·胜任力评价专题·

医学教育程序性评价的国际多案例比较研究

王文涛[1]　鲁子越[1]　潘俊颐[1]　沈健[2]　李响[2,3]
[1]重庆医科大学莱斯特大学联合学院，重庆 400016；[2]重庆医科大学附属第一医院心血管内科，重庆 400016；[3]重庆医科大学教务处，重庆 400016
通信作者：李响，Email：lzd9288@163.com

【摘要】 目的 通过跨国多案例比较，分析医学教育程序性评价的实施取向、评估工具、反馈支持、进展评议和信息系统支撑，为我国医学教育评价改革提供参考。**方法** 检索PubMed，并结合Web of Science和Scopus补充核查，检索截至2025年11月10日的相关文献。纳入明确报告实施程序性评价或具备其核心特征的医学教育实证研究，经项目归并后，依据项目成熟度、文献支撑度、代表性、情境差异和资料完整性遴选典型案例，并围绕实施取向、证据来源、反馈支持、进展评议和信息系统进行比较。**结果** 共检索文献3 863篇，最终纳入55篇实证研究，识别23个医学教育项目，遴选8个典型案例，涉及6个国家。其中本科医学教育项目4个、研究生医学教育项目2个、住院医师规范化培训项目2个。持续反馈或导师反馈覆盖8个案例，覆盖率为100.00%；信息系统或学习档案系统覆盖7个案例，覆盖率为87.50%；临床表现评估、基于工作场所的评估或临床现场观察覆盖6个案例，覆盖率为75.00%；阶段性进展评议或综合审议机制、反思写作或基于学习档案的自我分析均覆盖5个案例，覆盖率均为62.50%。**结论** 程序性评价是以持续反馈、纵向证据积累、信息系统支持和关键节点综合判断为核心的系统性评价框架。我国可结合本土资源，从局部试点、最小数据集建设、电子学习档案、导师反馈和阶段性评议机制等方面分阶段推进。

【关键词】 程序性评价；　医学教育；　跨国比较；　胜任力导向；　教育评估；　反馈机制
【中图分类号】 R-0；R-4
基金项目：重庆市高等教育教学改革研究项目（232047）；重庆市高等教育学会高等教育科学研究课题（cqgj25zx10C）；重庆医科大学校级教育教学改革研究项目重大项目（JY20250202）；重庆医科大学附属第一医院研究生教育教学改革项目（CYYY-YJSGXM-202309）
DOI：10.3760/cma.j.cn116021-20260210-02317

An international multi-case comparative study of programmatic assessment in medical education

Wang Wentao[1], Lu Ziyue[1], Pan Junyi[1], Shen Jian[2], Li Xiang[2,3]
[1]Chongqing Medical University-University of Leicester Joint Institute, Chongqing 400016, China; [2]Department of Cardiovascular Medicine, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China; [3]Academic Affairs Office, Chongqing Medical University, Chongqing 400016, China
Corresponding author: Li Xiang, Email: lzd9288@163.com

【Abstract】 Objective To analyze the implementation orientation, assessment tools, feedback support, progress review, and information system support of programmatic assessment in medical education through a multi-case international comparison, and to provide a reference for medical education assessment reform in China. **Methods** PubMed was searched, with supplementary verification using Web of Science and Scopus, for relevant literature published up to November 10, 2025. Empirical studies in medical education reporting programmatic assessment or its core features were included. Studies were merged at the program level, and typical cases were selected according to program maturity, literature support, representativeness, contextual variation, and data completeness. The comparison of these cases focused on implementation orientation, evidence sources, feedback support, progress review, and information systems.

Results　A total of 3 863 records were retrieved. Finally, 55 empirical studies were included, 23 medical education programs were identified, and eight typical cases from six countries were selected. These included four undergraduate medical education programs, two postgraduate medical education programs, and two standardized residency training programs. Continuous feedback or mentor feedback was reported in eight cases, with a coverage rate of 100.00%; information systems or portfolio systems in seven cases, with a coverage rate of 87.50%; clinical performance assessment, workplace-based assessment, or direct clinical observation in six cases, with a coverage rate of 75.00%; and progress review or comprehensive review mechanisms and reflective writing or portfolio-based self-analysis in five cases each, with a coverage rate of 62.50%. **Conclusions**　Programmatic assessment is a systematic assessment framework centered on continuous feedback, longitudinal evidence accumulation, information system support, and comprehensive judgement at key progression points. In China, advancement can be made in stages by drawing on local resources in areas such as local pilots, minimum dataset construction, electronic portfolios, mentor feedback, and progress review mechanisms.

