampling Design)使得试题考查的知识点覆盖面更广, 测试结果更准确, 但是对教师的心理计量学的知识要求比较高。平衡指的是任何两个试题组都会在某个试题本中同时出现一次, 不完全指的是同一个试题本中不可能包含所有的试题组”(陆璟, 2013)。从以上的研究中不难发现, 基于题库组卷的方式也在不断发生变化, 即智能技术不断整合到题库系统或者题库应用软件中。

高中生物题库组卷系统是基于预先建立的题库进行组卷的教学辅助工具。当前国内市场上的主要组卷平台各具特色且功能日趋完善, 主要有二一组卷、e 卷通、猿题库、菁优网等, 各平台组卷功能亦各具特色, 但从各题库更新迭代版本来看, 呈现一个特点就是覆盖学科类型、考区范围、受众学生越来越大, 并且逐渐在系统中加入 AI 功能模块。新技术的应用有效减少了教学组卷中的重复性低效工作, 教师可以对生成的试卷进行二次修改编辑, 既保证了效率又保持了组卷的自由性和选择性。

虽然题库组卷备受欢迎, 但题库组卷在实际应用中面临诸多限制。题库中的题目往往按照一般性教学需求设计, 很难满足特定班级或教学阶段的个性化要求, 教师常常需要根据学生实际情况对题目进行调整, 这在很大程度上影响了题库的直接使用效果。长期依赖题库容易让教师形成惯性思维, 减少了原创题目的开发, 使试题形式趋于固化, 难以跟上教学创新的步伐; 虽然题库具备更新功能, 但更新速度往往跟不上教学发展的实际需要, 新的教学理念、实验技术和科学发现无法及时融入题库内容, 导致题库的时效性和实用性受到影响。

2.3. 生成式人工智能(AIGC)组卷

生成式人工智能(AI-Generated Content, 以下简称 AIGC)是一类能够创造新内容的人工智能技术, 它可以根据输入的提示或指令生成各种形式的原创内容。人工智能在教育领域的应用始于 20 世纪 80 年代, 随着机器学习、自然语言处理、知识图谱等人工智能技术的发展, 通过算法分析题目的知识点、难度、

题型、认知层次等特征,根据预设的组卷目标和约束条件自动生成试卷,它不是简单的知识点堆砌,而是基于科学算法模型,能够精准控制试题知识点分布、难易程度、分值分布等多个维度,满足线上模拟考试、课堂测试等多种应用场景的需求。王志军等(2019)提出了基于深度学习的智能组卷算法,通过分析题目的语义特征和难度分布,实现了自动化组卷。李晓明(2020)研究了基于知识图谱的智能组卷系统,能够根据学科知识结构自动生成符合教学目标的试卷;王磊(2025)将 Hilbert Huang 算法,即一种处理非线性和非平稳号并进行模态分析的方法,通过信号分析和人工智能技术实现高效语言评测和试卷生成[8];随着新技术在各行业领域的渗透,如今,国外 Claude、国内 DeepSeek 等 AI 软件具备组卷的功能。AIGC 技术能够根据学生的学习进度和理解能力智能生成符合其需求的教学内容和练习题,从而有效提升教学效果[9]。

AI 智能组卷是基于机器学习、自然语言处理、知识图谱等人工智能技术,有以下几个优势:首先,相较于传统人工组卷,AI 组卷在效率方面能够在几分钟内完成原本需要数小时甚至数天的工作,并可同时生成多个版本满足不同需求;其次,AI 系统严格按照预设算法和标准执行,有效避免了教师主观偏好、情绪状态和时间压力等人为因素的干扰,确保了组卷的客观性和一致性;最后,AI 系统强大的大数据处理能力使其能够同时考虑并优化难度分布、知识点覆盖、题型平衡等多个复杂因素,实现人工组卷难以达到的多目标协调,并具备自适应学习能力,通过分析学生答题数据和教师反馈持续优化组卷策略。

