Advances in Education 教育进展, 2026, 16(11), 215-226
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ae
https://doi.org/10.12677/ae.2026.16112377

Hans汉斯

TFLM-BME：生物医学工程专业生物化学课程形成性评价机制构建与实证

刘　超

中南民族大学生物医学工程学院，湖北　武汉

收稿日期：2026年9月30日；录用日期：2026年10月28日；发布日期：2026年11月5日

摘　要

2025年《高等教育学科专业设置调整优化行动方案（2025—2027年）》与《卓越工程师教育认证标准》对交叉学科专业基础课程的评价改革提出了系统性要求，新工科建设进入学科交叉融合与工程能力导向深化的新阶段。针对生物医学工程(BME)专业生物化学课程普遍存在的“评价偏重知识记忆、与工程能力培养脱节、学生被动接受评价”共性问题，本文以“目标－改进－结果”的“以评促学”理论与课堂本位形成性评价(CBFA)框架为基础，融合“评价即学习”理念，构建三阶反馈闭环教学设计模型(Three-Stage Feedback Loop Model for BME, TFLM-BME)，从目标锚定、证据引出与反馈推进、自我调节三个层级重构课程评价体系。基于97份本科生学情数据的诊断表明，传统评价体系在“目标－引出－解释－反馈－行动”全链条均存在功能失调，难以支撑“问题分析”“设计/开发解决方案”等工程能力目标。经过一个完整教学周期的实践检验，TFLM-BME模型显著提升了学生对考核要求的认知清晰度、作业完成质量与元认知反思水平，课堂形成性互动的质量与完整度明显改善；重复测量方差分析验证了作业成绩提升的统计显著性。研究表明，将评价从学科知识的矫正工具重构为跨学科融合的认知支架，通过“评价实施－结果反馈－教学改进 + 学习改进－效用生成”的完整链条，可有效推动“评价即学习”在专业教育场景的落地。同时，模型落地存在教师负荷、学生适应性等现实挑战，需配套策略缓释。研究为新工科背景下同类交叉学科基础课程的评价改革提供实践范式。

关键词

TFLM-BME，形成性评价，以评促学，评价即学习，生物医学工程

TFLM-BME: Construction and Empirical Study of Formative Assessment Mechanism for Biochemistry Course in Biomedical Engineering

Chao Liu

文章引用: 刘超. TFLM-BME: 生物医学工程专业生物化学课程形成性评价机制构建与实证[J]. 教育进展, 2026, 16(11): 215-226. DOI: 10.12677/ae.2026.16112377

刘超

College of Biomedical Engineering, South-Central Minzu University, Wuhan Hubei

Received: September 30, 2026; accepted: October 28, 2026; published: November 5, 2026

Abstract

As the construction of new engineering disciplines enters a new phase characterized by interdisciplinary integration and engineering competency orientation, the "Action Plan for Adjusting and Optimizing Higher Education Discipline and Program Settings (2025~2027)" and the "Accreditation Standards for Excellent Engineer Education" issued in 2025 have put forward systematic requirements for assessment reform in foundational courses of interdisciplinary programs. Addressing the common problems in the Biochemistry course for Biomedical Engineering (BME) majors—namely, "assessment emphasizing knowledge memorization, disconnection from engineering competency development, and students being passive recipients of assessment"—this study constructs a Three-Stage Feedback Loop Model for BME (TFLM-BME) based on the "objectives-improvement-outcomes" framework of "Assessment for Learning" and the Classroom-Based Formative Assessment (CBFA) framework, integrating the concept of "Assessment as Learning". The model restructures the course assessment system through three stages: goal anchoring, evidence elicitation and feedback advancement, and self-regulation. Diagnosis based on 97 undergraduate survey responses reveals that the traditional assessment system suffers from functional dysfunctions across the entire chain of "objectives-elicitation-interpretation-feedback-action", failing to support engineering competency objectives such as "problem analysis" and "design/development solutions". After one complete teaching cycle, the TFLM-BME model significantly improved students' clarity of assessment requirements, quality of assignment completion, and level of metacognitive reflection, while the quality and completeness of classroom formative interactions were notably enhanced. Repeated measures ANOVA verified the statistical significance of assignment score improvement. The study demonstrates that reconstructing assessment from a corrective tool for disciplinary knowledge into a cognitive scaffold for interdisciplinary integration—through the complete chain of "assessment implementation - result feedback - teaching improvement + learning improvement - utility generation"—can effectively promote the implementation of "Assessment as Learning" in professional education contexts. Meanwhile, the model faces practical challenges such as teacher workload and student adaptability, which require supporting strategies. The research provides a practical paradigm for assessment reform in similar interdisciplinary foundational courses under the new engineering education framework.

