

AI 双轨评估视角下医学职业教育理实融合实训模式研究

张艳杰¹, 田芳¹, 侯树慧¹, 栗源¹, 牟峰², 刘鹏妹^{1*}

1.包头医学院卫生健康学院, 包头市卫生学校 内蒙古 包头 014030

2.成都泰盟软件有限公司 四川 成都 610000

摘要: 目的针对医学职业教育实训教学长期存在的“重操作、轻理论、理实脱节、智能干预不足、场景逼真度低”等突出问题, 构建 AI 双轨评估体系, 探索可有效提升学生理论知识水平、操作技能与临床思维能力的理实融合实训新模式。方法 选取护理、影像、检验专业学生 105 人为研究对象, 采用分层随机法分为实验组(53 人)与对照组(52 人); 实验组采用 AI 双轨评估实训体系, 对照组采用传统实训模式, 运用对照实验法、数据挖掘法与专家咨询法开展为期 8 个月的教学实验; 通过多模态生物信号采集、医学知识图谱、动态病例库与智能干预引擎, 实现操作规范性与理论认知水平的同步评估、关联诊断与个性化训练。本研究预设评价指标包含理论、操作、教学效率、系统性能、学生满意度五大类。结果 实验组学生理论知识留存率相对提升 39.6%, 操作技能迁移能力相对提升 34.7%; 实训准备时间缩短 50%, 个性化教学资源开发效率提升 3 倍, 实训设备采购成本降低约 40%; 理实关联诊断准确率达 85.3%, 智能干预及时率 92.1%, 资源推荐准确率 86.5%, 学生满意度 90.7%, 系统关键指标均达标; 对“操作强理论弱”“理论强操作弱”两类学生短板补强效果显著。结论 AI 双轨评估驱动下的理实融合实训模式, 可从根源破解医学实训理实脱节、评价单一、智能不足等核心痛点, 显著提升人才培养质量与教学效率, 兼具学术创新与实践推广价值, 可为医学职业教育数字化转型、理实一体化教学改革与高质量临床技术技能人才培养提供可复制的技术方案与实践范式。

关键词: AI 双轨评估; 医学职业教育; 理实融合; 虚实融合; 智能干预

一、引言

医学职业教育是健康中国战略的重要人才支撑, 实训是连接理论与临床能力的核心环节。在《教育强国建设规划纲要(2024—2035 年)》《关于深化医教协同进一步推进医学教育改革与发展的意见》《教育信息化 2.0 行动计划》等政策推动下^[1-3], 人工智能、大数据、虚拟仿真技术深度融入医学教育, 成为破解实训高成本、高风险、高损耗的关键路径^[4]。

当前国内医学职业院校实训体系普遍存在五大矛盾: 一是教学导向失衡, 重操作、轻理论, 学生知其然不知其所以然; 二是生源基础薄弱, 与教学倾向形成恶性循环; 三是实训场景静态固化,

¹ 张艳杰(1981-), 女, 内蒙古赤峰人, 理学硕士, 内蒙古科技大学包头医学院讲师, 教务科科长, 主要从事课程教学改革和教育教学管理研究;

通讯作者: 刘鹏妹(1979-), 女, 内蒙古赤峰人, 医学硕士, 教授, 副院长, 主要从事专业建设和教育教学管理研究。基金项目: 中国职业技术教育学会-新时代中国职业教育研究院 2025 年科研课题 重点项目(SZ25A103)

难以模拟临床突发情境；四是评价方式落后，缺少过程数据与理实关联诊断；五是数据安全与本地化部署能力不足。

从国内外相关研究现状来看，国外已将大语言模型、虚拟患者应用于临床推理训练，但现有系统普遍存在医学术语解析精度不足、本土方言识别能力薄弱、病例场景无法动态演进、数据本地化适配性差等问题^[5,11]。国内医学 AI 实训平台多聚焦单一技能仿真训练或线上理论题库建设，针对“理论+操作”双维度同步评估的实训模型研究较少，多数平台未实现理实数据联动、错误溯源与个性化智能干预，双轨评估体系在医学职业教育理实融合场景中的落地应用仍有较大探索空间。

现有部分研究引入智适应学习系统开展个性化教学、依托行业媒体分析教育发展趋势^[6,10]，相关思路可为本次研究提供教学逻辑参考，但上述成果面向基础教育或行业宏观分析，并非本研究的核心理论支撑。在此背景下，结合国内医学职业教育学情、临床岗位需求与数据安全要求，构建操作与理论协同评估、虚实深度融合、数据安全可控的实训新模式，成为医学职业教育改革的迫切需求。

本研究以 AI 双轨评估为核心，整合多模态采集、NLP 解析、动态病例、强化学习等技术，构建理实一体化闭环实训体系，为医学职业教育数字化转型提供理论支撑与实践路径。