AI 组卷的局限性。AIGC 是能通过分析大量训练数据进行学习并处理和生成用户请求的相关内容,有时生成看似合理但实际上错误、虚构的信息,这些内容可能表现为编造事实、虚构数据、引用不存在的文献,甚至创造出逻辑矛盾的描述,即 AI 幻觉[10];另外,图表作为高考生物学试题常用的信息载体,从图表中获取和转化出关键信息是认识和分析问题的前提,图表中往往蕴含着解题的关键信息,学会识图(表)、读图(表)至关重要,这种信息处理能力是高考考查的重要目标之一[11]。但国外的 ChatGPT、Claude 还是国内 DeepSeek、Kimi,曲线图生成困难,如突触结构图,或生成图示指向不明或只有文字描述,如遗传谱系图,这是人工智能自身技术问题造成,相信随着技术的发展,其多模态转换能力会得到增强;AI 系统虽能处理大量数据,但缺乏教师在组卷时所具备的教育直觉和专业判断,难以充分考虑学生具体学情、近期教学重点和学习氛围等微妙因素。目前 DeepSeek Chat (包括 DeepSeek-V3)不支持直接生成或绘制图表、曲线图等可视化内容,遗传谱系图等可通过绘图软件等来间接实现,另外 DeepSeek 可提供可以提供文本和代码形式在其他环境中生成简单图形。

生成式人工智能对高中生物学教学领域中简单问题、复杂问题、情境问题、实验方案设计的解答,发现生成式人工智能对简单问题、复杂问题回答较好,很难对新情境问题进行正确解答[12]。AI 工具由于依赖海量数据训练,在处理复杂语境或需要高精度专业知识时,可能因数据复杂性和技术局限性出现偏差,存在语义理解不准确和生成内容质量参差不齐等问题,需要进一步的技术研发和优化[13][14]。

3. 高中生物学混合模式组卷路径探索

传统人工组卷、高中生物题库组卷和 AI 智能组卷三种模式各具优势,AI 系统的快速处理能力大大减少了教师的重复性劳动,凭借强大数据处理能力,保证试卷难度、知识点覆盖分析的客观性,题库系统则提供标准化的题目资源和质量保障,而教师的专业审核确保了试卷的教育质量,使试卷质量稳定性明显提高,教师可以在 AI 提供的框架基础上进行创新性调整,避免无效的重复劳动。功能互补进一步强化了各组卷方式的协同效应,即利用生成式人工智能组卷的高效性,参照题库真题优势,最后教师动态调整,体现出混合组卷的灵活性。另外,教师面对不同教学阶段、学生群体和考试类型时,三种组卷方式的权重配比需要相应调整,日常练习可以更多依赖 AI 系统和题库来提高组卷效率,重要考试则应增加人工参与比重以确保试卷质量,而在创新教学环境中,人工组卷的主导地位能够充分发挥教师的创造性

和教学智慧,实现教学目标的精准达成,因此可整合三种组卷方式,整合路径“见图1”。

本文探索的混合组卷模式充分发挥教师主导作用既可以避免题库与人工组卷的不足,又提高组卷效率,避免 AI 幻觉。通过在组卷不同阶段采用相应的技术手段来实现最优效果。基于 AIGC 高中生物学混合式组卷经历了基于根据高中生物课程标准等确立目标和内容-根据目标内容设计提示词后 AIGC 组卷-教师校验试做三个阶段。第一阶段,根据高中生物课程标准,确定考试性质(如诊断性测试、形成性测试等),进而选择内容;第二阶段,在第一阶段的基础上,设计提示词,并通过 Claude 形成试卷,教师根据试卷质量内容等调整提示指令;第三阶段,教师试做实验,教师通过试做或者同学试做,对试卷难度、时间及知识点准确性等评价。