Keywords

TFLM-BME, Formative Assessment, Assessment for Learning, Assessment as Learning, Biomedical Engineering

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

2025 年，中央教育工作领导小组印发《高等教育学科专业设置调整优化行动方案（2025—2027 年）》，明确提出“实施新兴学科和交叉学科孵化行动，布局建设一批示范性学科交叉中心”，并强调“加快教学内容迭代，强化人工智能赋能教育教学”[1]。同年 12 月，教育部发布《卓越工程师教育认证标准》，强调“以能力提升为目标”“以学位评价改革牵引教育评价重建”[2]。2026 年，教育部发布《普通高等

学校本科专业目录（2026 年）》，首次在“交叉学科”门类中集中布局，明确“更好适应新兴交叉学科发展和复合型人才培养需求”[3]。上述系列政策指向推动高等工程教育从“知识传授导向”向“能力培养导向”深化转型，同时以“学科交叉融合”为关键路径，加快培养具备跨学科视野与转化应用实践能力的高层次复合型人才。生物医学工程(Biomedical Engineering, BME)作为典型的医工交叉学科，兼具理学、工学、医学三重属性，其人才培养的核心在于“运用工程技术方法解决生物与医学领域问题”。生物化学课程是 BME 专业知识体系的重要基石，承担着为后续生物医学传感、医学仪器、分子诊断等工程类课程提供重要理论支撑的作用[4] [5]。

长期以来，BME 专业的生物类基础课程基本沿用传统生物科学专业的教学与评价模式，以终结性考试为核心评价方式，侧重考查知识点的记忆与复述。该评价模式与 BME 专业“工程能力导向”的培养目标存在系统性错位。学生熟练记忆代谢通路与分子机制后，仍然缺乏将其转化为分析工程问题、设计技术方案的能力。与此同时，学生普遍将基础课视为“突击应付的学分课”，学习动机以外部驱动为主，缺乏主动将生物学知识与工程场景建立关联的意识。

学习评价理论的发展为破解这一困境提供了思路。从泰勒的“目标模式”、斯塔弗尔比姆的“改进模式”到斯凯瑞文的“目标游离模式”，早期研究构建了“目标－改进－结果”的以评促学框架[6]；从“对学习的评价”(Assessment of Learning, AoL)到“为学习的评价”(Assessment for Learning, AfL)，再到“评价即学习”(Assessment as Learning, AaL)，评价的功能逐步从甄别排序转向促进学习、赋能学生。但现有研究多面向普通教育与通用学科，针对课堂微观层面的课堂本位形成性评价研究，在工程教育领域尤其匮乏，缺少结合 BME 专业工程属性、可落地、可推广的形成性评价设计框架与实证证据。

针对 BME 专业的特殊性，本研究以中南民族大学生物医学工程专业《生物化学》课程为实践载体，引入以评促学的系统理论与 Classroom-Based Formative Assessment (CBFA)的完整循环框架，针对 BME 专业的特殊性构建三阶反馈闭环教学设计模型(Three-Stage Feedback Loop Model for BME, TFLM-BME)，系统设计形成性评价反馈机制，并通过教学实证检验其效果。

2. 文献综述与理论基础

2.1. 以评促学的理论演进：从目标模式到评价即学习

“评价如何促进学习”是教育评价领域的经典命题。以评促学的理论发展经历了从“评价有什么用”到“评价如何起作用”的深化过程，形成了“目标－改进－结果”的经典分析框架与“AoL-AfL-AaL”的现代学本评价体系[6]。

2.1.1. 早期探索：“目标－改进－结果”的三维评价框架

20 世纪早中期的研究围绕“评价的功能”展开，形成了三种经典模式。一是泰勒的目标模式，提出评价本质是“确定课程与教学计划实际达到教育目标的程度”，建立了“确定教育目标－选择学习经验－组织学习经验－评价教学结果”的闭环，首次将评价从事实测量拓展为价值评价[7]。二是斯塔弗尔比姆的改进模式，提出“评价不是为了证明，而是为了改进”的核心命题，构建了背景、输入、过程、结果四环节评价模型，将评价定位为决策服务的工具，强化了评价的过程性与改进功能[8]。三是斯凯瑞文的目标游离模式，批判泰勒模式仅关注预期效应，提出评价重点应从“课程计划的预期结果”转向“课程计划实际的结果”，关注评价带给学生的真实发展[9]。三种模式分别对应目标效用、改进效用、发展效用三个维度，共同构成了“目标－过程－结果”的以评促学分析框架。

2.1.2. 现代转向：AoL-AfL-AaL 的学本评价体系

20 世纪 90 年代以来，布莱克和威廉姆等学者将研究深入课堂内部，推动形成性评价成为主流，逐步

形成了“关于学习的评价–促进学习的评价–作为学习的评价”的三级体系。关于学习的评价(AoL)以总结性评价为核心，侧重对学习结果的鉴定与分级[10]；促进学习的评价(AfL)核心是教师通过反馈帮助学生明确学习差距、改进学习过程，评价从“裁判”转向“支架”[11]；作为学习的评价(AaL)进一步将评价的主体转向学生，提出评价本身就是学习过程，学生通过自我评价、同伴评价与自我反思，主动监控学习进程、调整学习策略，核心是元认知能力的发展[12]。三者并非相互替代，而是功能协同：AoL 诊断现状，AfL 指明方向，AaL 赋能自主学习[6]。

