

学评融合：教育评价的理念转向与落地

辛涛^{1,2} 聂竹明² 施文静³ 张生¹

(1. 北京师范大学中国基础教育质量监测协同创新中心, 北京 100875; 2. 安徽师范大学教育科学学院, 安徽芜湖 241000; 3. 江苏省昆山市第二中学, 江苏昆山 215300)

[摘要] 技术发展推动教育生态变革, 要求评价功能从选拔转向促进学习, 学评融合成为回应个性化教育发展与技术赋能的必然选择。学评融合理念的核心在于范式转向: 评价目的转向促进发展, 主体转向学习者参与, 技术转向动态诊断。其本质是评价的学习性与诊断性统一, 评价过程即学习过程, 旨在通过精准剖析能力实现以评促学; 核心目标是从知识技能评价转向核心素养导向评价。驱动该理念落地的核心技术路径是个性化学习引擎, 即构建闭环系统: 数据层整合多源异构数据, 诊断层应用认知诊断模型动态解析知识状态并预测趋势, 推荐层依托算法生成个性化学习路径, 形成“数据采集→状态追踪→智能推荐→反思改进”的循环, 将评价重构为认知元监控过程, 使之进化为素养生长引擎。然而, 该路径面临三重挑战: 技术效度需平衡算法复杂性与可解释性, 构建可解释人工智能框架以提升信任度; 技术伦理需协调数据价值与隐私安全, 应对大规模远程测评风险; 资源普惠需破解数字鸿沟衍生的公平困境, 防止技术资源配置失衡等。

[关键词] 学评融合; 素养评价; 教育公平; 个性化推荐; 评价伦理

[中图分类号] G40-058.1

[文献标识码] A

[文章编号] 1007-2179(2025)05-0030-09

一、教育新生态呼唤教育评价变革

第四次工业革命浪潮下, 人工智能、大数据与区块链技术的聚合效应正从底层逻辑熔铸传统教育生态。联合国教科文组织《教育 2030 行动框架》与中国《教育现代化 2035》均指出教育生态变革是回应技术赋能驱动的社会形态跃迁、产业结构升级与高质量人才供给的系统性应答。作为教育体系核心环节的教育评价, 也正随之发生深刻而广泛的变革。

(一)对“未来教育”的回应

教育形态随着社会发展而变化。原始社会的教育以口耳相传生存技能为核心, 农业时代的教育依托家庭与师徒制传递生产经验和地方性知识, 工业革命催生大规模教育, 以班级授课制实现高效知识传授。随着社会经济技术的发展, 尤其是以人工智能为核心的颠覆性技术创新, 正以前所未有的速度和规模解构传统职业版图并重塑人力资本需求。未来五年, 全球 23% 的工作岗位将发生变革, 近半数现有工作面临消失或剧烈转型, 同时低技能岗位

[收稿日期] 2025-07-25

[修回日期] 2025-08-25

[DOI编码] 10.13966/j.cnki.kfjyyj.2025.05.004

[基金项目] 科技部“社会治理与智慧社会科技支撑”2021年度“揭榜挂帅”项目“大规模学生跨学段成长跟踪研究”(2021YFC3340800)。

[作者简介] 辛涛, 教授, 博士生导师, 北京师范大学中国基础教育质量监测协同创新中心主任, 安徽师范大学学术副校长, 研究方向: 心理测量、认知诊断与自适应学习理论、教育评价等; 聂竹明, 教授, 博士生导师, 研究方向: 智能学习技术与系统、移动与学习环境建构(zmnie@126.com); 施文静, 中学信息技术教师, 研究方向: 教育数字化、课程教学; 张生, 教授, 博士生导师, 研究方向: 自适应学习、教育大数据等。

[引用信息] 辛涛, 聂竹明, 施文静, 张生(2025). 学评融合: 教育评价的理念转向与落地[J]. 开放教育研究, 31(5): 30-38.

加速被市场淘汰(经济合作与发展组织, 2019), 并催生前所未有的新兴职业。

在此背景下, 全球“未来教育”聚焦促进人的全面发展与回应社会人才需求两大核心目标。联合国教科文组织将教育重新定义为每个人终生拓展可行能力的历程, 并把批判思维、伦理判断与跨文化沟通置于核心, 直接挑战了传统以标准答案和常模排位为核心的评价逻辑。欧盟数字素养框架(DigComp 2.3)将“负责任地使用人工智能”列为公民关键素养。芬兰以立法形式将义务教育延伸至 18 岁并打通职业教育与普通教育的学分互认通道, 构建学习→工作→再学习的路径循环。中国通过教育新基建与智能化全域评价回应同一命题, 强调以数据驱动的差异化资源配置破解“均值陷阱”, 并以核心素养为导向重塑课程与评价。各国战略表面形态各异, 深层逻辑却一致: 未来教育评价必须从选拔工具回归人的主体性建构, 在保障机会公平的同时释放学习者的潜能, 最终培育兼具创新精神与社会责任感的数字时代新主体。