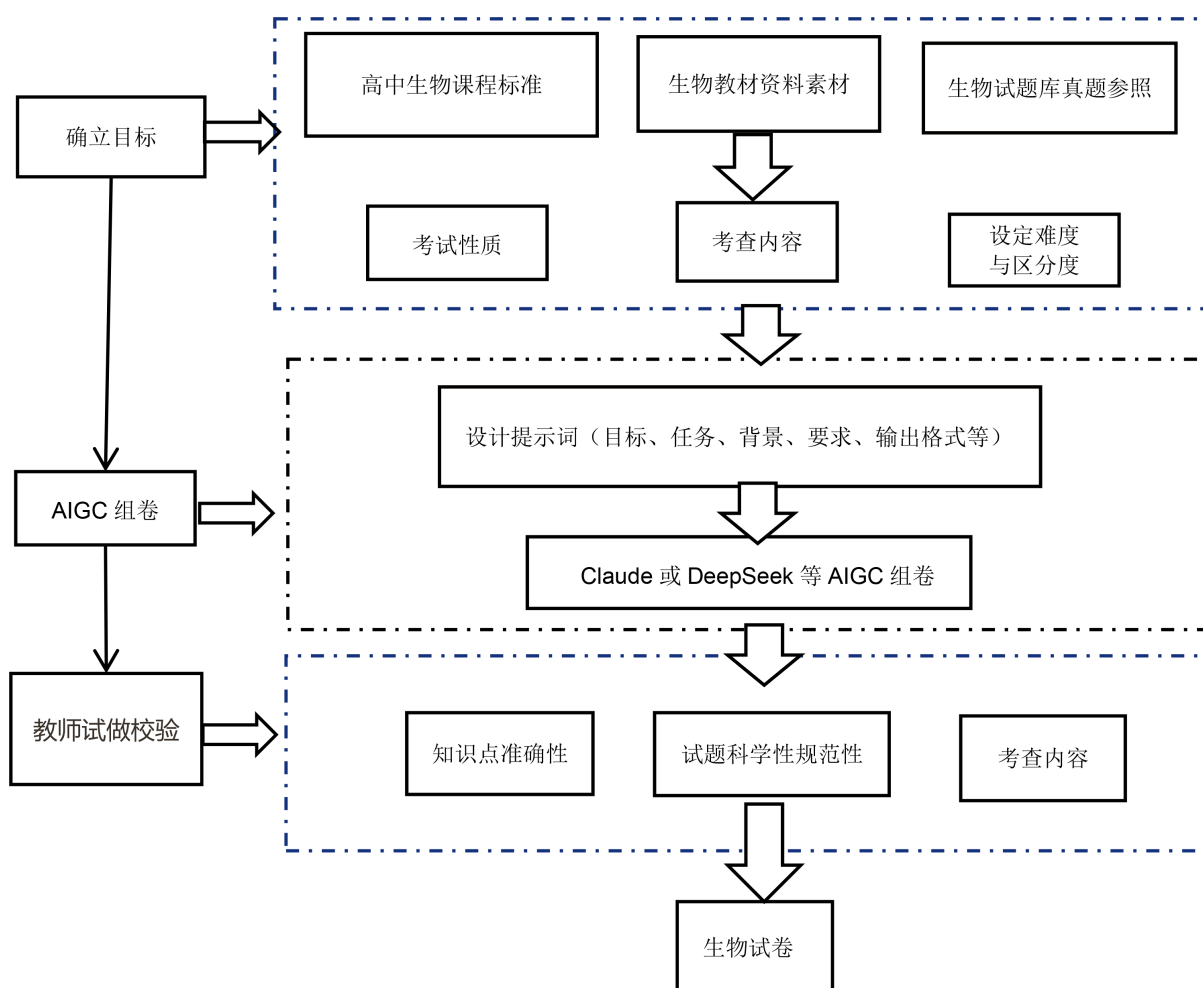


Figure 1. Pathway for high school biology test paper generation based on AIGC

图 1. 基于 AIGC 高中生物学组卷路径

3.1. 前期准备与试卷规划阶段

前期准备和试卷规划阶段,教师主要有三个任务:第一,明确考试目标,根据教学大纲要求、高中生物课程标准、高中生物教材中的资料素材、及参照高考生物真题,确定考试的性质(诊断性考试、形成性考试还是总结性考试),然后确定考试范围、内容、时间、难度等;第二,分析考试内容,梳理教学大纲和课程标准,确定考察的生物核心素养,根据实际学情分析,设计知识点分布、难度要求、题型比例