2.2. 课堂本位形成性评价的操作化框架

课堂本位形成性评价(CBFA)是形成性评价在课堂场景的落地形态。Gu 和 Yu 系统提出了 CBFA 的操作化定义与核心框架，指出完整的 CBFA 事件包含四个不可或缺的要素：学习证据引出、学习结果解释、针对性反馈、后续改进行动，四个环节以学习目标为参照，形成循环递进的螺旋结构[13]。该研究同时指出，真实课堂中普遍存在“形成性循环断裂”的问题：教师通常完成了“提问–判断–反馈”的前半段，但缺少后续的改进行动，多数反馈停留在确认与表扬层面，未能真正指向学习差距的弥补。现有 CBFA 研究多面向语言教育等单一学科，针对工科类交叉专业的适配性研究明显不足。

2.3. 反馈机制的核心转向：从“教师给予”到“学生使用”

反馈是形成性评价发挥效能的核心杠杆。Hattie 和 Timperley 提出的经典三功能框架——feed-up (目标导向)、feed-back (现状诊断)、feed-forward (改进路径)——奠定了反馈研究的基础[14]。但近年研究的焦点正在发生“接收端转向”：反馈的有效性不取决于教师输出了多少，而取决于学生是否“听到、理解并能够据此行动”[6]。Carless 和 Young 的纵向研究表明，成功的反馈接收体验是提升自我效能感、发展反馈素养的关键，而反馈素养的发展呈现“理解价值–增强韧性–管理接收”的阶段性特征[15]。这一转向对教学设计具有重要启示：形成性评价的设计不能只关注“教师如何给反馈”，更要关注“学生如何用反馈”。

2.4. 新工科交叉专业评价的特殊要求与研究缺口

将形成性评价应用于 BME 专业生物类基础课程，必须回应专业的三个核心特殊性。第一，工科属性：评价标准必须从“学科知识掌握”转向“工程能力发展”。第二，能力导向：评价必须支撑毕业要求中的“问题分析”“设计/开发解决方案”等高阶能力指标。《卓越工程师教育认证标准》明确要求“以能力提升为目标”，聚焦“核心课程、校企课题、师资队伍和工程师技术中心等培养要素”[2]。第三，跨学科整合：评价应促进生物学知识与工程思维的图式融合。2026 年本科专业目录首次在“交叉学科”门类中集中布局，明确“更好适应新兴交叉学科发展和复合型人才培养需求”[3]。BME 的核心能力是跨学科整合，评价任务需要引导学生在生物学原理与工程应用场景之间建立联结，而非孤立考查学科知识。现有以评促学研究多面向单一学科，针对上述交叉工程教育特殊性的适配设计明显不足。本研究正是针对这一缺口，构建适配 BME 专业的形成性评价模型(TFLM-BME)，开发 5 级高阶能力评分量规，并开展实证检验。

3. BME 专业生物类基础课程评价现状诊断

3.1. 研究设计与数据来源

采用自编《生物化学课程学习调查问卷》，面向某高校生物医学工程学院选修该课程的 2024 级本科生发放，共发放问卷 125 份，回收有效问卷 97 份，有效回收率 77.6%。问卷覆盖学习动机、学习投入、

知识掌握自评、学习难点归因、考核评价反馈等维度，同时结合超星泛雅教学平台的学习行为数据、课堂观察记录进行交叉验证。研究以 CBFA 的“目标–引出–解释–反馈–行动”五要素为分析框架，系统诊断传统评价体系的功能失调。

3.2. 传统评价体系的三重功能失调

3.2.1. 目标锚定偏差：评价标准偏离工程能力导向

学生选课动机与学习投入呈现明显的外部驱动与短期突击特征，具体结果如表 1 所示。仅 29.9%的学生因对学科内容感兴趣选修课程，超六成学生以获取学分、规避难度为主要选课动机；学习投入方面，51.5%的学生在测试前集中复习少于 6 小时，12 小时以上系统复习的学生仅占 17.5%。

Table 1. Survey on learning motivation and learning engagement of BME students (N = 97)
表 1. BME 专业学生学习动机与学习投入调查(N = 97)

维度	具体选项	人数	占比
选课动机	兴趣主导	29	29.9%
	学分主导	40	41.2%
	课程难度主导	20	20.6%
	其他	8	8.2%
复习时长	2~3 小时	10	10.3%
	4~6 小时	40	41.2%
	8~12 小时	30	30.9%
	12 小时以上	17	17.5%

从以评促学的框架来看，这一现象本质是目标锚定的系统性偏差：当评价的成功标准以知识记忆为核心、未锚定工程能力时，学生无法感知这门课对自身专业能力发展的价值，自然将其定位为“突击应付即可的学分课”。《行动方案（2025—2027 年）》明确要求“加快教学内容迭代”[1]，其深层逻辑正是推动评价标准与能力目标的系统性对齐。同时，63.9%的学生将“学习课程太多，没时间复习”列为首要学习难点，也反映出在多课程并行的情境下，学生倾向于将时间资源优先分配给“与工程能力直接相关”的课程。