【**Key words**】 Programmatic assessment;　Medical education;　International comparison;　Competency orientation;　Educational assessment;　Feedback mechanism

Fund program: Chongqing Higher Education Teaching Reform Research Project (232047); Chongqing Higher Education Association Higher Education Scientific Research Project (cqgj25zx10C); Major Project of Chongqing Medical University Educational and Teaching Reform Research (JY20250202); Graduate Education and Teaching Reform Project of The First Affiliated Hospital of Chongqing Medical University (CYYY-YJSGXM-202309)

DOI: 10.3760/cma.j.cn116021-20260210-02317

程序性评价(programmatic assessment)是近年来医学教育领域兴起的一种评估框架,旨在通过持续、多元、综合地收集评估数据,支持学习者发展,并为阶段性决策提供更可靠的依据。这一理念最早由范德弗卢滕等人在2005年提出,并于2012年进一步发展为较为系统的理论模型[1-2]。该理念强调同步发挥评估的促学功能和决策功能,在整个培养过程中有计划地使用大量低利害、小规模的评估以获取反馈,支持学生持续改进;同时将累积的多次评估结果进行纵向整合,以全局视角做出高利害的学业判断[1-2]。

齐心[3]指出,医学教育评价已从“学习评价”向“以评促学”再到“学评融合”演进,其关注点由终结性评价与形成性评价逐步拓展至程序性评价,评价目的也从简单判定合格转向促进学习和提升教育效能,实现基于多元证据对核心胜任力进行系统评估。作为胜任力导向医学教育(competency-based medical education)的关键工具,程序性评价以“低利害评价—中期回顾—高利害决策”为核心,通过多元化、多维度的评价方法促进学习者自主学习和行为改进,并基于教学设计原则整合评价与教学,但其应用需应对信息系统改进、教师理念转变等挑战[4-5]。与传统仅依赖少数高利害考试的模式相比,程序性评价通过多来源评价数据和持续反馈为每个学生绘制能力发展的“全景图”[6]。近年来,随着胜任力导向医学教育的推广,程序性评价因其强调以多元证据支持学习者发展,并改进阶段性综合判断而受到关注,已在多个国家的医学教育项目中得到应用。然而,不同国家和院校在落实这一理念时呈现出多样化特征,既有共同的原则,也有因情境而异的设计和实践差异[6-7]。

在国内的实证研究中,程序性评价在部分本科临床教学和全科医学课程中已开展初步试点。一些研究已显示,其在提高学习参与度、促进学生反思和提供结构化反馈方面具有较大潜力。例如,在呼吸科临床教学和全科临床诊疗思维课程中,程序性评价通过混合评价方式、学员反思与多源反馈提升了学习参与度、自我评估与课程表现,在促进学习与结构化反馈方面呈现出显著的优势[8-10]。国际上的研究同样表明程序性评价能促进学生设定合理的目标、形成自身的专业身份认知和提升学习自主性。部分医学教育项目通过建设程序性评价工具、构建反馈机制和提供心理安全的环境,综合做出更为可信的高利害决策。部分研究也揭示了评

估方式对成绩的影响，在强调形成性评价的同时也需保持适当的激励，确保学生的能力水平达到相应培养标准[11]。此外，已有关联研究从测量学与评估方法学视角探讨了程序性评价的纵向评估功能，但也提示，现有模型对学习者能力变化的解释能力和预测能力仍有提升空间[12]。

综上所述，程序性评价作为一种系统性的、发展导向的评估框架，正逐渐成为胜任力导向医学教育的重要支撑工具。我国在程序性评价的研究与实践方面起步较晚、经验相对不足，因此系统总结国际先进做法对于完善我国医学教育评价体系具有重要意义。尽管我国已在部分课程与临床教学环节开展程序性评价的初步探索，但总体仍以局部改革为主，缺乏跨培养阶段、系统化的实施经验[8-10]。为此，本文拟系统梳理国际医学教育中程序性评价的代表性实践，比较不同项目的设计理念、实施路径、评估工具、反馈支持和教学融合方式，分析其运行特征、实施条件及对我国医学教育评价改革的启示。