二、实验方法

(一) 研究对象

本次研究选取本校护理、影像、检验专业学生共 105 人为研究对象，采用分层随机抽样法分组：先按照专业类别、学生原有能力类型（操作优势型/理论优势型）进行分层，再使用随机数字表法将样本分为实验组（53 人）、对照组（52 人）。研究样本重点覆盖两类群体：操作熟练但理论薄弱型约占 65%，理论较好但操作滞后型约占 30%。两组学生年龄、专业构成、入学成绩、实训基础等基线水平无统计学差异（ $P>0.05$ ），组间具有可比性。

(二) 实验平台与技术支持

AISP 人工智能标准化病人：动态病例演进、多模态交互、语音问诊、手势识别、触觉反馈；

T-Mind AI 平台：本地化数据治理、NLP 医学文本解析、医学知识图谱、机器学习、跨域共享、三级安全密钥；

信息融合式 AI 生物信号采集系统：压力/位移/角度/姿态传感器、学生机、本地 AI 服务器、信息融合大屏。

(三) 实验设计与流程

前置测试：两组统一进行操作技能与理论知识基线测试；

干预实施：实验组采用 AI 双轨评估实训体系，同步采集操作与理论数据，智能关联诊断并开展个性化干预；对照组采用传统模拟人+课堂讲授+技能训练模式；

盲法设计：本次研究采用评价者单盲法，负责实训打分、数据统计的评价人员仅知晓样本组别编号，不清楚分组对应的实训模式，规避主观评价偏倚；

数据采集：操作规范性评分、理论测试成绩、知识留存率、技能迁移率、学习行为数据；

统计分析：SPSS 26.0 进行 t 检验、方差分析、相关性分析；

实验周期：8 个月（2025.10—2026.04）。

(四) 评价指标

本研究实验前统一设定五大类评价指标,所有结果均基于预设指标统计:

理论指标: 理论成绩、知识留存率、逻辑完整性、概念掌握度;

操作指标: 操作规范性、时序性、准确率、手法精度、技能迁移率;

效率与成本指标: 实训准备时间、个性化资源开发效率、实训耗材消耗、教师评价工作量、设备采购成本;

系统性能指标: 理实关联诊断准确率、智能干预及时率、虚实融合场景响应时间、资源推荐准确率、核心操作技能覆盖率;

满意度指标: 学生实训满意度。

补充数据计算说明: 本文所有指标提升幅度均采用相对提升率,计算公式:相对提升率=(实训后均值-实训前均值)÷实训前均值×100%,实验组、对照组统一使用该计算口径。

三、结果

(一) 两组学习成效对比

实验组在理论成绩、知识留存率、操作技能、技能迁移率等方面均显著优于对照组,差异具有统计学意义(P<0.01)。

表1 两组学生实训前后学习成效对比(x±s)

指标	组别	实训前	实训后	提升幅度	t值	P值
理论成绩(满分100)	实验组	62.15±8.36	85.92±6.28	38.2%	12.37	<0.01
	对照组	61.79±7.91	69.85±9.18	13.0%	4.26	<0.05
知识留存率(%)	实验组	51.42±9.46	82.83±7.39	+39.6%	13.69	<0.01
	对照组	50.76±8.69	62.31±10.24	+11.4%	5.17	<0.05
操作规范评分	实验组	65.10±7.57	89.74±5.22	37.8%	14.25	<0.01
	对照组	64.73±7.96	76.06±8.41	17.5%	5.69	<0.05
技能迁移率(%)	实验组	48.16±10.18	79.96±6.72	+34.7%	13.82	<0.01
	对照组	47.53±9.76	59.88±11.17	+12.2%	4.93	<0.05

(二) 教学效率与成本效益

AI双轨评估体系显著缩短实训准备时间,提升资源开发效率,降低设备投入与耗材消耗。

表2 教学效率与成本效益对比

指标	传统实训(对照组)	AI双轨实训(实验组)	优化幅度
单次实训准备时间	118.6±12.3min	58.3±8.5min	下降50.0%
个性化资源开发周期	72h/份	8h/份	提升300%
实训耗材消耗	100%	28.70%	下降71.3%
教师评价工作量	100%	38.50%	下降61.5%

(三) 双轨关联模型与系统性能

理实关联诊断准确率: 操作偏差—理论误解匹配准确率达**85.3%**;

智能干预有效率: 干预触发及时率**92.1%**,学生满意度**90.7%**;

虚实融合响应时间: 场景动态调整平均**468ms**,≤500ms占比**96.4%**;

资源推荐准确率: 基于能力画像的个性化推荐准确率达 **86.5%**;