(二) 虚实结合的信息生态环境形成

信息技术正以前所未有的渗透力和重构力, 驱动社会形态发生深刻的本体论转向——从物理世界主导向物理—数字二元世界深度互嵌演进。这一转向的深层逻辑在于, 数据与算法驱动的计算能力已跃升为数字世界的核心生产力, 其本质是信息处理范式的革命性跃迁。教育作为社会系统的有机组成部分, 被置于全新的技术—社会基座之上, 催生出对“教育新基建”的迫切需求, 并引发三重根本性变革: 生活空间变革为实体物理空间与虚拟数字空间的深度融合, 认知交互方式从传统的线性、单向接收转向高度非线性、多源交互的模式, 学习突破时空限制, 形成泛在化学习生态。这些变革要求教育体系重新审视“知识”的本质与“学习”的路径, 将传统物理场所的教学活动延伸、拓展乃至重构至数字场域, 使学习成为跨越虚实界限的连续体验。正如尼古拉斯·尼葛洛庞帝(Negroponte, 1997)预言的, “计算不再只和计算机有关, 它决定我们的生存”。因此, 教育新基建必须通过构建开放共享的数字平台, 使虚实融合从技术概念转化为普惠性教育实践(李海峰等, 2024)。

这一转化过程本身是推动大规模教育覆盖与

个性化精准适配辩证统一的关键枢纽。物理世界涵盖活动、表现、终端与设施等实体要素, 数字世界则是由模型、数据、图像与算法构成的虚拟空间(袁磊等, 2025)。这种耦合通过双重路径实现: 一是对物理世界的数字化映射, 二是对数字世界原生形态的独立建构。伴随着学校课程形态、教学范式、学习方式及教育与心理测量方式的根本变革, 科技赋能重构学评生态体系成为教育的核心任务。通过数据挖掘技术实现知识集成、语义集成与事物集成, 教育得以在虚拟网络世界中探寻个性化学习路径, 建立学习者特征库, 实现对学生发展潜能的精准识别、动态诊断与有效赋能。

(三) 科技赋能评价改革

科技赋能教育评价不仅是工具升级, 更是价值理性的系统重塑, 其核心在于以数据智能破解传统评价的结构性失灵。现有评价体系深陷三重困境: 一是目的异化, 选拔功能凌驾于发展功能, 导致认知资源多被用于应试策略而非能力建构(Black et al., 1998; 张应强等, 2017; 明月等, 2023); 二是方法滞后, 经典测量理论与项目反应理论均以单维静态的方式刻画能力, 难以揭示知识组件的微观结构, 更无法追踪核心素养的生成过程(袁建林等, 2017); 三是伦理失衡, 高利害考试与资源分配刚性挂钩, 数字鸿沟放大结果不平等, 形成“技术—资本”双重马太效应(钟祥铭, 2022; 韦路, 2024; Wayne, 2025)。科技介入的正当性恰恰在于它对上述困境的系统回应。科技赋能教育评价是支撑国家百年大计的核心战略, 其功能定位涵盖导向、诊断、指导与激励四大维度(郑永和等, 2023)。人工智能与大数据技术通过多模态数据实时采集、认知诊断模型动态建模、可解释算法即时反馈, 实现评价目的从甄别选拔转向促进发展、评价主体从教师主导转向学习者参与、评价过程从静态测量转向伴随式诊断的范式跃迁(张生等, 2021)。中共中央、国务院(2020)印发的《深化新时代教育评价改革总体方案》, 提出“四个评价”实施路径: 构建聚焦学习成果的结果评价, 强化关注学习历程的过程评价, 探索衡量进步幅度的增值评价, 健全涵盖多维素养的综合评价。这一改革方向将技术赋能嵌入评价全链条, 使人工智能、大数据等创新工具服务于“五育并举”的教育目标。由此, 科技赋能教育

评价完成从工具理性到价值理性的升维, 为学评融合提供底层逻辑与技术基石。

在教育新生态的背景下, 推进“学评融合”研究既是顺应智能时代教育发展的必然要求, 也是破解传统评价体系结构性困境的重要路径。本研究旨在通过系统构建“以评促学、以学优评”的发展性评价范式, 突破传统甄别式评价的局限, 为素养本位教育评价提供坚实的学理基础和方法论支持; 探索学评融合的落地路径与技术支持体系, 为区域和学校推动评价改革、优化教学过程提供可操作的解决方案, 切实推进“因材施教”和“五育并举”从理念走向实践, 助力实现高水平教育公平与高质量发展目标。

二、学评融合理念转向

教育生态变革催生评价理念的系统性重构, 学评融合由此成为衔接个性化学习与智能化测评的关键枢纽。教育生态的深层变革使评价不再是衡量学习的“后置镜头”, 而成为驱动学习的“内生引擎”。学评融合的提出, 正是对这一历史转向的系统回应: 评价目的从甄别选拔升维为促进发展, 评价主体从教师中心让位于学习者参与, 评价技术从静态测量演变为动态诊断, 评价内容从知识再现转向核心素养生成, 评价范式从“关于学习的评价”跃迁为“作为学习的评价”。这五个维度的同步转向根植于虚实融合的数字生态和“以学习者为中心”的教育哲学, 标志着评价从关于学习的终结性诊断, 进化为作为学习过程的发展性伙伴。