等基本条件设定组卷框架，以知识点为横轴，认知层次为纵轴制定双向细目表；第三，确定题型结构，根据考试目标选择合适的题型组合，数量等，如单项选择题、不定向选择题、非选择题等，并确定各题型的分值比例。

3.2. AIGC 命题阶段

在进行生成式人工智能命题之前，需要设计提示词，AI 组卷质量好坏，除了 AI 自身技术影响外，提示词也是影响试卷质量的因素之一，AI 存在语义理解不准确和生成内容质量参差不齐等问题，需要不断调整或者修改提示词，一个清晰明确的表达有助于提高人机交互质量。可以通过以下方式提高 AI 对需求的理解。首先，结构化指令，使用清晰的结构组织需求，具体任务，相关背景知识，文档格式、字数等具体要求，输出方式等；其次，角色扮演，让 AI 扮演特别角色，便于其从专业视角生成相对专业的内容，比如“我是一名资深的高中生物教师，具有丰富的教学经验，现在要出一套人教版高中生物学必修 2 遗传与进化第 3 单元测试试卷。”教学经验丰富的高中生物教师就是 AI 的角色定位；最后，迭代改进，在提示词设计完后，进行初次测试，即将提示词导入 AI 对话框，然后根据输出情况调整指令，优化提示词，进一步细化需求。另一种方法是让 AIGC 生成试卷提示词，然后教师在根据具体需求进行二次调整修改其结果。以下是要求 GCAI 辅助出一套人教版高中生物学必修 2 遗传与进化第三章单元试题的提示词。

不论是 AIGC 生成的提示词还是教师设计的提示词，都需要在初次测试后，根据生成内容的不足做调整，使 AIGC 深入思考，对试卷进一步完善。两种方式的不同点在于 AIGC 生成的提示词相对来说，更系统，全面，清晰，从试卷基本信息、知识覆盖要求、重难点及题型结构要求，以及最后对生成的试卷做评价，为教师对试卷的二次修改提供了一定参考。

3.3. 质量检验阶段

教师凭借自身专业知识初步审核试卷内容，根据评价量表检查试题覆盖面，知识点准确性，试题科学性、规范性、难易程度等六个维度(见表 1)，将试卷划分四个等级(优秀：85~100 分；良好：70~84 分；合格：60~69 分；不合格：小于 60 分)。

Table 1. Quality evaluation scale for AIGC-assisted generation of high school biology test papers
表 1. AIGC 辅助生成高中生物试卷质量评价量表

维度	子项目	评分标准
知识点覆盖度 (权重 20%)	内容分布均衡性(5)	5 分：完全覆盖课程标准要求的核心概念，各模块(分子与细胞、遗传与进化、稳态与调节、生物与环境)分布比例合理(误差 ≤ 5%)
		4 分：覆盖 80%以上核心概念，模块分布基本合理(误差 ≤ 10%)
		3 分：覆盖 60%~80%核心概念，存在明显的模块失衡(某模块占比 > 45%或 < 10%)
		2 分：覆盖 40%~60%核心概念，模块分布严重失衡
		1 分：覆盖 < 40%核心概念，内容分布极不均衡
知识点关联度(5 分)		5 分：试题间存在合理的知识网络关联，能考查知识迁移能力，无简单重复
		4 分：大部分试题有关联性，少量重复(<10%)
		3 分：知识点较为孤立，存在一定重复(10%~20%)
		2 分：知识点孤立明显，重复率 20%~30%
		1 分：知识点杂乱无章，重复率 > 30%