1 研究方法

本研究通过系统文献检索与多案例比较分析的方法，对国际医学教育领域具有代表性的程序性评价项目进行识别与比较。文献检索以 PubMed 为主要数据库，并结合 Web of Science 和 Scopus 对相关主题文献及已识别项目进行补充核查，检索时间截至 2025 年 11 月 10 日。PubMed 检索式为：((("programmatic"[All Fields] OR "programmatically"[All Fields]) AND ("assess"[All Fields] OR "assessed"[All Fields] OR "assessement"[All Fields] OR "assesses"[All Fields] OR "assessing"[All Fields] OR "assessment"[All Fields] OR "assessment s"[All Fields] OR "assessments"[All Fields])) AND (1977:2025[pdat])。纳入标准为：①研究对象为具体医学教育项目，且文献明确报告其实施程序性评价，或其评估体系具备程序性评价的核心特征；②研究类型为定量、定性或混合方法研究；③能够较清晰呈现项目背景、结构设计或实施机制。概念性论文、综述、观点类文章、会议摘要，以及项目描述过于笼统、难以识别或同时涉及多个项目且未对单一项目进行清晰界定的研究，均予以排除。

文献筛选包括去重、题名与摘要筛选和全文审阅 3 个阶段。首先将检索结果导入文献管理软件进行去重，随后依据纳入和排除标准进行题名与摘要筛选，并对可能符合标准的文献进行全文审阅。筛选工作由一名研究者初步完成，另一名研究者对最终纳入文献、边界性文献及项目归并结果进行复核；对于纳入判断、排除原因或项目归属存在疑问的情况，由 2 名研究者讨论后确定最终结果。

基于纳入文献，本研究提取项目层面信息，并形成项目层面数据提取表。提取内容包括国家、机构、项目名称、培养阶段、专业方向、评估工具、反馈方式、学习支持机制、进展评议机制、信息系统支持及主要结构特征。对于同一项目的多篇文献，先在文献层面提取信息，再合并至项目层面。初筛过程中，同时记录候选项目的纳入判断、主要依据及项目概要，以支持后续归并与排除过程的追溯。在此基础上，本研究从全部候选项目中进一步筛选用于跨国比较的典型案例。筛选时主要考虑 5 个方面，即项目成熟度、文献支撑情况、案例代表性、情境差异以及资料可获得性。项目成熟度主要依据项目是否已形成相对完整的程序性评价框架进行判断，重点包括其是否包含多元低利害评估、持续反馈、纵向证据整合和阶段性进展评议等关键要素。文献支撑情况主要考察同一项目是否有较充分的实证研究、报告或资料，能够较完整地反映其设计与实施。案例代表性主要考察该项目在本国、本机构或特定培养阶段中是否具有一定典型性。情境差异主要体现于国家或地区、培养阶段、专业领域及制度环境等维度，从而提升案例比较的解释力。资料可获得性则主要指项目是否具备足够公开且可识别的信息，便于进行比较。研究者据此对候选项目进行结构化比较，并作出综合判断，最终遴选出 8 个典型案例纳入跨国比较分析。

在分析阶段，本研究采用多案例比较的方法，依据程序性评价的核心理念建立结构化的分析框架，分析维度包括理念取向、系统结构、实施机制、反馈方式、进展评议以及相关情境因素。研究者首先对各案例进行逐案审阅，再开展案例比较，以识别不同项目中程序性评价实施的共性与差异。为补充定性比较结果，本研究还对 8 个典型案例中反复出现的评估工具和关键实施机制进行描述性统计，统计其出现的案例数和覆盖率。仅在文献中明确报告相关工具或机制时，才纳入频次统计。

2　结果

2.1　文献筛选、项目识别与典型案例纳入情况

经数据库检索共获得文献 3 863 篇，去重后余 3 856 篇文献进入题名与摘要筛选阶段。经题名与摘要筛选，排除与研究主题不符、非医学教育领域、非实证研究或信息不足的文献 3 677 篇，剩余 179 篇文献进入全文审阅。全文审阅后，进一步排除概念性论文或综述、会议摘要、项目描述不清或无法识别具体实施项目的文献 124 篇，最终纳入 55 篇符合纳排标准的实证研究。通过对同一项目相关文献进行归并，共识别出 23 个实施背景明确、结构特征较清晰的医学教育项目。在进一步依据项目成熟度、文献支撑度、代表性、情境差异和资料完整性等标准进行筛选后，最终遴选出来自 6 个国家的 8 个典型案例，构成本研究的跨国比较案例集。文献筛选与案例遴选流程见图 1。