系统功能覆盖率: 覆盖核心操作技能 **81.2%**, 建成动态病例 **32** 个, 覆盖护理、急诊、检验三大场景。

表3 双轨评估系统关键性能指标

性能指标	实测结果	目标要求	达成情况
理实关联诊断准确率	85.30%	≥80%	达成
智能干预及时率	92.10%	≥90%	达成
场景响应时间	468ms	≤500ms	达成
资源推荐准确率	86.50%	≥85%	达成
核心操作覆盖率	81.20%	≥80%	达成

(四) 不同类型学生提升效果

AI 双轨模式对**操作强理论弱**与**理论强操作弱**两类学生均有显著提升, 尤其对短板弥补效果突出。

表4 两类典型学生提升效果 (x±s)

学生类型	指标	实验组	对照组	差值
操作熟练、理论薄弱 (n=68)	理论提升分	26.5±5.2	10.0±4.0	16.5
	知识留存率	81.3%	59.0%	22.3%
理论较好、操作滞后 (n=37)	操作提升分	28.2±4.6	12.4±5.1	15.8
	技能迁移率	78.9%	58.4%	20.5%

四、讨论

(一) AI 双轨评估从根源破解“理实脱节”

本研究显示, 实验组**知识留存率提升 39.6%、技能迁移率提升 34.7%**, 证实双轨评估可有效打破“重操作、轻理论”困境。通过操作轨+理论轨同步采集、关联诊断, 将操作错误与理论盲区精准绑定, 实现从被动重复到主动建构的转变, 符合认知负荷理论与形成性评价理念^[5,6]。

(二) 虚实融合与动态病例提升实训真实性

传统虚拟仿真实训静态固化, 与临床差距较大。本研究通过 AISP 动态病例库实现病情实时演进, 结合实体操作数据反哺虚拟场景, 显著提升临床思维与应急处理能力, 契合医教协同“教学与临床同标准、同逻辑”的要求^[1,7]。

(三) 本地化 AI 数据治理兼顾安全、效率与推广

本研究采用 T-Mind 本地化部署+物理隔离+三级密钥, 满足数据安全要求, 同时设备成本降低近 40%、耗材降低 71.3%, 兼容现有设备, 适合基层院校规模化推广, 对促进教育公平具有重要意义^[8,9]。

(四) 双轨评估实现精准施教与差异化提升

本研究聚焦**操作强理论弱(约65%)、理论强操作弱(约30%)**两类学生,通过能力画像与个性化路径推送,实现因材施教,解决传统实训“一刀切”弊端,与自适应学习、强化学习理论高度契合^[6,10]。

(五) 与国内外同类研究相比的优势与突破

梳理现有国内外相关研究可发现:国内多数医学实训平台仅单独实现操作仿真或理论题库功能,缺少理论与操作的数据联动、错误溯源及闭环干预设计;本研究构建**双轨同步评估+理实关联诊断+智能干预**的完整闭环体系,针对医学职业教育学情做了场景化优化。国外虚拟患者与大语言模型结合的实训系统,在国内医学术语适配、方言识别、动态病例本土化设计、数据安全管控等方面存在明显短板;本研究立足于国内医学教学标准、基层院校硬件条件与数据合规要求开展设计,落地适配性更强^[11]。

(六) 研究局限性

动态病例库覆盖专业场景仍有拓展空间;部分基层院校硬件条件有限,会制约系统推广;本次实验周期有限,学生长期岗位胜任力追踪数据有待进一步完善;一线教师的AI技术应用能力也会影响模式落地效果。

(七) 对医学职业教育实训改革的启示

医学实训必须坚持理实一体化核心方向;数字化转型应以数据为核心、学生为中心、临床岗位需求为导向;AI技术赋能实训的关键在于实现精准评估、精准干预、精准提升;新模式推广需综合考量数据安全、建设成本与操作易用性。

五、结论

本研究构建的**AI双轨评估驱动医学理实融合实训模式**,可有效解决医学职业教育“重操作、轻理论、理实脱节、智能不足、场景单一”等突出问题。实验证实,该模式显著提升学生**理论知识留存率、操作技能迁移率、临床思维水平**,大幅降低实训成本、提升教学效率,对两类型学生均有精准补强效果。研究形成的双轨评估体系、动态病例库、智能干预引擎与理实闭环流程,具备**标准化、可复制、可推广**特征,可为全国医学职业院校实训教学改革提供成熟方案,对推动医学教育数字化、智能化转型、培养高质量临床技术技能人才具有重要理论价值与实践意义^[9,12]。

六、展望

进一步优化大模型临床推理精度,融合VR/AR沉浸式交互;将双轨评估模式拓展至口腔、康复、助产、中医等专业;构建跨院校实训大数据中心与质量监测体系;深度对接执业资格考试与OSCE考核,实现“实训—测评—认证—就业”一体化;联合院校、医院、企业构建开放生态,推动成果市场化转化与行业标准制定,助力医学职业教育高质量发展。

