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

魏　非[1]　杨可欣[2]　杨淑婷[2]

（1. 华东师范大学 教师发展学院，上海 200062；2. 华东师范大学 教育信息技术学系，上海 200062）

[摘要] 场景化评估为教师能力评估从评判向赋能转向提供了新的可能路径。本文首先明晰了教师能力场景化评估的内涵、特征与价值取向，探讨了生成式人工智能赋能教育评估的现状与技术机制，提出了生成式人工智能赋能教师能力场景化评估的框架与路径；然后阐释了生成式人工智能辅助评估场景生成的三种方式：非结构化指令驱动生成、结构化框架引导生成和动态指令驱动的自适应生成，并说明了数据采集与预处理、多模态数据融合、行为模式识别与推断、人机协同分析四个关键数据融合分析环节，为动态场景生成和多维数据分析提供支撑；最后通过三个典型案例阐释了如何将理论框架转化为可操作的评估实践。

[关键词] 场景化评估；教师能力评估；生成式人工智能；人机协同评估设计；人机协同评估实施

[中图分类号] G420　　**[文献标识码]** A　　**[文章编号]** 1007-2179(2026)02-0066-10

一、引 言

能力评估是撬动教师专业发展的重要举措。在人工智能深刻变革教与学模式的背景下，教师能力评估范式创新已成为教师发展数字化的重要任务。传统的教师能力评估主要沿着两条技术路径展开：一是基于自陈量表（如问卷、知识测试与级差表等）的评估。它便于规模化实施，却因依赖主观报告、简化真实教学情境、难以捕捉外显行为等而受质疑。二是基于客观证据的专家评估。它的外部效度较好，但高度依赖专家资源，成本高昂、效率低下，难以实现常态化与规模化推广。二条路径不同，但都面临规模化效率与评估精准度难以兼顾的矛盾，都未构建起“诊断—反馈—改进”的发展闭环，削弱了评估赋能教师学习的价值。

评估应与真实教育教学情景相融合，同时应与教师专业发展机会相连（Darling-Hammond et al., 2012）。场景化评估（scenario-based assessment, SBA）作为一种新兴评估范式，通过构建真实或拟真任务情境，通过采集过程性行为数据，实现对复杂能力的多维度测量。近年来，这种评估在教育、工程、医学、自动驾驶等领域展现出显著优势，弥合了抽象测评与现实应用之间的鸿沟，使评估结果更具生态效度和预测价值。生成式人工智能的发展，尤其

[收稿日期]2026-01-27　　**[修回日期]**2026-02-09　　**[DOI编码]**10.13966/j.cnki.kfjyyj.2026.02.007

[基金项目] 2023 年度全国教育科学规划一般课题“面向教师数字能力发展的场景构建与应用研究”(BCA230283)。

[作者简介] 魏非，博士，副研究员，华东师范大学教师发展学院，研究方向：教师数字素养发展、教师发展数字化（fwei@dec.ecnu.edu.cn）；杨可欣，硕士研究生，华东师范大学教育信息技术学系，研究方向：教师能力发展、大语言模型教育应用；杨淑婷，博士研究生，华东师范大学教育信息技术学系，研究方向：数智时代教师专业发展、教师能力评估等。

[引用信息] 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.

情境理解与生成、自然语言交互、复杂模式识别与分析、个性化反馈及高效场景模拟等技术的进步，为场景化评估提供了强有力的技术支持，使其能走向个性化、规模化与智能化，破解教师能力评估中情境简化、过程忽略、反馈滞后与支持有限等困境，推动教师能力评估从评判向赋能的范式转变。

二、基本内涵

（一）定义与特征

场景化评估，亦称基于场景的评估，是一种依托真实性评价与证据中心设计理论，以情境为载体、任务为驱动、行为表现为核心证据的创新评估方法。真实性评价强调在拟真情境中激发教师的实践性知识与深度反思，证据中心设计理论构建了“从行为证据到能力推断”的评估逻辑。这两种理论的融合应用，再结合人工智能技术的支持，场景化评估能实现对复杂能力的全方位分析与评价。

围绕场景化评估，顾小清（2024）提出场景化测评是通过深入模拟、整合真实情境来全面评估个体在特定情境中的认知行为、决策能力或高级思维等的测评方式。郭炯等（2025）将其定义为依托真实任务情境，通过采集学习者行为流数据，结合学生模型、任务模型与证据模型实现多维度能力测评的评估范式。在语言学习领域，研究者认为这是一种基于技术的创新评估方法，试图将评估与环境（如学校）、角色（如同学和老师）和总体目标（如完成科学博览会的项目）相结合，通过模拟现实生活的语言使用环境激发学习者独立的和综合的语言技能（Banerjee, 2019）。可见，场景化评估强调通过模拟真实情境、采集多源数据并运用多种分析模型，实现对能力的多维度诊断，具有场景真实性、数据融合性与反馈及时性三大核心特征。

1）场景真实性。这是场景化评估区别于传统标准化测试的核心特征。在教师能力评估中，构建真实或高仿真教学情境，让评估任务反映测评对象的实际环境和条件，能提高评估的表面效度和生态效度。但真实性不限于物理拟真，更强调逻辑约束和情境压力，以使测评结果具有现实迁移价值。

2）数据融合性。场景化评估侧重观察和评估受测者在模拟情境中实际做了什么（行为表现）及如何做（动态过程），然后通过分析多源数据确保结果全面和客观。这些多源异构数据包括认知、行为、情感、教师教学行为日志与教学设计方案等，再结合教师的反应、操作步骤、解决方案、决策过程和任务结果等进行分析和判断。