(一) 理念跃迁: 从关于学习的评价到学评融合

教育测量与评价范式正经历从关于学习的评价向以评促学最终迈向学评融合的三阶跃迁(金婉莹等, 2025)。第一代学习评价范式聚焦对知识掌握水平与优秀学生选拔的判断, 以学习结果为核心, 体现工具理性主导的标准化、程序化特征。第二代以评促学范式利用评价反馈优化教学全过程, 强调诊断学习过程以促进改进。第三代学评融合范式将评价视为学习方式纳入教学过程, 通过学生自评与反思发展元认知能力, 呼应在线学习与个性化学习常态化趋势。这一演进源于三重驱动: 社会形态变革与中国特色社会主义发展需求, 国家教育战略对评价改革的顶层要求, 以及以学习者为中心的

教育转型。传统评价范式存在显著局限: 仅依赖考试阻碍学生发展进步, 多样化证据整合困难, 诊断性关注不足导致实践应用障碍。其根本矛盾在于评价工具性与发展性的割裂(Black et al., 1998)。范式演进的内在逻辑是教育评价从选拔工具向成长伙伴的功能性升华, 这种转型根植于社会虚实融合生态的形成与“以学习者为中心”的教育战略, 标志着评价从关于学习的终结性判断, 进化为作为学习过程的发展性伙伴。

(二) 基本逻辑: 学习性与诊断性的统一

学评融合的本质是评价的学习性与诊断性的深度统一。评价的学习性强调评价过程本身即学习的发生过程, 学习者通过参与评价活动实现高阶思维发展, 具体表现为反思与改进的及时性迭代、生成与展示的开放性深层次思维加工(丁邦平, 2008)。评价的诊断性聚焦对学生核心能力的精准剖析, 其核心在于基于过程数据的动态诊断(黄涛等, 2021)。布鲁姆认为“评价即教育过程”。格特·比斯塔等(2025)提出“诊断性评价应服务于人的主体性发展”的主张。二者融合使评价兼具双重功能: 既是能力诊断工具, 更是认知建构活动, 形成以人为核心的自评与互评机制。

技术赋能使学评融合从理念走向实践, 核心在于构建动态诊断与学习干预的闭环系统(张生, 2021)。物联网传感网络实时捕获的课堂音视频流、可穿戴设备收集的生理参数和在线学习行为轨迹, 形成多源异构教育大数据池, 为精准诊断提供实证基础。知识追踪技术作为底层引擎(Abdelrahman et al., 2023), 通过隐马尔可夫过程建模学习者认知状态转移路径。深度知识追踪模型利用长短期记忆网络处理学习序列中的时序依赖关系(刘海鸥等, 2019), 增强学生对复杂知识结构动态演变过程的建模能力。动态键值记忆网络则借助注意力机制实现知识掌握度的可视化映射(Zhang et al., 2017)。这种基于机器学习的诊断能突破传统测评的静态局限, 使知识状态图谱具备实时更新与预测功能, 如预测未来三周知识点遗忘曲线。诊断结果会驱动个性化推荐引擎, 生成适应性学习路径。更重要的是, 该系统通过知识掌握热力图、能力成长仪表盘等可视化工具, 可将诊断数据即时转化为可操作的反思支架——学习者可清晰识别知识盲区并自

主选择补救方案,教师依据群体诊断图谱调整教学节奏。这种“数据采集→状态追踪→智能推荐→反思改进”的闭环,本质上是将评价活动重构为认知元监控过程,使诊断性与学习性在技术赋能下实现辩证统一。学习过程数据与历史数据经由信息采集、管理和存储形成知识图谱,进而支撑诊断报告的图表化生成。

(三)关键指向:核心素养的生成导向

学评融合的核心目标是从基本知识技能评价转向核心素养导向的评价。欧盟率先将核心素养转化为可观察可测量的外显表现,开发了配套指标框架与测量工具,涵盖认知领域的知识技能运用与非认知领域的情感、态度和价值观(European Commission, 2017)。国际测评实践在逐年变革:PISA 评估体系 2003 年新增问题解决评估,2009 年融入数字化素养,2012 年拓展财经素养与创造性问题解决,2015 年纳入协作能力,2022 年首创创造性思维评估,2025 年计划增加外语素养与数字世界学习素养(见表 1)。测评构念呈高阶化转向:写作评估从纠正文通字顺的浅层能力,升级为测量谋篇布局、论证策略与创造性的写作高阶能力。总体来看,未来评价的核心使命是将抽象素养转化为可操作、可验证的发展轨迹(杨宗凯, 2025)。

核心素养导向的评价需突破纸笔测试对高阶能力测量的桎梏,其技术实现依赖于多模态情境建模与跨域能力画像。个性化学习引擎可用于在虚拟仿真场景中部署传感器网络:在科学探究实验中捕捉操作序列与验证假设路径,在社会问题解决任务中记录决策树与协作对话流,形成富含语义的行为日志库。深度学习算法可发挥关键作用:卷积神经网络可用于分析实验操作视频的规范性动作(He et al., 2016),神经网络可用于解码协作讨论的角色互动模式,Transformer 架构能对创造性产出的新颖性与实用性进行语义评分。为保障测量的生态效度,学习者模型持续整合多维度特征——基于 Felder-Silverman 量表(姜强等, 2010)构建认知风格向量,通过知识追踪更新领域掌握度矩阵,结合眼动数据推断元认知策略使用频率。这些特征驱动推荐引擎可执行素养适配策略:为序列型思维者设计阶梯式论证任务,为全局型学习者生成跨学科整合项目,并通过动态难度调节,使其维持在最