3.2.2. 解释标准错位：评价内容与毕业要求不匹配

在知识掌握自评中，超过 70%的学生表示“知识记不住”或“掌握一般”，“听懂了但记不住”是高频反馈。学习难点的多选归因统计如表 2 所示，“考察的知识点太多”占比达 59.8%，“花了大量时间复习，但效果不佳”占比 26.8%；更值得关注的是，52.6%的学生认为“考试通过完全靠记忆”。

Table 2. Attribution statistics of students' learning difficulties in the course (Multiple choice, N = 97)
表 2. 学生课程学习难点归因统计(多选，N = 97)

归因选项	选择人次	占比
学习课程太多，没时间复习	62	63.9%
考察的知识点太多	58	59.8%
花了大量时间复习，但效果不佳	26	26.8%
内容理解困难	22	22.7%

这一数据揭示了 CBFA“解释环节”的标准偏差：传统评价对学习结果的解释，仅围绕“知识点记忆是否准确”展开，未能有效覆盖“问题分析”“设计解决方案”等工程能力指标。《卓越工程师教育认证标准》强调“以学位评价改革牵引教育评价重建”[2]，正是针对这一问题的政策回应。当评价的解释标尺指向“记忆”时，学生的学习策略自然向机械记忆倾斜。

3.2.3. 反馈主体单一：自我调节训练的缺失

调查同时呈现一组反差：84.6%的学生认可课程知识的重要性，74.2%肯定课堂讲授的效果，但 52.6%认为考试通过完全靠记忆。这说明教学在“知识理解”层面有效，但评价未能引导学生走向更高阶的能力发展。从 CBFA 的完整循环来看，传统评价存在两处断裂：一是反馈主体单一，学生是被动的“被评价者”，缺乏自我评价、同伴评价的参与；二是后续行动缺失，评价结束后缺少学生的反思调整与策略优化，评价没有转化为学习改进的行动。从“AoL-AfL-AaL”的体系视角看，传统评价仍停留在 AoL 为主的阶段，AfL 的反馈改进功能不足，AaL 的自我调节功能基本缺失。而 BME 专业毕业要求明确包含“终身学习”指标，评价循环的断裂使得学生失去了在评价中锻炼元认知能力的机会。

4. TFLM-BME 形成性评价反馈机制的构建

针对诊断出的三重功能失调，本研究基于以评促学的“目标－改进－结果”框架，结合 CBFA 的核心要素与 BME 专业的工程属性进行适配性改造，构建三阶反馈闭环教学设计模型(TFLM-BME)，分别通过“目标锚定、证据引出与反馈推进、自我调节”三阶结构逐一回应，形成“目标－诊断－调节－再目标”的完整螺旋闭环。模型整体框架如表 3 所示。

Table 3. Overall framework of the three-stage feedback loop model (TFLM-BME)
表 3. TFLM-BME 三阶反馈闭环模型整体框架

模型阶段	对应以评促学环节	对应 CBFA 核心要素	对应解决的失调问题	核心实施策略	BME 专业性体现
第一阶：目标锚定	目标效用	学习目标与成功标准	反馈迟滞、目标缺失	三维度评分标准、两类范例对比、目标内化卡	评价标准从“知识复述”转向“工程应用分析”
第二阶：证据引出 + 反馈推进	改进效用	证据引出、结果解释、针对性反馈	反馈信号失真、内容错位	三类医工融合任务、“3 + 1”反馈模式、两阶段测验	反馈聚焦“跨学科整合薄弱点”
第三阶：自我调节	发展效用	后续改进行动、元认知调控	反馈主体单一、循环断裂	三层反思框架、反馈采纳标注、阶段性目标重设	反思指向“医工融合学习策略”

模型设计遵循“教－学－评”一体化原则，将课程目标、教学过程、评价活动系统性统一，确保评价嵌入教学全过程，而非独立于教学之外。《行动方案（2025—2027 年）》提出“实施学科专业内涵更新行动，加快教学内容迭代，强化人工智能赋能教育教学”。TFLM-BME 模型的评价任务设计正是这一要求在课程微观层面的落实。

4.1. 第一阶：目标锚定——建立工程能力导向的评价标准

设计原则：将评价的“成功标准”从“掌握生物化学知识”重构为“能够运用生物化学知识分析 BME 场景中的问题”。与通用形成性评价的区别在于：不是泛化地“明确学习目标”，而是完成评价标准从

“学科知识维度”向“工程能力维度”的转换。

为解决高阶能力评价模糊、一致性不足的问题，本研究开发 BME 专业高阶能力 5 级评分量规，将“医工融合”“综合应用”两个核心维度拆解为 1~5 分的可观测行为表现，同时配套“知识理解”维度分级标准，整体构成三维度评价体系。量规既作为教师评分依据，也作为学生自评、互评与自我改进的参照，具体标准如表 4 所示。

Table 4. Five-level scoring rubric for high-order competencies of BME majors (dimensions of medical-engineering integration and comprehensive application)
表 4. BME 专业高阶能力 5 级评分量规(医工融合与综合应用维度)