3）反馈及时性。场景化评估能实现及时、个性化的分析和反馈，方便教师对照目标自我修正和反思，并逐步学会自我监控、自我评估与自我调节，激发其学习内驱力，进而实现深度学习。

（二）评估范式与场景化评估定位

教师能力评估范式随着教育理念、学习方式和技术条件的演变而不断演进，大致可分为四种：基于心理测量学的标准化测试、成果导向的绩效评估、情景嵌入的发展性评估和人机协同的智能化评估。

基于心理测量学的标准化测试将教师能力视为一系列可观测、可量化的孤立行为或知识点的集合，评估方式以纸笔测验（如学科知识测试）和结构化观察量表为主，追求客观、统一和常模参照，便于大规模实施，但存在主观性强、易受自我认知偏差影响等局限（Paulhus et al., 2007）。成果导向的绩效评估以教师能力模型为核心，在界定优秀教师应具备的知识、技能和特质基础上，依据教师实践表现与教育教学成果（如教学设计、课堂实录、学生成果等）开展评价。该范式依赖评估者的专业素质，成本高，难以规模化实施。情景嵌入的发展性评估以促进教师反思、改进与专业自主为目的，强调融入教师生活情境及过程性表现，关注教师的整体素质及未来规划，有助于实现从评判到发展的功能转向（刘尧, 2001）。它虽然可以借助数字技术进行诊断和分析，但存在实施路径不明、标准不清晰和开发成本较高等问题。智能技术的快速发展使得教育评价迈入人机协同的崭新阶段（郭炯等, 2025），教育评价既能兼顾效率、规模与客观性，又具备支持诊断、反馈与赋能发展的潜力。

场景化评估本质上是对上述范式的整合和升华（见图 1）。它融合了发展性评估的赋能价值导向与智能化评估的动态增强技术特质，旨在解决传统发展性评估难以规模化、精准化的痼疾，即通过智能技术的运用，低成本、自动化创设真实、互动与复杂的情景，实现大规模个性化评估，再依托多模态的过程性数据分析技术，持续捕捉教学行为、

认知决策与情感互动等证据，在真实的情境中动态评估教师的知识转化与实践应变能力。这种方式不仅能超越标准化测试的语境剥离与绩效评估情景简化的不足，有效衡量教师将理论知识转化为教学实践的水平，更致力于将单一评价工作转化为促进教师持续发展的支持活动，推动评价活动从评判转向赋能，是践行“评价促发展”理念的创新路径。

三、技术机制

（一）GenAI 的应用

人工智能为教育评估突破现有局限带来可能。经济合作与发展组织强调生成式人工智能在提供即时、个性化反馈方面的巨大潜力：可支持持续的形成性评价，帮助教师调整教学以满足学生个体需求（OECD, 2023）；动态内容生成、多模态数据融合与个性化反馈能力可用于开发多元评价方法（蒋慧芳等, 2025）；监测、挖掘和分析学生项目作业、角色扮演、实际场景模拟等数据，更好地了解学生的学习过程和效果（张峰等, 2023）。

大语言模型的内容生成能力也被广泛应用于自动化题目构建。一些生成式人工智能工具（如 Eduaide.ai 和 Quizgecko）能依据教学目标与指定内容，便捷生成多选题、判断题、简答题等，如麻省理工学院利用人工智能自动生成同一概念不同难度的问题来实现个性化评估（Circi et al., 2023）；能实现对批判性思维等的测评，帮助学生在各学科领域乃至跨学科领域将课程知识与现实情境相结合（冷静等, 2024）。大语言模型的快速发展，推动题目自动生成向更智能、更灵活的方向迈进，且能生成较好的题目质量（韩雨婷等, 2025），确保与教学目标更匹配；还可创建可扩展、适应性和包容性的评估，满足不同学生群体接受公平评估的需求（Kuang et al., 2024）。教学情境模拟是人工智能应用于评估的重要方向。有研究通过在三维课堂环境中嵌入具有自适应交互能力的学生智能体，构建面向复杂问题解决的拟真教学情境（Lim et al., 2025），还有研究指出，虚拟教师在建立社会临场感方面，与真实教师相比仍存在差距（Xu et al., 2025）。

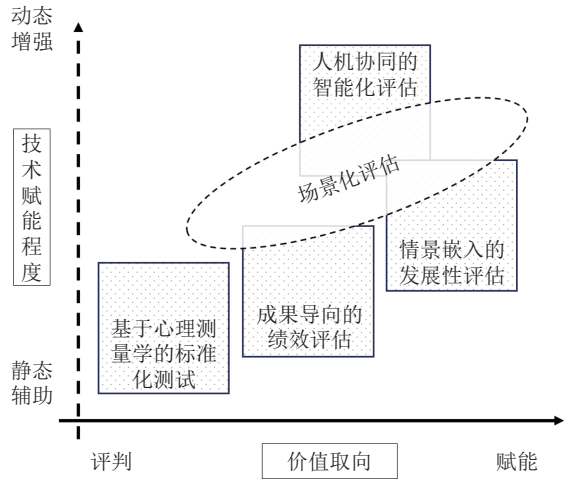

图 1　场景化评估在评估范式谱系中的定位

面对人工智能应用带来的人才培养新要求，联合国教科文组织（UNESCO, 2023）强调，教育系统需重新设计评价体系，更加注重价值观、基础知识和技能、高阶思维及与人工智能协作所需的职业技能，同时明确提出教育评估的最终判断和责任必须由人类教师承担，并要求对人工智能系统进行严格的伦理审查，确保符合教育的基本价值观。