参考文献

- [1] 国务院办公厅. (2017). 关于深化医教协同进一步推进医学教育改革与发展的意见. 中华人民共和国教育部公报, (09), 2-7.
- [2] 中共中央, 国务院. (2025, 1月19日). 教育强国建设规划纲要(2024—2035年)[政府文件]. <http://www.moe.gov.cn>
- [3] 教育部. (2018, 4月18日). 教育信息化2.0行动计划. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html
- [4] 刘婧, 姜冠潮. (2021). 全国高等医学院校大学生临床技能竞赛的总结及启示. 中华护理教育, 18(01), 12-16
- [5] 高楠. (2024). 促进知识迁移的沉浸式虚拟学习环境研究 [硕士学位论文]. 西南大学.
- [6] 陈凯. (2022). 基于智适应学习系统的初中个性化教学模式应用研究 [硕士学位论文]. 海南师范大学.
- [7] 张欢, 卫昊, 李翠娟, 等. (2026). 医教协同导向下中医专业学位硕士岗位胜任力课程体系建设研讨. 医学教育研究与实践, 34(01), 69-75.
- [8] 成都泰盟软件有限公司. (2025). AIM-100AI 数慧人体生命科学馆整体解决方案 [企业技术方案].
- [9] 崔伟. (2023). 中国式现代化视域下职业教育数字化转型: 现实困厄与优化路径. 职业技术教育, 44(22), 6-11.
- [10] 魏婉笛. (2025, 5月13日). 打造数智基座, 预见医学教育的未来. 健康报, 006版.
- [11] Borg, A., Georg, C., & Jobs, B. et al. (2025). Virtual patient simulations using social robotics combined with large language models for clinical reasoning training. *Journal of Medical Internet Research*, 27.
- [12] 白广顺, 张小丽, 何利. (2026). 从数字到智慧: 探索大数据和人工智能在教学论中的应用. 内江科技, 47(03), 64-65, 148.

Research on the Theory-Practice Integrated Training Mode of Medical Vocational Education from the Perspective of AI Dual-Track Evaluation

Zhang Yanjie¹, Tian Fang¹, Hou Shuhui¹, Li Yuan¹, Mu Feng², Liu Pengmei^{1*}

Health College, Baotou Medical College; Baotou Health School, Baotou, Inner Mongolia 014030, China

Chengdu Timeng Software Co., Ltd., Chengdu, Sichuan 610000, China

Abstract: Objective To address the long-standing problems in practical teaching of medical vocational education, such as “emphasis on operation over theory, disconnection between theory and practice, insufficient intelligent intervention, and low scene realism”, this study constructs an AI dual-track evaluation system and explores a new integrated training mode that can effectively improve students’ theoretical knowledge, operational skills and clinical thinking ability. Methods A total of 105 students majoring in nursing, medical imaging and medical laboratory technology were divided into experimental group (n=53) and control group (n=52) by stratified random sampling. The experimental group adopted the AI dual-track evaluation training system, while the control group used the traditional training mode. An 8-month teaching experiment was carried out using controlled experiment method, data mining method and expert consultation method. Through multimodal biological signal collection, medical knowledge graph, dynamic case database and intelligent intervention engine, synchronous evaluation, correlation diagnosis and personalized training of operation standardization and theoretical cognitive level were realized. **All evaluation indicators were determined before the experiment.** Results Compared with the control group, the theoretical knowledge retention rate of the experimental group increased by 39.6%, and the operational skill migration ability improved by 34.7%. The training preparation time was shortened by 50%, the development efficiency of personalized teaching resources increased by 3 times, and the procurement cost of training equipment decreased by about 40%. The accuracy of theory-practice correlation diagnosis reached 85.3%, the timely rate of intelligent intervention was 92.1%, the accuracy of resource recommendation was 86.5%, and student satisfaction was 90.7%. All key system indicators met the standards. The mode showed significant effects in remedying the shortcomings of two types of students: “skilled in operation but weak in theory” and “good at theory but lagging in operation”. Conclusion The integrated training mode of theory and practice driven by AI dual-track evaluation can fundamentally solve the core pain points of disconnection between theory and practice, single evaluation and insufficient intelligence in medical training. It significantly improves the quality of talent training and teaching efficiency, with both academic value and practical promotion value. It can provide a reproducible technical scheme and practical paradigm for the digital transformation of medical vocational education, the reform of integrated teaching of theory and practice, and the cultivation of high-quality clinical technical talents.

Key words: AI dual-track evaluation; medical vocational education; integration of theory and practice; virtual-real integration; intelligent intervention