近发展区。例如,系统依据保罗·迪恩(Deane, 2011)认知过程模型,分解为立意生成、证据组织、修辞策略等 12 项子能力,利用 BERT 模型(Devlin et al., 2019)对文本进行分层标注,最终生成涵盖逻辑严谨性、文化敏感性和社会洞察力的三维素养雷达图(杨向东, 2022)。这种技术路径使批判性思维、创造性解决问题等抽象素养转化为可量化、可干预的发展轨迹。

(四)突出特征:被评价者的主体参与

学评融合的核心特征在于被评价者参与性的层级跃升:从教师主导的“关于学习的评价”中学生被动接受结果,到“以评促学”中学生参与改进学习过程,最终发展为“作为学习的评价”中以学习者为中心的多方协同参与模式。这种演进直接体现为评价权力的再分配:教师与教育管理机构逐步授权让学习者发挥主体作用,学习者通过发展自我评价与反思能力,从评价对象转变为评价活动的共同设计者。其实施路径包括培养自我监控能力与评他能力:自我监控聚焦内省式思维训练,借助圆圈图、气泡图、双气泡图、树形图等八类思维导图激发深度反思(如双气泡图引导学习者对比目标状态与实际状态的差距,树形图帮助结构化反思认知策略得失);评他能力通过判断选择、标准意识、认知反馈等构建外部诊断框架,达成诊断性与学习性交融的目的。更深层的参与体现在目标协商机制:推荐引擎基于当前能力状态和素养发展目标,通过多臂老虎机算法(Slivkins, A, 2019)生成备选学习路径及预期收益曲线。学习者可在风险偏好框架下自主调整路径参数,如选择激进型挑战还是保守型巩固。这种参与性变革标志着评价从外部测量工具演化为学习者认知发展的内生动力机制,其理论基础源于大卫·海勒(Hyerle, 2009)的思维地图理论,即通过可视化思维结构提升元认知效率。

(五)技术转向:动态诊断的理念支撑

学评融合的技术创新首先体现为评价形式多样化,评价载体从纸笔作答经增强现实技术向自动化、交互式和虚拟化跃迁。终结性单一测评被交互式复杂任务解决能力评价所取代,典型实践包括游戏化创造力测量、基于大脑功能模块的学习能力测量、人机交互科学探究能力测量及虚拟场景审辩式思维测量。这些形式通过知识地图整合核

表 1 国际教育测评素养维度演进

年份	新增素养	测评方法论创新	技术载体	核心文献依据
2003	问题解决能力测评	根据问题构成和学生完成相关任务的绩效划分测评	PISA2003 问题解决能力构成及等级	PISA (2003)
2009	数字化素养	CBT 交互任务设计	计算机模拟系统	PISA 2009 技术报告
2012	财经素养 创造性问题解决	情境化决策模型+开放解构式评分	多线程任务平台	OECD (2013)
2015	协作能力	人机协同行为图谱	CPS 云端协作系统	PISA 技术报告 (2017)
2022	创造性思维	跨学科方案设计指标 VR	原型构建工具	OECD 预发布 (2021)
2025	外语交际素养	数字世界学习	沉浸式语言应用评估	元宇宙情境实验室 (庄雷等, 2024)

心素养与非智力因素,实现心理健康筛查与能力诊断的同步。技术推动评价数据多元化。PISA 测评系统提供两类关键数据:教育系统运行状况与产出,如学生学习心理和生理状况,以及影响因素,如区域经济环境。具体数据源包括记录教师授课过程与学生课堂表现的课堂音视频数据,反映教学模式的动态演变;网络痕迹数据,反映学生学习平台行为轨迹;可穿戴设备数据通过脑电、心、电和体温等生理参数映射学习者注意力、记忆力和认知负荷。为驾驭复杂数据,评价方法融合化成为关键突破点,即依据认知科学理论,融合贝叶斯网络行为过程数据,采用 DINA、NIDA 等认知诊断模型,实现知识状态精细化建模(刘铁园等, 2022)。认知诊断与机器学习方法融合可解决能力追踪问题,与大数据情境融合可优化跨情境适配,与脑科学融合可深化生理机制解析。误差处理依赖匹配标准化、随机化与统计调整技术(杨丽萍等, 2025)。在技术实现层面,学习材料模型能对多模态资源进行特征标识,支撑评价任务精准匹配;教育学规则定义能力状态与评价形式的关联逻辑;知识追踪动态能更新能力状态,驱动推荐引擎按需组合测评形式;多模态数据融合分析可从复杂行为中提取素养证据(Kipf et al., 2017)。