（二）GenAI 赋能教师能力场景化评估的路径

1）复杂情境创设与规模化

基于生成对抗网络（GANs）、变分自编码器（VAEs）等深度学习技术，生成式人工智能可通过对抗训练与潜在空间学习实现高质量内容生成与情境建构，动态创设贴近真实实践的复杂任务环境，且虚拟教师在语音、交流与外观生成质量上亦达到一定水准，促进的学习成效已与真实教师相当（Xu et al., 2025），为创建模拟环境、基于情景的任务问题解决环境提供了可能（Ilieva et al., 2025）。

2）评估任务的动态调整

通过语义理解与自适应生成，生成式人工智能能根据被评估者的经验背景、能力水平和发展阶段，动态调整任务难度，实现高度个性化的测评。这不仅能提升被评估者的动机与参与度，还能使反馈更契合实际需求（Arslan et al., 2024）。

3）多模态数据分析与交互

生成式人工智能基于多模态 Transformer 架构，能够统一编码和处理不同模态的数据流，包括文本对话、语音、面部表情、互动日志等，并通过跨模态注意力机制建立模态间的语义关联（Radford et al., 2021）。在教师能力评估中，系统可同步捕获教师的语言表达、非言语行为、认知过程和互动模式，形成教师行为的多维证据，进而揭示教师在复杂情

境中的思维路径与决策逻辑。

4)即时性与发展性反馈

生成式人工智能凭借上下文感知与推理生成能力，能支持评估过程中的近实时分析与反馈生成。在教师能力评估中，系统能从教师的多模态行为数据中提取稳定模式，识别关键优势与薄弱环节(Fütterer et al., 2026)，生成诊断报告与发展建议。这使评估真正成为教师专业成长的驱动力。

四、设计与实施

教师能力评估是一项基于证据的、对专业实践智慧进行推断与解释的系统性过程，场景的核心在于通过对现实世界的模拟或假设，为人与环境的互动提供行动框架。教师能力场景化评估要实现评判向赋能的转向，关键是通过构建真实、复杂和互动的教学实践场景，引发教师产生真实的认知和行为，从而为分析和诊断提供客观依据。结合生成式人工智能技术的特性，本研究提出教师能力场景化评估设计与实施的核心原则和系统架构。

(一)核心原则：人机协同

教师能力评估本质上是一种富含教育理念、实践智慧、伦理关怀与情境判断的专业性工作。研究人员普遍认为，必须将人工智能的高效生成、深度分析能力与人类的专业洞察、价值判断相结合，构建人机协同、优势互补的评估新范式。因德兰等(Indran et al., 2024)提出人机协同框架应由精心设计的提示词引导生成初步内容，然后由专家进行质量评估和必要修正。韩雨婷等(2025)认为，“AI辅助、人类主导”的协作方式不但保留了人工智能的效率优势，而且有效融合了专业人员的知识判断，特别适合对内容质量要求严格的高利害测验情境。

在教师能力场景化评估中，人机协同原则的核心实现机制是：技术负责规模化情境生成、数据采集与实时分析，并在输出端构建“证据追溯机制”；人类专家主导教育意图融入、复杂教学智慧的综合研判、评估规则制定及伦理价值的把握，最终实现“人类主导→AI执行→人类确认”的协同。在评估设计阶段，专家负责定义能力标准、设计核心评估框架，人工智能生成多样化场景及测评内容。在此基础上，人类再审核评估内容的教育适当性，并确认证据链的合理性。在评估实施阶段，专家提出标准，生成式人工智能依据标准采集数据，进行智能化分析，识别问题和模式，并自动生成包含“情境—行为—指标”映射关系的诊断报告(即可追溯的“证据链”)。在此基础上，专家进行情景化校准、深度解读和价值判断，最终形成对教师能力水平的质性评价与个性化发展建议(见图2)。

(二)系统架构：三层协同模型

为了实现“从行为表现到能力评判再到发展赋能”的核心目标，场景化评估必须完成两个关键环节：场景构建(场景创设与生成任务)与场景解析(数据采集与任务分析)。稳健、高效地支撑“发展赋能”的转向目标，需要构建层次清晰、功能衔接的系统性技术架构。

数据层：全面、原始地记录教师背景性数据，以及评估过程产生的过程性数据和成果性数据，并对涉及的多模态数据进行采集、存储与预处理，构成能力评估及发展需求推理的原始证据；存储构建能力评估所依据的相关标准、典型场景和任务案例，以支持不同场景生成方式。生成式人工智能通过多模态理解技术和分析模型，可将非结构化数据转化为可分析的语义信息。

模式层：负责从原始数据中提取证据，进而分析、推理与决策，将证据与内在的、不可直接测量的能力构念联系起来。依据证据中心设计评估理论，模型可分为能力型、任务型和证据型。能力模型定义评估对象，任务模型依据评估能力定义激发目标能力行为具体情境的创设要求与规则，证据模型设定证据规则和测量模式。生成式人工智能利用强大的语义分析、情感计算、模式识别等能力，从复杂、自然的行为数据中提取证据并进行推理，为评估情境的“无限供给”和“个性化定制”提