续表

时效性与前沿性 (5 分)		5 分: 有机融入近 3 年生物学前沿进展(如 CRISPR、mRNA 疫苗、合成生物学等), 比例适中(10%~20%)
		4 分: 包含部分前沿内容(5%~10%)
		3 分: 偶有涉及前沿内容(<5%)
		2 分: 内容相对陈旧, 缺乏时代感
		1 分: 完全脱离当代生物学发展
认知层次梯度(7 分)		7 分: 认知层次分布完全符合理想比例(误差 $\leq 3\%$), 梯度合理
		6 分: 基本符合理想比例(误差 $\leq 5\%$)
能力层级分布 (权重 20%)		5 分: 较为符合(误差 $\leq 8\%$), 但某层次略有不足
		3~4 分: 分布失衡, 低阶或高阶思维明显过多/过少
		1~2 分: 严重失衡, 如全是识记题或过度追求高阶思维
生物学科能力考查 (7 分)		7 分: 认知层次分布完全符合理想比例(误差 $\leq 3\%$), 梯度合理
		6 分: 基本符合理想比例(误差 $\leq 5\%$)
		5 分: 较为符合(误差 $\leq 8\%$), 但某层次略有不足
		3~4 分: 分布失衡, 低阶或高阶思维明显过多/过少
		1~2 分: 严重失衡, 如全是识记题或过度追求高阶思维
情境真实性与新颖性(7 分)		7 分: 情境来源于真实科研、生产、生活, 贴近学生经验, 80%以上为原创情境
		6 分: 大部分情境真实新颖(60%~80%原创)
		5 分: 部分情境新颖(40%~60%原创), 其余为常见情境改编
		3~4 分: 多为常规情境(<40%原创), 缺乏新意
		1~2 分: 情境陈旧重复, 或脱离实际
题干情境创新性(权重 15%)	情境复杂度(4 分)	4 分: 情境层次丰富, 包含简单~中等~复杂情境, 比例合理(3:5:2)
		3 分: 情境复杂度有层次, 但分布不够均衡
		2 分: 情境复杂度单一, 全简单或全复杂
		1 分: 情境表述不清, 影响理解
图表数据质量(4 分)		4 分: 图表(曲线、模式图等)设计科学, 数据合理, 具有分析价值, 无常见错误
		3 分: 图表基本合理, 数据可信
		2 分: 图表设计简单, 数据价值有限
		1 分: 图表存在科学性错误或数据不合理
选项迷惑性 (权重 15%)	干扰项设计合理性 (8 分)	8 分: 所有干扰项均基于典型错误概念、易混知识点或不完整思维设计, 具有强迷惑性, 无明显破绽
		7 分: 90%以上干扰项设计合理
		6 分: 80%~90%干扰项合理
		4~5 分: 60%~80%干扰项合理, 部分过于简单或明显错误
		2~3 分: <60%干扰项合理, 多为凑数选项

续表

		1分: 干扰项设计差, 存在多个正确答案或全部错误
		4分: 各选项长度相近(差异 $\leq 20\%$), 结构平行, 无明显提示性差异
	选项长度与结构均衡性(4分)	3分: 选项长度基本均衡(差异 $\leq 30\%$)
		2分: 存在明显长度差异(30%~50%), 可能暗示答案
		1分: 选项长度严重不均(>50%), 存在明显提示
	答案分布随机性(3分)	3分: A/B/C/D 答案分布均匀(每个选项 18%~32%), 无明显规律
		2分: 分布基本合理, 但某选项偏多/偏少(15%~35%)
		1分: 答案分布严重失衡(如 A 答案>40%), 存在明显规律
	概念表述准确性(10分)	10分: 所有生物学概念、术语、原理表述完全准确, 符合学术规范
		8~9分: 存在 1~2 处表述不够严谨但不影响理解
		6~7分: 存在 3~5 处概念表述不清或欠准确
		3~5分: 存在 6~10 处概念错误, 部分影响理解
科学准确性 (权重 25%)	逻辑严密性(8分)	1~2分: 存在>10 处概念错误, 或有严重科学性错误
		8分: 题干逻辑链完整, 因果关系清晰, 推理严密, 无漏洞
		7分: 逻辑基本严密, 存在 1 处小瑕疵
		6分: 存在 2~3 处逻辑不够严密但可理解
		4~5分: 存在明显逻辑漏洞或循环论证
	数据与事实准确性(7分)	2~3分: 逻辑混乱, 因果颠倒
		1分: 严重逻辑错误, 无法作答
		7分: 所有数据、实验现象、生物学事实完全准确, 引用规范
		6分: 存在 1 处数据/事实小错误
		5分: 存在 2~3 处数据/事实错误
难度与区分度 (权重 10%)	难度分布合理性(5分)	3~4分: 存在 4~6 处错误
		1~2分: 数据事实错误 > 6 处
		5分: 难度分布完全符合标准, 呈正态分布
		4分: 基本符合, 误差 $\leq 5\%$
		3分: 较为符合, 误差 $\leq 10\%$
	预期区分度(5分)	2分: 分布不合理, 过易或过难
		1分: 难度严重失衡
		5分: 预期试卷整体区分度 $D \geq 0.3$, 核心题目 $D \geq 0.4$
		4分: 预期 D 在 0.25~0.3 之间
		3分: 预期 D 在 0.2~0.25 之间
		2分: 预期 $D < 0.2$, 区分度不足
		1分: 题目过易或过难, 无区分功能