综上,学评融合的实践落地依赖于动态诊断、多维画像、资源结构化与干预闭环的技术协同,将抽象理念转化为可操作的测量、诊断、推荐路径,使评价从静态选拔工具进化为素养生长引擎。

三、学评融合实践的技术引擎

学评融合的实践落地需依托个性化学习技术

的驱动。其引擎价值体现为双重突破:一是构建学习测量的闭环,即通过自适应评测诊断能力,生成个性化学习路径,实现精准干预;二是建立数据赋能的决策机制,以机器学习算法替代经验判断,动态优化资源推送策略。这种技术跃迁根植于认知诊断模型对知识空间的精细化建模,标志着评价从能力标定工具进化为成长赋能系统。

(一)从自适应评测到个性化学习

自适应评测作为个性化学习技术的起点,可根据被测者的知识水平,动态调整与学习者相匹配的评测难度。基于认知诊断的自适应评测更是突破单向能力诊断,通过构建语文、数学、艺术、科学等学科的多维知识空间模型,与学习内容建立关联。个性化学习则通过学习者模型与素材库的双轨架构驱动学评融合。学习者模型记录学习者的个体信息,是学习者特征的集合(危怡等, 2018)。核心引擎通过匹配教育学规则与学习者经验知识实现教学背景适配,依据视觉偏好等个体特征完成材料精准匹配。具体而言,学习系统可根据学习者个性特征呈现信息与资料,通过反馈建议构建智能学习环境,使评价嵌入学习全链条:学习环节关联学习者与学习材料,评价环节通过测试作业估计潜在属性,系统据此自适应决策,提供相应的后续学习材料,直至达到理想掌握水平。机器学习算法在此过程中替代教师制定学习策略,结合学习者测评表现建立个性化学习推荐系统。其技术跃迁体现为两点:一是学习进阶路径的显性化,即将知识状态演进轨迹转化为可量化指标;二是学习者个体画像的精准刻画,即基于历史数据与实时反馈生成定制化学习材料序列(Lord, 2021; 陈平等, 2011),形成“测量—决策—干预”闭环。

(二) 指向素养发展的学习者建模

自适应学习系统基于学习者模型, 对真实学习者构建抽象认识(顾小清等, 2024)。学习者模型的关键是特征选取和对特征的建模(王蒙蒙等, 2020)。其运行机制包含三阶设计: 组织设计规划内容结构、界面呈现与场景情境; 特征标识定义数据流、数字媒体与活动类型等属性; 路径设计与材料推荐基于预定学习路径或最优策略生成学习序列。学习者特征的完整性和准确性是系统推荐恰当学习材料的基础(见图 1)。学习者建模通过宏观→中观→微观三层架构实现素养发展追踪。宏观层挖掘不同地域年龄群体的特征规律, 构建学习规律模型与发展轨迹框架。国家 24365 平台基于学信网核验毕业生能力数据, 通过区块链存储实习成果、技能认证等非正式学习记录, 构建了从学业能力到职业素养的可信发展轨迹, 为宏观教育资源配置提供依据(田兆富, 2011)。中观层聚焦个性化学习者特征的动态演化机制, 解析心理、生理、认知等多维属性的交互关系。微观层整合学习者特征库, 系统归集心理品质、生理指标、认知能力和社会性发展数据。这种分层建模的核心价值在于将抽象素养转化为可量化的发展轨迹, 通过历史数据与实时反馈的迭代分析, 生成适应个体成长节奏的学习路径。

(三) 融合计算机技术的测量建模

测量建模的创新聚焦三类技术路径: 个人拟合统计量通过构建学习警示指数诊断异常学习模式, 如粗心、幸运猜测或特殊知识结构, 优化学习过程(Yu et al., 2024); 学习者模式诊断在认知诊断模型框架中融合机器学习算法, 建模作弊、睡眠模式与随机猜测等行为(朱佳等, 2022); 基于测量的分类以项目反应理论测量结果指标, 对学习过程开展聚类分析。具体而言, 个人拟合统计量识别偏离常规反应模式的学习者, 如通过反应向量分析区分正常作答与异常状态(心不在焉或作弊倾向)。学习者模式诊断结合认知属性掌握概率与行为特征, 识别系统性失误的认知根源。基于测量的分类则依据能力值变化轨迹将学习者分为不同发展类型, 分类一为全掌握状态但存在异常睡眠模式, 分类二为无异常, 分类三为未掌握且存在作弊行为。例如, 猿编程 AI Agent 工业级应用的抄袭检测模块采用代码相似度算法, 同步分析操作轨迹时间戳异常, 实时阻断作弊并推送网络安全教育动画, 实现行为诊断与教育干预的同步闭环(贾积有等, 2023)。