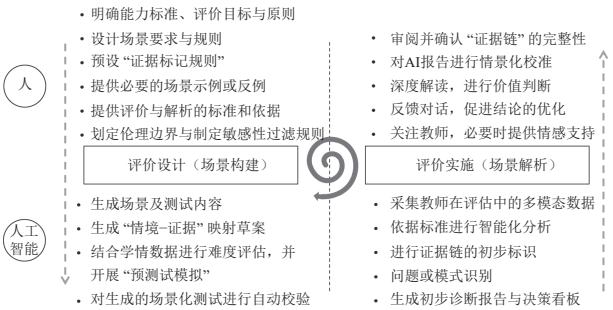

图2　人机协同的评估设计与实施原则

供设计蓝图和规范。

应用层：将抽象的评估设计转化为教师可感知、可交互的具体场景任务与反馈报告，是评估发生的“交互界面”。该层包括四个组件：动态场景智能生成器、自适应评估引擎、智能反馈与推荐系统和人机协同决策看板。生成式人工智能在统一框架下生成海量个性化情境内容；基于教师实时表现，动态构建适配的能力诊断路径；即时分析行为数据，生成诊断报告并推送匹配的学习资源与练习，提供发展性指导；提炼分析结论、关键片段与证据链，辅助人类专家进行高效复核与干预决策。

三层模型（见图 3）体现了高内聚、低耦合的工程设计原则，共同构成生成式人工智能赋能教师能力场景化评估方法：数据层负责多源证据的采集与组织，模式层承担能力构念建模与证据推理，应用层面向教师呈现具体任务与反馈结果。这一模型为重新审视和重构教师能力评估提供了范式。

五、实施路径与关键策略

（一）场景评估的智能生成

特拉伊科夫斯基等（Trajkovski et al., 2025）基于人机协作方式将生成式人工智能辅助的评估分为人工智能自主生成、人工智能与教师的协作创造和人工智能驱动增强三种。本研究将生成式人工智能辅助教师能力评估的场景生成方式分为非结构化指令驱动生成、结构化框架引导生成和动态指令驱动的自适应生成，关键要素和人机协同策略见表 1。

1. 非结构化指令驱动生成

非结构化指令驱动生成指由人类向人工智能提供整体性、自然语言形式的指令，从而生成多样化情景试题。此类指令常未对评估目标、任务结构或评价标准进行显式拆分，而是以语义描述的方式提出总体要求。人工智能需理解指令意图，自主补全场景构建及评估所需要素。为确保证据推理的透明性，人类专家需对人工智能生成的场景进行“证据回溯”式审核，即审视人工智能自主补全的情境要素是否隐含与预设能力标准相对应的表现机会，将人工智能的生成锚定在可解释的证据框架内。该方式运用自然语言生成模型和深度学习算法，独立生成评估方案，生成效率高、灵活性强，适用于评估设计的初步构想或示例性情境建构，但生成内容的可控性和一致性相对有限（见表 2）。

在该案例中，人工智能可依据清晰的提示工程，生成贴合中小学课堂真实教学情境的类型试题，以适应不同对象的需求。已有研究表明，提示工程作为一种更轻量级的方法，无需改变模型参数，而是通过设计合适的提示引导模型生成目标内容，有助于提升题目自动生成的质量（韩雨婷等, 2025）。