以人教版高中生物选择性必修 1 为列，组一份中等难度的综合试卷，对试卷进行质量评估(见表 2)，六个维度最后加权分为 71 分，试卷良好。如果发现与预期相差过大，可选择再次生成，或调整提示词，便于人工智能理解；教师试做验证，在生成预期试卷后，教师可自己或组织相关师试做，估算答题时间，发现潜在问题，对于 AIGC 组卷生成的不合适的题型，教师可以人工替换一些平时做题过程中比较典型的有代表性的好题(比如题库系统中筛选)，使试卷更加完善，确保试卷能够准确反映学生的学习水平，最后组织学生小范围测试，进行效果评估。答案有差异，主要原因可能与当前 AIGC 技术对于非选择题中图示的解读有待提高。

Table 2. Quality evaluation report of AIGC-generated high school biology exam papers
表 2. AIGC 生成高中生物试卷质量评价报告

维度	得分(S)	权重(W)	质性评价与改进建议
知识点覆盖度	14	20%	列：对细胞结构考察全面，但对光合作用考察部分缺失。建议增加暗反应具体过程。
能力层级分布	13	20%	列：记忆类、理解类题目过多，综合能力考察方面欠缺，
题干情境创新性	9	15%	列：以 2023 年诺贝尔生理学奖“RNA 疫苗”研发为背景设计免疫调节的题目，情景新颖，教师可以在生成指令中特意强调。
选项迷惑性	10	15%	列：部分选择题区分度不高。
科学准确性	19	25%	列：内容准确，但只能生成简单模型图，软件自身技术待提高，建议教师针对性处理。
难度与区分度	6	10%	列：中等难度指令下生成的试卷相对较好。

本研究对 Claude 生成的试卷进行试做测试，结果发现，试题难度及难度梯度与预期相符，单项选择题质量比多项选择题较高，AIGC 在设计非选择题时，对于真实情景等知识需要教师导入，引导 AI 思考出题。另外，为进一步了解 AIGC 的阅卷能力，在将“2025 年黑龙江、吉林、辽宁、内蒙古普通高等学校招生选择性考试生物卷”分别导入 DeepSeek 及 Claude 后，并给出“帮我批阅这份试卷，要求给出每道题的解析过程”的指令，第一次 AI 批阅后人工检查得出以下结果：选择中正确率为 80%(15 道单选题全对，5 道多选题错了 3 道)，进行重新批阅操作，对给出答案进行迭代后，发现其解析过程会有变化，但选项基本稳定；非选择题语句表述专业性有待进一步提高。