(四) 基于个性化推荐的技术闭环

学评融合的技术闭环以个性化学习的推荐算法为核心引擎, 涵盖五类算法: 学习理论和心理测

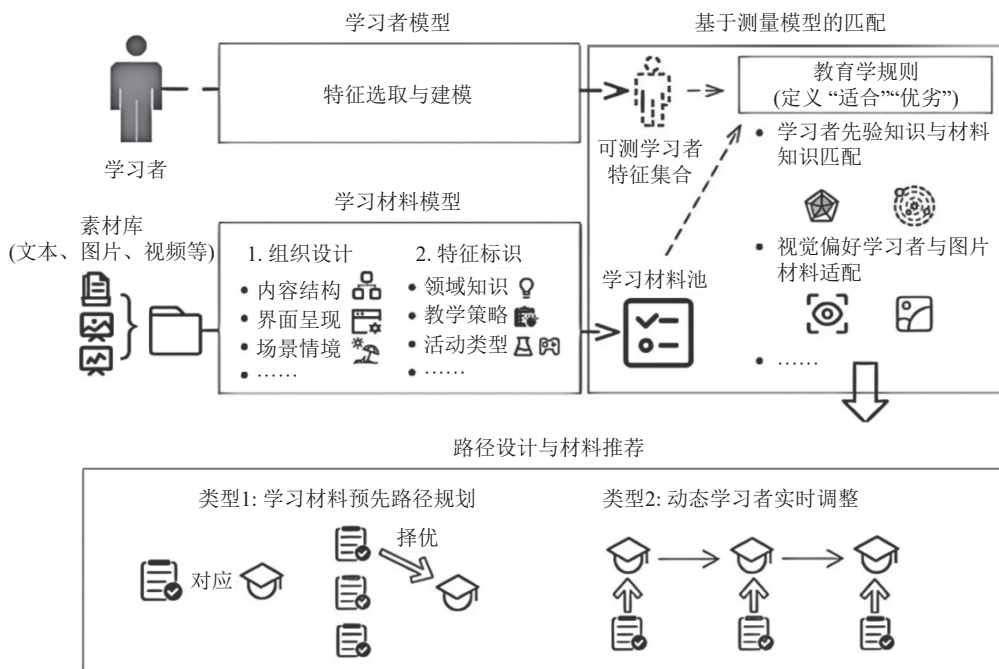


图 1 以个性化学习为引擎的学评融合设计

量理论、基于状态预测或知识追踪算法、基于距离计算的机器学习分类算法、基于奖励反馈的算法(强化学习)与基于损失函数的算法(蚁群算法、粒子群算法和遗传算法)(罗自强, 2019)。这一闭环通过三重映射实现系统集成: 学习者模型将真实学习者映射为潜变量表征, 材料模型将实际学习材料映射为标签化特征, 数据模型通过建模规则连接潜变量与特征标签。其运行逻辑体现为测量环节量化学习者潜在特质, 推荐环节依据特质匹配学习材料, 形成“测量—建模—推荐”的迭代循环(见图2)。技术闭环的关键价值在于以基于测量的个性化推荐作为人工智能与教育融合的核心发力点, 通过机器学习算法替代教师决策学习策略, 持续优化学习路径直至潜在属性达理想掌握水平。

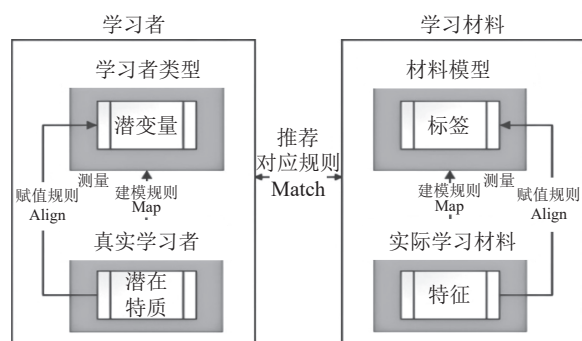


图2 基于测量的个性化推荐

四、总结与反思

学评融合需辩证调适技术理性与教育价值, 突破三重矛盾: 测量有效性需平衡算法复杂性与效度可解释性, 技术伦理性需协调数据价值与隐私安全, 资源普惠性需破解教育公平困境。

(一) 信度与效度

如何评价人工智能时代教育评价的信度与效度? 核心在于建立三阶段演进的效度框架: 由二十世纪中叶前基于标准与内容的效度主导, 转向基于结构的效度论证框架, 未来需融合效度与可解释性解决算法黑箱问题。美国教育考试服务中心调查显示, 仅66%受访者信任人工智能评分的学习评估, 72%信任人工智能技能改进建议(TeachAI Consortium, 2023)。这种信任赤字表明, 效度验证滞后于技术发展, 亟需构建可解释的、既满足测量严谨性, 又赢得用户信任的人工智能效度新框架。

(二) 伦理与安全

大规模在线学习环境下教育与心理测量面临严峻的数据安全挑战。自2020年起, 日常教学与高风险选拔考试广泛采用远程在线方式, 对数据安全及远程监考提出更高要求。应对策略需聚焦两重路径: 一是构建行为区块链, 记录学习过程数据, 在完全不透露隐私信息的同时维持信息一致性, 为精准化学习评价提供可信支撑; 二是利用基于区块链的学生评测数据安全技术, 确保参与计算的各方无需向验证者提供敏感信息, 即可证明论断正确性, 避免第三方窃取或窥视数据, 从根本上保护教育数据的安全与隐私。