2. 结构化框架引导生成

结构化指令引导生成指人类以结构化的指令形式明确评估的关键要素和生成约束，如评估目标、

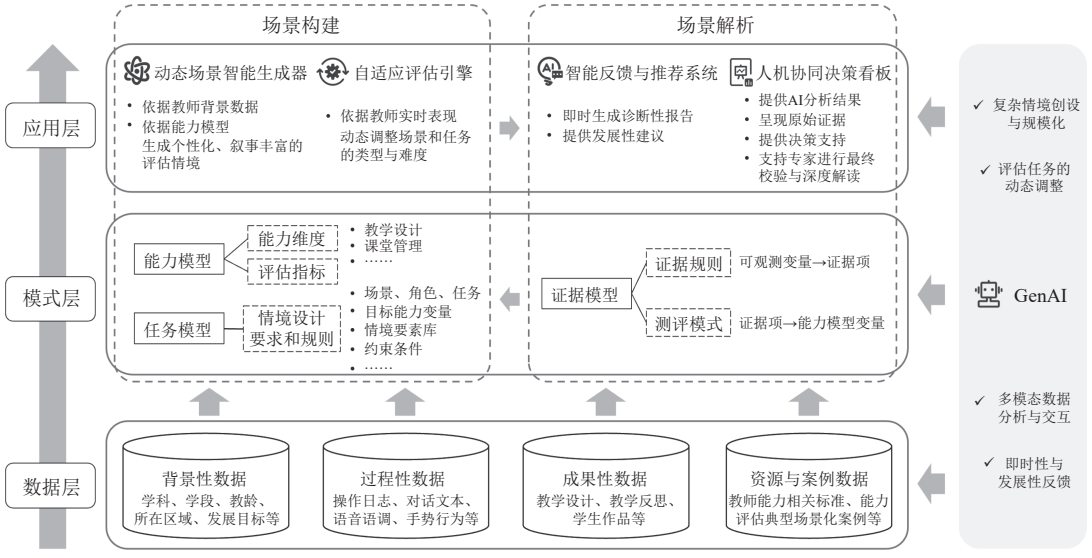

图 3　教师场景化评估的三层协同模型

表 1　GenAI 辅助场景化评估生成的三种方式

分析要素	非结构化指令驱动生成	结构化框架引导生成	动态指令驱动的自适应生成
概要	GenAI 根据人类提供的整体性、自然语言形式的提示语指令生成场景和评估内容	人类先制定包括场景化评估的关键要素和约束条件，形成具体框架，再由 GenAI 生成	预先设定结构化指令，GenAI 根据学习者的个人信息及学习和评估的具体表现进行针对性设计与调整
人类职责	定义场景需求 设定场景依据/标准 审核场景和问题的真实性、合理性与教学适配性	提炼场景关键要素、任务模块和评价维度，并根据需要提供示例、真实案例素材 调整场景复杂度	提供原始场景与评价维度 监控自适应调整过程 解释与确认评估结果
人工智能职责	理解指令语义 自动生成情景化试题 构建虚拟环境 生成多语言/无障碍版本	根据框架填充动态细节 角色扮演对话生成 自动化场景分支（根据选择触发不同情景）	行为数据收集与模式识别 动态调节难度或测试路径 标记潜在问题 生成对比视图（标准操作与用户操作）
设计路径	人类发出非结构化指令→人工智能生成场景试题→人类复核	人类设计框架→人工智能批量填充内容→人工调整→人工智能再优化	人类提供初始化指令与数据→人工智能分析并推荐优化→人工确认→人工智能实施调整
技术差异	依赖自然语言生成、深度学习算法、知识图谱	需要模板解析和语义理解	基于数据分析和动态决策机制

表 2　非结构化指令驱动生成示例

指令	生成结果（deepseek 生成）
作为教师培训专家，你正在给新教师讲授学习支架的概念和应用。可否请你从布鲁姆认知目标分类的应用、分析、综合、评价层面，设计一道情景化测试题，以考察他们对支架的理解，可以是单选题、判断题或多选题？	**情境：**王老师在教小学五年级学生写“说服性议论文”时，发现学生难以组织论点与论据，于是设计了一个“写作思维导图”模板，要求学生先填写核心观点，再列出三个理由，并为每个理由补充一个例子。 **问题：**王老师使用的“写作思维导图”模板属于哪种类型学习支架？主要对应布鲁姆认知目标分类的哪一层次？ **选项：** A. 程序性支架——应用层次；B. 概念性支架——分析层次；C. 策略性支架——综合层次（正确选项）；D. 元认知支架——评价层次

场景要素或评估维度等，生成式人工智能不再需要自行推断评估结构，而是在既定指令结构的引导下生成具体情境与评估内容。这种结构化但灵活的评估创建方法，依赖精心设计的内容结构，人工智能根据指令生成特定内容，或在预设框架内替换、调整具体要素和细节。这种方法既保留了题目自动生成技术“基于规则的方法”的要领，依赖专家知识和认知理论，通过框架预定义规则或模板控制生成题目，又发挥了生成式人工智能强大的语言理解和生成能力，以结构化框架为基础生成针对特定目标的多样化问题，并保持测试在构念和要素上一致。这种方法使生成结果更易于对齐教学目标与评价标准，适用于场景化评估的系统性设计。例如，针对教师“选择和运用评价数据采集工具”的评估，评估者可根据拟定的框架形成具体的提示语内容，并依据测试对象的学科、所在区域、能力水平等生成评估场景。评估框架的构成与示例见表 3。

3. 动态指令驱动的自适应生成

人工智能在快速分析大量数据方面表现出色，非常适合创建根据学生表现动态调整的自适应评估（Wang et al., 2020）。生成式人工智能接收的不仅是预先设定的指令，还包括随评估进程不断更新的指令。这些指令可根据学习者的表现、作答路径或反馈动态调整，从而引导生成式人工智能生成与个体相匹配的评估情境与任务。例如，学习者如果完成建构主义基本观点问题的解答，接下来可能会有更难的问题检验其对建构主义的理解，或结合某个真实问题对其考察。相反，如果学习者表现不佳，系统可能会提供较简单的题目以找出薄弱环节。自适应评估由状态跟踪、表现预测、策略决策和内容生成四个核心模块组成。当然，动态指令的更新并非人工智能自主推断，而是遵循人类专家预设的“任务选择规则”或决策树逻辑。人工智能根据这些透明的、预设的规则，实时生成与学习者状态相匹配的情境。这种方式将证据设计的路径牢牢掌握在人类专家设定的规则之内，人工智能则用于确保情境的丰富性和适应性。

动态指令驱动的自适应生成能够与学习过程

表 3　结构化框架引导生成示例

构成		描述	示例
评估任务		确定场景化评估的输出目标与核心意图	生成一套能有效甄别教师在复杂技术环境下，是否掌握“选择和运用评价数据采集工具”核心能力的场景化测试题，确保题目具有高区分度、高真实度，并紧密贴合能力核心特征中的行为表现。
评估内容	能力描述	准确定义评估的内容	“选择和运用评价数据采集工具”指教师基于教学目标，能够在数字化教学环境中，科学地选择、组合与运用多模态数据采集工具的能力。它不仅包括对传统测评与观察工具的合理使用，更强调人机协同下评价工具的深度融合。
	能力表现	明确能力的核心特征和行为表现，细化评估的具体要求，保证评估效度	生成的题目必须涉及以下评估维度：1）目标界定精准（匹配评价目标和内容、数据来源合理、评价流程设计清晰），2）工具选择恰当（根据工具特征和情境判断工具适用性），3）工具运用规范与灵活（告知学生数据采集事项、有序组织学生的数据采集、动态管理数据采集方案、检验数据的质量）。
评估场景	环境	描述事件发生的教学和技术环境（可变度最高）	预设一个具体到学科、学段、教学内容的教学环境，技术环境可以包含多媒体教学环境、智慧教室、学习管理系统、在线学习平台、人工智能评估系统等。
	行为主体	描述事件中的主要参与者，通常是教师本人（可变）	角色以教师为主体，同时可自然带入教师的个体属性与环境属性，设计符合教师实际需要的情景。
	教学事件	聚焦常规教学流程中的某个环节或具体情境	描述教师需要选择、组合和运用评价数据采集工具获取支持的具体情境，如问题分析、教学改进、学习成效评估等。
	触发冲突	描述触发决策和行动的关键冲突或异常困境（可变）	设计一个关键冲突，如数据解读偏差、工具使用不当、学生参与度不均等。
输出规范		规定评估题目的展现形式	请输出测试题，包含结构化的场景文本、针对冲突点的决策任务选项和评估标准。