分值	医工融合维度(可观测行为表现)	综合应用维度(可观测行为表现)
5 分(优秀)	能够精准建立生物化学原理与 BME 工程/临床场景的内在逻辑关联；主动辨析原理在工程场景中的约束条件与边界；可以多角度解释生物学原理如何影响器件、检测方案性能。	能够完整开展 BME 问题分析，清晰拆解问题要素；提出逻辑严密的初步技术方案；能够识别方案潜在局限，给出优化思路，论证充分。
4 分(良好)	可以建立生物化学原理与 BME 工程场景的明确联系；能够解释核心原理对应的工程意义，少量细节存在疏漏；基本识别原理对应的应用约束。	可以完成 BME 问题分析，给出可行的初步技术方案；方案逻辑通顺，但对局限性分析不够充分，论证较为完整。
3 分(合格)	能够建立生物学知识与工程场景的浅层对应关系；仅能简单罗列“知识点 + 工程例子”，缺少内在逻辑推导；未充分考虑应用约束条件。	可以完成基础问题拆解，给出基础方案框架；方案存在部分漏洞，缺少局限性分析，论证基本成立。
2 分(待改进)	医工关联非常薄弱，仅复述生物学知识点；工程场景举例牵强，知识点和工程应用之间缺少逻辑联结。	仅能碎片化回答部分问题，方案框架残缺；关键环节缺失，难以支撑问题解决。
1 分(较差)	完全无法建立生物学知识与 BME 工程场景的联系，只复述课本理论，没有任何工程视角。	不能识别 BME 问题核心，无法形成有效分析与方案构思。

量规使用说明：① 适用场景：课后作业、PBL 小组汇报、课程小项目的评价；② 评分方式：教师评分、同伴互评均对照量规独立打分，两个维度分别计分；③ 学生使用：作业布置时同步公开量规，学生可对照量规开展自评、预判作业等级，明确改进方向；④ 反馈对接：教师反馈直接对应量规等级，指出学生当前所处等级、具体行为短板以及进阶路径。

除评分量规外，目标锚定阶段还配套两项实施策略：第一、两类范例对比。每章开篇同时提供“学科型范例”和“工程型范例”。以酶的专一性为例，学科型范例聚焦“酶催化底物的特异性”，工程型范例则延伸至“葡萄糖氧化酶的专一性如何保证血糖传感器的检测准确性”，让学生直观看到“好的工程导向分析”是什么样。第二、目标内化卡。第一节课引导学生完成目标内化卡，思考“本课程中’好的分析’与’好的记忆’有何区别”“本课程能力对我未来的专业方向有什么用”，推动外部目标向内部动机转化。

4.2. 第二阶：证据引出与反馈推进——聚焦跨学科整合的多主体反馈

设计原则：评价任务本身具备“跨学科整合”的诊断功能，能够区分“知识记忆型理解”与“工程应用型理解”；反馈聚焦跨学科整合的薄弱点，而非仅纠正知识错误。与通用形成性评价的区别在于：反馈的核心不是“纠错”，而是“定位医工融合的断点”。

具体实施策略包括：第一，三类医工融合评价任务。设计原理分析型、方案设计型、文献分析型三

类任务，如“分析酶失活对血糖试纸检测结果的影响”、“基于乳酸氧化酶设计乳酸生物传感器的初步方案”等，强制学生同时调用生物学知识与工程思维。第二，“3 + 1”多主体反馈模式。构建“教师反馈 + 助教反馈 + 同伴反馈 + 范例参照”的反馈体系：教师反馈重点标注“跨学科整合薄弱点”，助教反馈聚焦基础性错误，同伴互评增设“工程关联度”指标。第三，两阶段测验与信心标注。线上形成性测验以知识理解为主，允许多次尝试，反馈附带“BME 应用提示”；纸质总结性测验以医工融合分析为主。同时要求学生标注“确定/不确定/猜测”以及“该知识点是否与工程应用相关”，帮助教师识别“掌握知识但不会应用”的学生群体。

4.3. 第三阶：自我调节——培养跨学科学习的元认知能力

设计原则：将反思从“知识掌握程度的复盘”升级为“跨学科学习策略的审视”，让学生在评价中学会学习，支撑终身学习能力培养。与通用形成性评价的区别在于：将行动主体从教师拓展到学生，反思不局限于“我学会了什么”，更指向“我如何更好地实现医工融合”。

具体实施策略包括：第一，三层反思框架。设计知识层、整合层、策略层三级反思问题：知识层聚焦“是否掌握核心概念”，整合层追问“能否识别概念在 BME 问题中的切入点”，策略层引导总结“分析医工融合问题的困难模式”。第二，反馈采纳标注。要求学生在作业修改稿中标注“采纳了哪条反馈”“如何调整了分析”“调整是否有效”，将隐性的反馈使用过程显性化。第三，阶段性目标重设。课程中段设置目标重设环节，学生基于前半学期的反思记录，识别自身医工融合的薄弱环节，制定后半学期的个人改进目标，教师提供针对性指导，形成“目标 - 诊断 - 行动 - 再目标”的完整螺旋闭环。