(三) 教育公平性

在线学习模式下教育公平的核心矛盾在于信息技术赋能与数字鸿沟风险并存。教育的基本价值是保障所有学生的学习权利, 教育信息化的发展本应助力弥合资源分配不均(Butucescu et al., 2020)。然而, 网络资源连通性差异实际上导致数字鸿沟持续存在, 发达地区可能享受更高质量的智能化教育服务, 从而扩大教育资源失衡。解决路径包括: 一是评估学生心理状态、社会资源支持度及需克服的客观困难; 二是加强偏远地区教育资源投入, 重点保障教育信息技术薄弱地区的公平权益; 三是通过政策干预防止技术资源配置引发教育马太效应。科技赋能教育评价的最终目标在于实现大规模个性化教育, 使教育精准服务科教兴国战略与现代化建设人才支撑目标。

[参考文献]

- [1] Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. [J]. ACM Computing Surveys; 1-37.
- [2] Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment[J]. Phi Delta Kappan, 80(2): 139-148.
- [3] Butucescu, A., & Iliescu, D. (2020). Fairness perceptions in educational assessment: The role of positive and negative affect[J]. Educational Studies, 48(3): 1-13.
- [4] 陈平, 辛涛 (2011). 认知诊断计算机化自适应测验中的项目增补 [J]. 心理学报, 43(7): 836-850.
- [5] Deane, P. (2011). Writing assessment and cognition (Research Report ETS RR-11-14)[R]. Princeton: Educational Testing Service: 8-11.
- [6] Devlin, J., Chang, M. W., Lee, K., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American

Chapter of the Association for Computational Linguistics. Minneapolis: Association for Computational Linguistics; 4171-4186.

[7] 丁邦平(2008). 从“形成性评价”到“学习性评价”: 课堂评价理论与实践的新发展[J]. 课程. 教材. 教法, (9): 20-25.

[8] European Commission(2017). DigCompEdu: Digital competence framework for educators[R]. Luxembourg: Publications Office of the European Union; 12-15.

[9] 格特·比斯塔, 赵康(2025). 测量时代和人工智能时代的教育——基于主体化教育展开的对话[J]. 教育研究, 46(5): 55-67.

[10] 顾小清, 刘桐(2024). 大模型时代的智适应学习研究: 进展、实例与展望[J]. 中国教育信息化, 30(5): 55-66.

[11] He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J. (2016). Deep residual learning for image recognition[A]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition[C]. Las Vegas: IEEE Computer Society; 770-778.

[12] 黄涛, 赵媛, 耿晶, 等(2021). 数据驱动的精准化学习评价机制与方法[J]. 现代远程教育研究, 33(1): 3-12.

[13] Hyerle, D. (2009). Visual tools for transforming information into knowledge[M]. Thousand Oaks: Corwin Press; 57.

[14] 贾积有, 乐惠晓, 张誉月等(2023). 基于大数据挖掘的智能评测和辅导系统设计[J]. 中国电化教育, (3): 112-119.

[15] 姜强, 赵蔚, 杜欣(2010). 基于 Felder-Silverman 量表用户学习风格模型的修正研究[J]. 现代远距离教育, (1): 62-66.

[16] 金婉莹, 庄榕霞(2025). 从“关于学习的评价”到“作为学习的评价”——元评价视域下智能技术支撑学习评价的有效性探索[J]. 中国远程教育, 45(5): 92-107.

[17] 经济合作与发展组织(2019). 2019年就业展望报告[R]. 巴黎: OECD Publishing; 35-38.

[18] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks[A]. Proceedings of the 5th International Conference on Learning Representations[C]. Toulon: ICLR; 5-7.

[19] 李海峰, 王炜(2024). 生成式人工智能赋能教育新基建智慧数字底座的样态[J]. 远程教育杂志, 42(6): 24-32.

[20] Lord, F. M. (1980). Applications of Item Response Theory to practical testing problems[M]. Hillsdale: Erlbaum; 97.

[21] 刘海鸥, 刘旭, 姚苏梅, 等(2019). 基于大数据深度画像的个性化学习精准服务研究[J]. 图书馆学研究, (15): 68-74.

[22] 刘铁园, 陈威, 常亮, 等(2022). 基于深度学习的知识追踪研究进展[J]. 计算机研究与发展, 59(1): 81-104.

[23] 罗自强(2019). 云进化计算[M]. 北京: 电子工业出版社; 29-32.

[24] 明月, 聂雅怡, 刘翔鹰, 等(2023). “双减”背景下学生认知负荷的表征、成因及优化[J]. 教学与管理, (12): 63-68.

[25] Negroponte, N. (1997). 数字化生存[M]. 胡泳译. 海口: 海南出版社; 12.

[26] OECD(2003). The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills[R]. Paris: OECD Publishing; 121-158.

[27] OECD(2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy[R]. Paris: OECD Publishing; 137-157.

[28] OECD(2021). AI and the future of skills, Volume 1: Capabilities and assessments[R]. Paris: OECD Publishing; 89-93.

[29] Slivkins, A. (2019). Introduction to multi-armed bandits[J]. Foundations and Trends in Machine Learning; 45-49.

[30] TeachAI(2023). Policy framework for AI in education[R]. Washington, D. C.: TeachAI Consortium; 36.