深度耦合。它通过即时捕捉和分析教师教学情境的表现数据，动态生成个性化的评估任务与反馈，从而精准揭示教师在教学设计、课堂应变或学科理解等方面的能力状态和待改进方面。例如，系统可根据教师处理课堂提问的方式，自动调整后续模拟教学的复杂度，引导其深化对“因材施教”的理解。这种高度定制化的评估不仅能避免标准化测试带来的局限，而且能发现测评对象的潜力，提升其自我效能感，最终将评估从静态测量转变为动态、综合的学习过程（Algerafi et al., 2023）。

上述三种方式中，生成式人工智能扮演了“情境创设引擎”的角色，但其运行始终遵循“人类主导意图与规则，人工智能负责情境生成与初析”的协同框架。核心优势在于能够根据预设的评估参数与证据规则，生成丰富、个性化、动态演进的模拟情境，并同步输出可追溯的“情境—证据”映射关系。这种方式超越统一、标准化的测试，能增强测评的真实感和体验感，提升评估场景设计的效率和质量，将人类专家从繁琐的情境细节构思中解放出来，使其聚焦更核心的证据解读、价值判断与发展建议，从而支持评估活动实现从“单一评判”走向“赋能发展”的协同目标。

（二）数据的融合分析

场景化评估的另一关键是依托系统性的技术流程，将情境互动中形成的多源过程数据转化为结构化证据，并据此形成可解释的教师专业能力报告。

1. 数据采集与预处理

系统应自动、无干扰地采集多维度过程性数据，包括对话文本、语音语调、表情、决策时间序列、虚拟环境操作路径等。这些数据共同刻画了教师在复杂情境中的认知、行为与情感反应，是智能化评估的起点。然而，原始数据具有异构、冗余与含噪等特性，必须经过系统化的预处理，包括数据清洗、时间对齐、语义标注与标准化，为后续多模态融合与高阶推理奠定基础。

2. 多模态数据融合

多模态数据融合是实现对教师复杂能力构念精准推断的关键技术步骤，其目标是将异构数据映射到统一的语义空间（即让不同模态的数据在数学表征上具有可比性），从而捕捉跨模态的互补与协

同信息。常用方法有特征融合、决策融合和混合融合（任泽裕等, 2021）。通过三种范式的融合应用，评估系统能模仿人类专家的专业分析方式，实现对教师行为全面、深刻的表征。

3. 行为模式识别与推断

完成多模态数据的融合与表征后，系统从整合后的行为语义中识别有价值的模式，并依据预设的理论框架，将这些行为语义推断为内在的、潜在的能力特质。该环节遵循证据中心设计的理论，先从教师的行为序列中挖掘稳定、有意义的过程性证据，基于证据规则与测量模型，将证据转化为能力度量，进而识别典型行为模式和教师能力水平。

4. 人机协同分析

系统基于行为模式识别与推断结果，关联整合关键证据形成证据链，由生成式人工智能依据标准生成结构化诊断要点与结论，并在人机协同决策看板中呈现。专家核查证据、开展情境化校准、修正偏差、补充情境解释与价值判断，最终形成可解释的教师专业能力诊断报告和改进建议。

六、创新实践与案例解析

人机协同的评估设计与实施作为教师专业发展的重要创新方向，近年来涌现出一系列实践案例。这些实践不仅验证了生成式人工智能在教师能力评估中的技术可行性，也明晰了场景化评估的设计思路与实施路径。为呈现前述框架的运作逻辑与实际效果，本研究以某高校的技术赋能教师发展工作为例，阐释理论框架转化为可操作的评估实践。

（一）动态情景判断测试

情境判断测试（situational judgment test）通过模拟实际工作情境评估个体胜任力，被认为在能力测评方面有较高的效度（Motowidlo et al., 1990），是测量个体胜任力的有效工具（漆书青等, 2003），但用于教师发展领域存在明显困境：有限的静态情境库难以覆盖教学实践的多样和复杂性，且保真性弱。

某高校的教师数字素养动态情景判断测试系统，基于生成式人工智能技术，通过动态场景生成与个性化适配机制实现了模式突破：人类专家先提出主题框架，包括教师数字素养维度的能力描述、评估目标、具体事件、核心冲突以及教学环境与教师的关键属性，明确试题生成规则，如优先选择与人工智能、智慧学习环境、数据驱动教学相关的高频场景；依据教师的学科、学段、角色等，采用框架引导生成方式自动生成高度拟真、贴合教学实践的情景与试题；依据预设标准，实时生成个性化测评反馈，并结合教师发展目标推送资源。