4. 结论与展望

通过对 AI 组卷、题库组卷和传统人工组卷三种方式的深入比较分析，发现单一组卷方式都有其独特优势和明显局限：AI 组卷效率高、客观性强，可以节省教师大量的时间，但曲线图和一些复杂的图形(如细胞图、系谱图)不能呈现出来，个别题有文字描述不准确；题库组卷标准化程度高、资源共享便利但个性化不足；传统人工组卷个性化和创新性突出但效率低且质量不稳定。基于此发现，本研究提出了混合组卷模式，通过功能互补和动态调整等策略，有效整合了三种方式的优势，提高组卷效率与质量，减轻教师负担，为教育评价提供了更有效的解决方案，另外，该模式在跨学科应用方面有一定范式意义，通过适应性调整，可用于其他学科(比如化学、地理)，辅助相关试卷的生成和批阅。混合式组卷是新技术的应用，应以谨慎和批判性的眼光审视这一模式，了解其自身的风险、局限性以及伦理问题，知识深度、情景创设及图表绘制是其自身局限性的体现，而试卷的版权归属、学术诚信以及公平性等潜在问题不容忽视。

相信随着人工智能技术的不断发展，Claude、DeepSeek 等 AIGC 软件的多模态场景转换应用会得到

发展,自然语言理解能力的提升将使 AI 系统能够更好地理解题目的语义内容和教学意图,多模态技术的应用将使系统具备处理图像、音频等多种形式题目的能力。

基金项目

赤峰学院 2025 年度研究生科研创新项目(Cfxyyjskeycx2025006)。

参考文献

- [1] 柳博. 考试命题制度研究[M]. 北京: 高等教育出版社, 2017.
- [2] 骆方, 田雪涛, 屠焯然, 等. 教育评价新趋向: 智能化测评研究综述[J]. 现代远程教育研究, 2021, 33(5): 42-52.
- [3] Huang, R. and Song, H. (2025) Comment On: “ChatGPT 35 Better Improves Comprehensibility of English, than Spanish, Generated Responses to Osteosarcoma Questions”. *Journal of Surgical Oncology*, **131**, 1491-1492.
<https://doi.org/10.1002/jso.28132>
- [4] 管宝云. 基于混合智能算法的高校时间表及自动组卷问题研究[D]: [博士学位论文]. 天津: 天津大学, 2005.
- [5] 张玉强, 李年琪, 尹学慧. 中小学教师命题能力: 现状剖析、提升路径与整体支持[J]. 华东师范大学学报(教育科学版), 2025, 43(6): 81-98.
- [6] 全惠云, 范国闯, 赵霆雷. 基于遗传算法的试题库智能组卷系统研究[J]. 武汉大学学报(自然科学版), 1999(5): 758-760.
- [7] 王友仁, 张砦, 施玉霞, 等. 题库系统智能成卷理论和组卷方法研究[J]. 电子科技大学学报, 2006(3): 363-366.
- [8] Lei, W. (2023) Automatic Test Paper Generation Technology for Mandarin Based on Hilbert Huang Algorithm. *Procedia Computer Science*, **228**, 808-816.
- [9] 聂慧超. 学术出版交手 AIGC 产生“双重魔法” [N]. 中国出版传媒商报, 2024-04-05(001).
- [10] 邱元阳. AI 的幻觉[J]. 中国信息技术教育, 2025(9): 22.
- [11] 任少铎. 基于项目命题落实立德树人考察核心素养培养关键能力——以 2024 年福建中考“项目式试题”为例[J]. 物理教学, 2025, 47(9): 34-37.
- [12] 蔡悦琪, 肖安庆. 生成式人工智能在高中生物学考试评价中的实践与探索[J]. 生物学教学, 2025, 50(7): 41-43.
- [13] Ngo, A., Gupta, S., Perrine, O., Reddy, R., Ershadi, S. and Remick, D. (2024) ChatGPT 3.5 Fails to Write Appropriate Multiple Choice Practice Exam Questions. *Academic Pathology*, **11**, Article 100099.
<https://doi.org/10.1016/j.acpath.2023.100099>
- [14] 闫白洋, 余建云. 生成式人工智能在高中生物学教学领域中的问答测试与使用建议[J]. 生物学教学, 2023, 48(9): 34-36.