[31] 田兆富(2011). 基于信息服务网络平台的大学生就业指导体系建设[J]. 教育与职业, (35): 93-94.

[32] Wayne, Au. (2025). 天生不平等: 美国的应试教育与社会分化[M]. 张凌, 译. 北京: 中国科学技术出版社; 13-123.

[33] 王蒙蒙, 徐久成(2020). 基于在线学习行为数据的学习者群体特征挖掘研究[J]. 软件导刊, 19(7): 153-157.

[34] 危怡, 胡梦华, 胡艺龄, 等(2018). 开放学习者模型: 让学习者参与构建——访国际知名教育人工智能专家朱迪·凯教授[J]. 开放教育研究, 24(3): 4-11.

[35] 韦路(2024). 数字鸿沟: 概念、成因与后果[M]. 杭州: 浙江大学出版社; 27.

[36] Yu X. F., Tang Q., Qin C. Y., Li Y. J.(2024). Person-Fit in cognitive diagnostic assessment[J]. Journal of Psychological Science, 47(3): 744-751.

[37] 杨丽萍, 辛涛, 李峰, 等(2025). 学生评价的数字化整合: 内涵特征与关键问题[J]. 中国远程教育, (4): 114-128.

[38] 杨向东(2022). 作为理论构念的素养及其模型构建[J]. 华东师范大学学报(教育科学版), 40(11): 41-57.

[39] 杨宗凯(2025). 智能时代高校学生能力评价及其实践[J]. 中国大学生就业, (6): 13-19.

[40] 袁建林, 刘红云(2017). 核心素养测量: 理论依据与实践指向[J]. 教育研究, 38(7): 21-28+36.

[41] 袁磊, 梁世松, 吴瑶, 等(2025). 高校学生的数智素养测评及培养策略研究[J]. 现代远距离教育, (3): 14-24.

[42] 中共中央, 国务院(2020). 深化新时代教育评价改革总体方案[N]. 人民日报, 2020-10-14(1).

[43] Zhang, J., Liu, Q., Huang, Z., Wu, S., Wang, L., Gao, M., Guo, Y., Liu, H., & Chen, E. (2017). Dynamic key-value memory networks for knowledge tracing[J]. Perth: International World Wide Web Conference Committee; 768-770.

[44] 张生, 王雪, 齐媛, 等(2021). 人工智能赋能教育评价: “学评融合”新理念及核心要素[J]. 中国远程教育, (2): 1-8+16+76.

[45] 张应强, 黄捷扬(2017). 培养大学生核心素养与深化高等教育评价改革[J]. 厦门大学学报(哲学社会科学版), (6): 62-71.

[46] 郑永和, 王一岩(2023). 科技赋能教育高质量发展: 价值内涵、表征样态与推进策略[J]. 中国电化教育, (1): 118-126.

[47] 钟祥铭, 方兴东(2022). 数字鸿沟演进历程与智能鸿沟的兴起——基于50年来互联网驱动人类社会信息传播机制变革与演进的视角[J]. 新闻记者, (8): 34-46.

[48] 朱佳, 黄昌勤(2022). 智慧教育技术与应用[M]. 北京: 清华大学出版社; 138-163.

[49] 庄雷(2024). 元宇宙视域下高校课程智慧化建设路径及案例研究[J]. 中国教育信息化, 30(10): 100-108.

(编辑: 魏志慧)

Transition to Assessment-Learning Integration in Educational Assessment and Technological Engine Construction

XIN Tao^{1,2}, NIE Zhuming², SHI Wenjing³ & ZHANG Sheng¹

(1. Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing 100875; 2. College of Education Science, Anhui Normal University, Wuhu 241000, China; 3. The Second Middle School of Kunshan, Kunshan 215300, China)

Abstract: *The transformative changes in the educational ecosystem bring forth a paradigm shift in the assessment from assessing to select to assessing to foster learning. The integration of learning and assessment has become an inevitable choice in the development of personalized education and technological empowerment. The shift from knowledge and skill-based evaluation to competency-oriented evaluation breaks from the limitations of paper-based tests through multi-modal modeling to convert abstract competencies into quantifiable trajectories. The technological pathway driving its implementation is the personalized learning engine, which constructs a closed-loop system with its assessment data, integrating multi-source heterogeneous data, its diagnostic applications using the cognitive diagnostic models to dynamically analyze knowledge statuses and predict trends, and its recommendations generating personalized learning paths based on algorithms. With a cycle of “data collection→state tracking→intelligent recommendation→reflection and improvement,” this system reconfigures evaluation as a cognitive meta-monitoring process, evolving it into a competency growth engine. There are three major challenges to implementing this evaluation system. First, the balance of algorithm complexity and interpretability is needed to build an interpretable AI framework to enhance its trust. Second, the coordination of data value and privacy security is essential to address the risks of large-scale remote assessments and explore privacy protection technologies. Finally, it is necessary to address the fairness challenges arising from the digital divide, prevent imbalances in the allocation of technical resources, and ensure that technological empowerment achieves both personalized assessment and promotes equal educational opportunities through compensation mechanisms.*

Key words: *Assessment-learning integration; Competency-oriented evaluation; Educational equity; Personalized recommendation; Assessment ethics*