（二）模拟对话互动评估

基于对话的评估（conversation-based assessment）（Yildirim-Erbasli et al., 2021）的思路与苏格拉底所倡导的“诘问法”一脉相承，是一种深入探究教师内隐知识和实践智慧的质性评价方法。美国西方教育研究实验室主导的“数据素养访谈评估项目”通过将教师置于模拟的、富有挑战性的工作情境中，借助对话观察和分析教师在真实教育情境中较真实的数据使用能力和态度（李艳等, 2020）。

2024 年推出的“知心慧语”师生沟通能力实训系统，采用生成式人工智能技术，通过角色模拟、语义深度解析与多模态证据整合重塑对话评估。第一，生成式人工智能基于儿童的不同气质沟通逻辑和语言风格，采用框架引导生成的方式创建不同的场景和角色对话，使基于对话的测评更具真实性和沉浸感；第二，利用大语言模型的深度语义分析能力，实时解构教师对话中的专业术语运用、问题解决策略、共情表达、决策逻辑等隐性能力指标，使能力分析更具专业深度；第三，详细记录教师对话情景的语言、行为和操作等表现，基于对话过程捕捉教师的思维痕迹和决策链，再结合语音情感开展的情绪分析，识别教师的行为模式，为教师能力诊断和提升提供支持。

（三）基于实训环境的技能评估

操作模拟测试曾被广泛用于计算机操作技能测试，然而应用空间有限：第一，测评内容有限，高成本导致模拟环境有限、敏捷性不足，难以匹配工具快速更新速度；第二，操作路径僵化，抑制个性化和创新解法；第三，聚焦软件或工具操作，未关联真实教育场景；第四，无法捕捉操作背后的决策逻辑。

在教师提示语工程测评与实训系统中，教师可自主选择不同的大语言模型环境，针对复杂的教育实践任务（如大单元教学方案设计、个性化作业设计等），通过多轮交互（包括提示词设计、需求补充、迭代改进、结果优化等）完成任务。在此过程中，系统自动记录交互过程，依据设定的任务类型，记

录输入的提示语内容，将机械界面点击记录转变为操作语义分析，实现多元路径的灵活认定，并从准确性、一致性、相关性、效率等维度评估用户的提示词设计能力和动态交互能力。目前该系统可结合教师的差异化需求与实践问题生成针对性任务，并根据教师操作成效进行动态调整，通过自适应生成场景的方式完成基于场景的实训。

七、挑战与未来图景

本文结合教师能力评估的现实困境和生成式人工智能技术特性，提出人机协同的场景化评估设计与实施新思路。生成式人工智能赋能的场景化评估不仅是技术工具的创新，更是教师能力评价范式的重构。它使评估从“标准检验”延伸为“情境建构”，从“专家中心”过渡到“人机协同”，从“能力判定”转向为“成长赋能”。在这一过程中，教师得以在高拟真的情境中持续开展实践反思与能力迭代，并获得精准、即时、发展性的专业支持。这种评估模式既回应了教育数字化转型的新要求，也为破解教师能力评价中长期存在的“学评分离”“情境脱嵌”“反馈滞后”等难题提供了可行路径。

尽管生成式人工智能在还原真实场景的有效性方面有待验证，但随着多模态大模型、教育知识图谱、情感计算等技术的深度融合，其赋能的场景化评估将迎来系统性变革。未来，该领域有望在自适应情境生成、多模态场景生成、跨模态证据融合、认知情感协同诊断等关键方向上取得突破，实现从“静态辅助”向“动态增强”的根本性演进。

［参考文献］

[1] Algerafi, M. A. M., Zhou, Y., Oubibi, M., & Wijaya, T. T.(2023). Unlocking the potential: A comprehensive evaluation of augmented reality and virtual reality in education[J]. Electronics, 12(18): 3953.

[2] Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D.(2024). Opportunities and challenges of using generative AI to personalize educational assessment[J]. Frontiers in Artificial Intelligence, 7: 1460651.

[3] Banerjee, H. L.(2019). Investigating the construct of topical knowledge in second language assessment: A Scenario-Based Assessment Approach[J]. Language Assessment Quarterly, 16(2): 133-160.

[4] Circi, R. , Hicks, J. , & Sikali, E. (2023). Automatic item generation: foundations and machine learning-based approaches for assessments[C]//Frontiers in Education. Frontiers Media SA, 8: 858273.

[5] Darling-Hammond, L. , Jaquith, A. , & Hamilton, M. (2012). Creating a comprehensive system for evaluating and supporting effective teaching[M]. Stanford, CA: Stanford Center for Opportunity Policy in Education: iv.

[6] Fütterer, T., Hou, R., Bühler, B., Bozkir, E., Bell, C., Kasneci, E., Gerjets, P., & Trautwein, U. (2026). Validating automated assessments of teaching effectiveness using multimodal data[J]. Learning and Instruction, 101: 102264.

[7] 顾小清 (2024). 基于场景的测评：内涵特征、实践应用与未来展望 [J]. 上海教育(8), 30-33.

[8] 郭炯，邹佳人(2025). 场景化评价：技术赋能新时代教育评价改革的新趋向 [J]. 中国远程教育, 45(1): 71-85.

[9] 韩雨婷，王文轩，刘红云，游晓锋(2025). 题目自动生成的技术革新与现实挑战 [J]. 心理科学进展, 33(10): 1766-1782.