TFLM-BME 模型的核心价值不在于针对生物化学课程的具体设计，而在于其“能力锚定 - 整合反馈 - 元认知调节”的三阶逻辑。这一逻辑可迁移至其他交叉学科基础课程，只需根据不同专业的培养目标，调整评价任务的专业场景与能力指向，即可快速形成适配的形成性评价方案。

5. 实践效果与分析

5.1. 实施设计与数据来源

在《生物化学》课程一个完整教学周期(48 学时)中系统实施上述机制，实践对象为生物医学工程专业 2025 级本科生，共 128 人选修课程。数据来源包括：课程平台学习行为数据、四次课后作业评分、学生反思文本、课程结束后的效果调查问卷，回收有效效果问卷 117 份，回收率 91.4%。

5.2. 量化效果分析

5.2.1. 考核要求清晰度显著提升

课程结束后的效果调查显示，学生对考核体系、评分标准、改进方向的认知度均处于较高水平，具体结果如表 5 所示。100%的学生表示“非常清楚”或“比较清楚”课程的考核要求，68.4%的学生“非常清楚”作业评分标准，65.8%的学生“非常清楚”反馈的改进方向。这表明明确的能力导向标准与多主体反馈，有效消除了学生对“教师期望什么”的模糊认知，解决了传统评价中“信号失真”的问题。

Table 5. Survey on perceived effectiveness of the formative assessment mechanism (N = 117)
表 5. 形成性评价机制效果认知调查(N = 117)

调查项	非常清楚	比较清楚	不太清楚	完全不清楚
课程整体考核要求	53.0%	47.0%	0%	0%
作业评分标准	68.4%	28.2%	3.4%	0%
反馈指出的改进方向	65.8%	30.8%	3.4%	0%

5.2.2. 作业质量持续提升，统计显著性验证

对四次课后作业的教师评分进行纵向比较，结果如表 6 所示。作业平均分从第一次的 13.4 分(满分 20 分)提升至第四次的 16.7 分，累计提升 3.4 分；评分标准差从 2.7 降至 1.9，说明学生间的作业质量差距逐步缩小，形成性反馈对中后段学生的提升作用更为明显，体现了评价促公平的价值。

Table 6. Statistics of quality changes of four after-class assignments
表 6. 四次课后作业质量变化统计

作业批次	平均分(满分 20 分)	标准差	虚假掌握率
第一次作业	13.4	2.7	21.4%
第二次作业	14.2	2.3	15.3%
第三次作业	15.6	2.1	11.2%
第四次作业	16.7	1.9	7.4%

为验证作业成绩提升的统计显著性，对四次作业得分进行单因素重复测量方差分析。结果显示：作业得分的时间主效应显著，$F(3, 348) = 52.37$，$p < 0.001$，偏 $\eta^2 = 0.31$，表明随着形成性评价机制的持续推进，学生作业质量存在统计学意义上的显著提升，效应量中等偏上。事后两两比较显示，相邻两次作业得分差异均达到显著水平(p 值均 < 0.05)，说明反馈的改进作用稳定、持续显现。

5.2.3. 知识虚假掌握率下降，学习诊断更精准

信心标注机制的数据显示，“标注确定但实际错误”的虚假掌握率从第一次作业的 21.4%持续下降至第四次作业的 7.4%，累计下降 14.0 个百分点(表 6)。这说明该机制帮助学生更准确地认知自身学习状态，也帮助教师实现了更精准的学习诊断，反馈的针对性显著增强。

5.2.4. 课堂形成性互动质量改善

课堂观察数据显示，改革后课堂中工程应用类提问占比从 12%提升至 37%，开放式、分析性问题的频次显著增加；小组研讨环节中，围绕工程场景的讨论占比达 63%，改变了传统课堂“教师问知识点、学生背答案”的单一互动模式。

5.2.5. 作用路径的初步相关性证据

为初步检验 TFLM-BME 模型的内部作用逻辑，对“考核要求清晰度”“元认知反思深度”与“第四次作业成绩”进行皮尔逊相关性分析。结果显示：学生对考核要求的清晰度与作业成绩呈显著正相关($r = 0.42, p < 0.001$)；学生元认知反思的频次与作业成绩呈显著正相关($r = 0.38, p < 0.001$)。

这一结果初步支持了“目标清晰→反馈改进→元认知发展→学业提升”的链式作用路径假设。受本研究单组行动研究的设计局限，尚未能建立严格的因果推断；后续扩大样本量、设置平行对照组后，可进一步通过路径分析或结构方程模型，系统验证目标锚定、反馈质量、反思深度与学业表现之间的中介与链式效应，从数据层面完整验证模型的内部逻辑。

5.3. 质性分析：反思深度的层级演进

对四次作业中 356 条学生反思文本进行主题编码分析，结果如表 7 所示。表层反思占 31.5%，策略性反思占 50.0%，元认知反思占 18.5%。策略性反思占比最高，表明多数学生能够将反馈转化为具体的学习行动。

Table 7. Distribution of depth levels of students' reflection texts
表 7. 学生反思文本深度层次分布

反思层次	特征描述	频次	占比
表层反思	仅复述反馈内容，无具体改进计划	112	31.5%
策略性反思	提出具体的学习策略调整与改进计划	178	50.0%
元认知反思	对自身学习方式、思维模式进行批判性审视	66	18.5%

更具意义的是变化趋势：第一次作业中元认知反思仅占 8.2%，第四次作业中上升至 28.1%。这一递增趋势表明，持续的结构化反思训练，能够逐步引导学生从“执行学习策略”走向“审视学习策略”，学生的反馈素养与元认知能力在评价过程中得到了发展，“评价即学习”的效应逐步显现。

5.4. 基于以评促学机制的效果解释

从“评价实施–结果反馈–教学改进 + 学习改进–效用生成”的作用机制来看，TFLM-BME 模型之所以有效，核心在于实现了三个机制的协同：第一，目标锚定机制。工程导向的评价标准为学生提供了清晰的能力发展坐标系，学习动机从“应付考试”转向“能力建设”。第二，反馈改进机制。医工融合的评价任务与多主体反馈，精准诊断了学生的学习断点，推动教师优化教学、学生调整策略。第三，元认知发展机制。三层反思框架将终身学习能力的培养嵌入日常评价循环，让评价过程本身成为能力训练的过程。

5.5. TFLM-BME 模型实施的现实挑战与应对策略

本研究在完整教学周期的实践中也发现，TFLM-BME 模型的落地并非零成本，在教师投入、学生适应、评价信度等方面存在现实挑战。结合实践经验，本研究提出对应应对策略，为同类课程实践者提供参考。

5.5.1. 挑战一：多主体反馈与精细化评分显著增加教师工作负荷

问题表现：相较于传统作业仅需判断对错、给出总分，实施 5 级量规评分、跨学科整合点反馈、反思文本批阅后，教师与助教的单份作业批改时间增加约 40%；“3 + 1”多主体反馈也增加了教学团队的协调成本。在高校教师教学科研双重压力的现实下，全学期、全作业量推行会面临人力约束。

应对策略：① 分层实施，抓大放小：选取 2~3 次核心作业(如 PBL 项目、综合性大作业)开展完整的量规评分、深度反馈与反思训练，普通作业采用“要点式反馈 + 范例参照”的简化模式，控制总工作量。② 发挥同伴互评的分流作用：基础等级判断、问题初步甄别的工作交由学生同伴互评完成，教师仅对高阶能力维度进行二次点评与抽样复核，减少重复性劳动。③ 建设可复用反馈资源库：针对医工融合常见问题、典型薄弱点预制标准化反馈片段与改进建议，减少重复撰写；教研组集体共建量规、范例、反馈语料，实现资源共享，降低单人负担。

5.5.2. 挑战二：学生初期不适应深度反思与量规导向评价

问题表现：课程初期，部分学生长期习惯于传统“对错打分”模式，对 5 级量规的描述性评价、深度元认知反思任务存在抵触：约 23%的学生在首次反思中仅摘抄知识点，无法开展策略层面反思；少数学生认为反思、自评属于额外课业负担。学生反馈素养的建立需要适应周期。

应对策略：① 前置培训与范例示范：课程开篇安排 1 课时专门讲解评分量规、反思框架，结合“优秀/待改进”样例作业直观展示不同等级的差异，让学生明确“好的标准是什么”。② 循序渐进降低门槛：前期提供反思填空模板，引导学生逐步学会结构化表达；反思深度从知识层切入，中期逐步过渡到整合

层、策略层，避免一开始设置过高要求。③ 明确价值与成绩挂钩：明确告知学生量规、反思不是额外作业，而是过程性评价的组成部分，直接纳入课程成绩；同时说明该模式对工程思维、反馈素养的长期增益，提升学生认同度。

5.5.3. 挑战三：同伴互评中高阶能力评分存在信度漂移

问题表现：医工融合、综合应用属于高阶主观能力，学生同伴互评时容易出现打分偏松或偏严的个体偏差，导致互评分数信度不足，难以直接作为最终成绩依据。

应对策略：① 明确定位：互评作为诊断工具而非评分工具：同伴互评的核心目的是让学生在评判他人的过程中学习、诊断自身问题，互评结果仅作为学习参考，最终成绩以教师评分为准。② 开展评分校准练习：正式互评前，用 2~3 份典型样例组织全体学生集体打分、讨论，统一对量规等级的理解，缩小评判偏差。③ 教师抽样复核：教师按比例抽核互评结果，对明显偏离量规描述的评分进行纠正与说明，引导学生逐步提升评价能力。

6. 结论与展望

6.1. 研究结论

本研究引入以评促学的系统理论与课堂本位形成性评价框架，针对 BME 专业生物化学课程评价的现实困境，开发高阶能力 5 级评分量规，构建并检验了 TFLM-BME 模型，主要结论如下：

首先，新工科交叉专业基础课程的传统评价体系，普遍存在目标锚定偏差、评价内容错位、形成循环断裂三重功能失调，本质是评价逻辑与专业能力培养逻辑的脱节。开发可观测的高阶能力评分量规，是破解高阶能力评价模糊、评价一致性不足的重要手段。