[10] Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S.(2025). A framework for generative AI-driven assessment in higher education[J]. Information, 16(6): 472.

[11] Indran, I. R., Paranthaman, P., Gupta, N., & Mustafa, N.(2024). Twelve tips to leverage AI for efficient and effective medical question generation: A guide for educators using Chat GPT[J]. Medical Teacher, 46(8): 1021-1026.

[12] 蒋慧芳，曾文婕(2025). 生成式人工智能推动教育评价转型 [J]. 中国教育学刊, (8): 41-48.

[13] Kuang, Y. , Tang, Y. , & Xie, T. (2024). Objectives, methods, and challenges of applying intelligent assessment in education: A systematic review[C]//2024 International Symposium on Educational Technology (ISET). IEEE, 185-190.

[14] 冷静，卢弘焕，代琳(2024). 生成式人工智能赋能批判性思维测评——基于 ChatGPT 的应用实验 [J]. 现代远程教育研究, 36(6): 102-111.

[15] 李艳，刘淑君(2020). 国外教师数据素养测评研究及启示 [J]. 开放教育研究, 26(1): 37-49.

[16] Lim, J., Lee, U., Koh, J., Jeong, Y., Lee, Y., Byun, G., Jung, H., Jang, Y., Lee, S., & Moon, J.(2025). Development and implementation of a generative artificial intelligence-enhanced simulation to enhance problem-solving skills for pre-service teachers[J]. Computers & Education, 232: 105306.

[17] 刘尧(2001). 发展性教师评价的理论与模式 [J]. 教育理论与实践, (12): 28-32.

[18] Motowidlo, S. J., Dunnette, M. D., & Carter, G. W.(1990). An alternative selection procedure: The low-fidelity simulation[J]. Journal of Applied Psychology, 75(6): 640-647.

[19] OECD (2023). OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem. Paris: OECD Publishing: 110.

[20] Paulhus, D. L., & Vazire, S.(2007). The self-report method[J]. Handbook of Research Methods in Personality Psychology, 1(2007): 226-

233.

[21] 漆书青，戴海琦(2003). 情景判断测验的性质、功能与开发编制 [J]. 心理学探新，(4)：42-46.

[22] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I.（2021）. Learning transferable visual models from natural language supervision[C]//International conference on machine learning. PmLR, 8748-8763.

[23] 任泽裕，王振超，柯尊旺，李哲，吾守尔·斯拉木(2021). 多模态数据融合综述 [J]. 计算机工程与应用，57(18)：49-64.

[24] Trajkovski, G. , & Hayes, H.（2025）. The AI-Assisted Assessment Creation Framework[M]//AI-Assisted Assessment in Education: Transforming Assessment and Measuring Learning. Cham: Springer Nature Switzerland, 59-114.

[25] UNESCO(2023). Guidance for generative AI in education and research[M]. Paris: UNESCO: 41.

[26] Wang, N., Wang, D., & Zhang, Y.(2020). Design of an adaptive examination system based on artificial intelligence recognition model[J]. Mechanical Systems and Signal Processing, 142: 106656.

[27] Xu, T., Chen, Q., Zhang, Z., Dong, B., Zhang, H., Bai, J., & Zhou, Y.(2025). Maximizing effectiveness of AI-generated instructors through human-like behavior and dynamic visual cues in instructional videos: Evidence from an eye-tracking study[J]. The Internet and Higher Education, 67: 101034.

[28] Yildirim-Erbasli, S. N. , & Bulut, O.（2021）. Conversation-based assessments: Real-time assessment and feedback[J]. ELearn, 2021(12).

[29] 张峰，陈玮(2023). ChatGPT 与高等教育：人工智能如何驱动学习变革 [J]. 重庆理工大学学报(社会科学)，37(5)：26-33.

（编辑：赵晓丽）

GenAI Empowering Scenario-Based Assessment of Teacher Competence: Framework, Pathways, and Practice

WEI Fei[1], YANG Kexin[2] & YANG Shuting[2]

（1. *East China Normal University*, *School of Open and Learning Education*, *Shanghai* 200062, *China*； 2. *Department of Education Information Technology*, *East China Normal University*, *Shanghai* 200062, *China*）

Abstract: *Scenario-based assessment offers a new pathway for shifting teacher competence evaluation from judgment to empowerment. Firstly, this paper clarifies the connotation, characteristics, and value Orientation of scenario-based assessment for teacher competence, and explores the status and technical mechanisms of generative artificial intelligence that empower educational assessment. Building on this, a framework and pathways for GenAI-empowered scenario-based assessment of teacher competence are constructed. Guided by the principle of human-computer collaboration, the framework adopts a dual perspective of design and implementation. It establishes a three-layer architecture comprising a data layer, a pattern layer, and an application layer to support the construction and implementation of assessment scenarios. In the implementation pathways section, the paper explains three approaches to generating assessment scenarios for GenAI-assisted assessment: unstructured instruction-driven generation, structured framework-guided generation, and dynamic instruction-driven adaptive generation. It also elaborates on four key data integration and analysis steps: data collection and preprocessing, multimodal data fusion, behavioral pattern recognition and inference, and human-computer collaborative analysis, thereby supporting dynamic scenario generation and multidimensional data analysis. Finally, the paper illustrates how to translate the theoretical framework into actionable assessment practices through three typical cases in teacher development.*

Key words: *scenario-based assessment*; *teacher competence assessment*; *generative artificial intelligence (GenAI)*; *human-computer collaborative assessment design*; *human-computer collaborative assessment implementation*