其次，TFLM-BME 模型通过“目标锚定－整合反馈－自我调节”三阶结构，能够系统性破解上述三重失调，将评价从学科知识的矫正工具，重构为工程思维与生物学知识融合的认知支架。重复测量方差分析验证了作业成绩提升的统计显著性，相关性分析初步支持了模型的内部作用路径。

最后，“评价即学习”在专业教育场景的落地，不是靠增加反馈数量，而是靠构建“目标内化－跨学科反馈－自我调节”的完整认知循环。同时模型落地存在教师工作负荷增加、学生初期适应困难、同伴互评信度不足等现实挑战，需通过分层实施、循序渐进、校准培训等配套策略予以缓释。

6.2. 实践启示

本研究对新工科背景下的基础课程教学改革具有三点普遍启示：一是基础课评价必须锚定专业能力目标，配套可操作的评分量规。《卓越工程师教育认证标准》强调“以能力提升为目标”[2]。基础课评价不能照搬对应学科专业的评价标准，要以本专业的毕业要求为反向牵引，完成评价标准从“知识”到“能力”的转换，从源头发挥评价的正向指挥棒作用。二是形成性评价的核心是反馈质量而非数量。反馈要聚焦专业核心能力的发展断点，而非停留在知识点纠错，才能真正支撑人才培养目标，实现以评促学的改进效用。同时要平衡改革收益与教师人力成本，分层落地改革举措。三是要将自我调节纳入评价体系，兼顾学生适应性。评价不仅要诊断学习，更要培养学生的元认知能力，为终身学习奠定基础；对于深度反思、量规自评等新模式，应当循序渐进，给学生适应周期。

6.3. 研究局限与展望

本研究的局限在于实践周期较短、样本局限于单一课程，缺少平行对照班级，目前仅通过重复测量方差分析验证了成绩变化的显著性，通过相关性分析初步验证作用路径。后续研究将扩大实践样本，在 BME 专业多门生物基础课中推广 TFLM-BME 模型与评分量规，增设对照班级，采集更大样本数据，进

一步开展路径分析或结构方程模型，系统验证“目标清晰度–反馈质量–反思深度–学业表现”的链式作用机制；同时探索学习分析与AI技术在反馈机制中的应用，提升反馈的个性化与精准度，追踪学生跨学科能力发展的长期效应。

基金项目

中南民族大学教学研究项目“新工科学科交叉背景下生物类基础课程教学探索与实践”(JYX24019)。

参考文献

[1] 中央教育工作领导小组印发《高等教育学科专业设置调整优化行动方案（2025—2027年）》[EB/OL]. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202508/t20250828_1410769.html, 2025-08-28.

[2] 教育部关于发布《卓越工程师教育认证标准》教育行业标准的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A22/s7065/202512/t20251209_1422841.html, 2025-11-25.

[3] 教育部关于公布《普通高等学校本科专业目录（2026年）》的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A08/moe_1034/s3882/202604/t20260427_1434931.html, 2026-04-09.

[4] 朱春楠, 郑冬云, 刘超. 新工科背景下生物医学工程专业化学类基础课程教学改革探索[J]. 化工管理, 2021(33): 15-16.

[5] 莫光权, 阳范文. “新医科”背景下生物医学工程非电子方向的化学课程体系构建研究[J]. 科技风, 2023(35): 35-37.

[6] 李鹏. 评价如何促进学习?——从泰勒到厄尔的探索与反思[J]. 外国教育研究, 2020, 47(1): 31-44.

[7] 拉尔夫·泰勒. 课程与教学的基本原理[M]. 施良方, 译. 北京: 人民教育出版社, 1994: 87-88.

[8] Stufflebeam, D.L., Foley, W.J., Gephart, W.J., *et al*. (1971) Educational Evaluation and Decision Making. F.E. Peacock Publishers.

[9] Scriven, M. (1991) Prose and Cons about Goal-Free Evaluation. *Evaluation Practice*, **12**, 55-62. https://doi.org/10.1177/109821409101200108

[10] Harlen, W. (2007) Assessment of Learning. Sage Publications Ltd. https://doi.org/10.4135/9781446214695

[11] Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M. and Stobart, G. (2002) Assessment for Learning: 10 Principles. University of Cambridge School of Education.

[12] Earl, L.M. (2003) Assessment as Learning: Using Classroom Assessment to Maximize Student Learning. Corwin Press.

[13] Gu, P.Y. and Yu, G. (2020) Researching Classroom-Based Assessment for Formative Purposes. *Chinese Journal of Applied Linguistics*, **43**, 150-168. https://doi.org/10.1515/cjal-2020-0010

[14] Hattie, J. and Timperley, H. (2007) The Power of Feedback. *Review of Educational Research*, **77**, 81-112. https://doi.org/10.3102/003465430298487

[15] Carless, D. and Young, S. (2025) Towards an Idealized Model of Feedback Uptake: Longitudinal Development of Feedback Literacy from a Social Constructivist Perspective. *Studies in Higher Education*, **50**, 1641-1653. https://doi.org/10.1080/03075079.2024.2389436