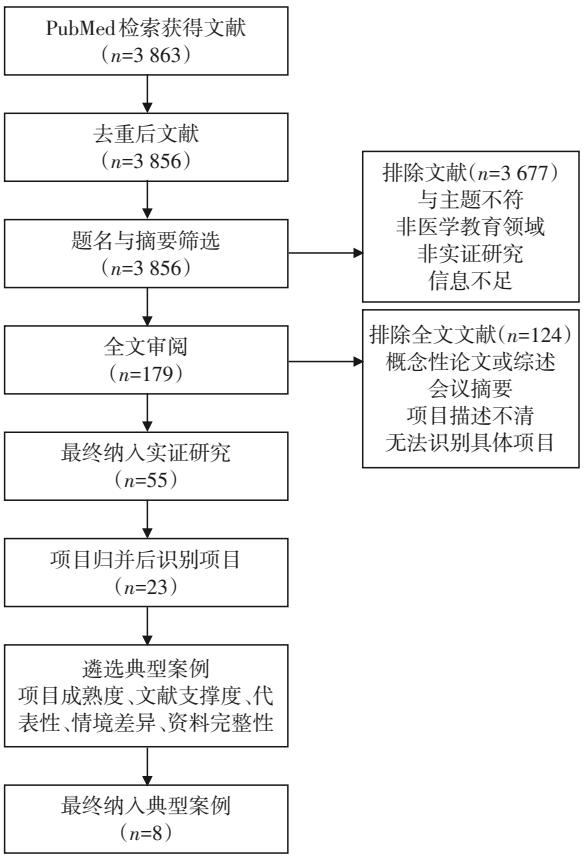


图 1　文献筛选与案例遴选流程

2.2　典型案例的基本特征

8 个典型案例来自 6 个国家，其中加拿大和澳大利亚各 2 个，各占 25.00%；英国、新西兰、荷兰和爱尔兰各 1 个，各占 12.50%。从培养阶段看，本科医学教育项目 4 个，占 50.00%；研究生医学教育项目 2 个，占 25.00%；住院医师规范化培训项目 2 个，占 25.00%。所选案例覆盖院校医学教育和毕业后医学教育 2 个主要阶段。8 个典型案例的基本特征见表 1。

2.3　评估工具与实施机制的出现情况

为进一步揭示程序性评价项目的共同组成要素，本研究对 8 个典型案例中反复出现的评估工具和关键实施机制进行描述性统计。结果显示，持续反馈或导师反馈在 8 个案例中均有体现，覆盖率为

表 1　国际程序性评价典型案例基本信息

案例编号	国家	机构	项目名称	教育阶段	学制（年）	专业领域
C1	英国	伦敦帝国理工学院	帝国理工学院临床与科学整合案例（CSI）课程	本科医学教育	6	临床医学
C2	新西兰	奥塔哥大学医学院	程序性评价与条件通过评估改革项目	本科医学教育	6	临床医学
C3	加拿大	麦克马斯特大学	麦克马斯特模块化评估项目（McMAP）急诊医学住院医师培训项目	住院医师规范化培训	5	急诊医学
C4	加拿大	多伦多大学	多伦多大学医学博士（MD）项目	本科医学教育	4	临床医学
C5	荷兰	马斯特里赫特大学	马斯特里赫特大学六年制学士—硕士医学课程	本科医学教育	6	临床医学
C6	澳大利亚	墨尔本大学医学院	墨尔本大学医学博士（MD）项目	研究生医学教育	4	临床医学
C7	澳大利亚	澳大利亚圣母大学弗里曼特尔医学院	圣母大学澳大利亚校区弗里曼特尔医学院医学博士（MD）项目	研究生医学教育	4	临床医学
C8	爱尔兰	爱尔兰全科医师学院	爱尔兰全科医师学院（Irish College of General Practitioners，ICGP）全国全科医学培训“基于工作场所的学习导向评估”体系项目	住院医师规范化培训	4	全科医学

注：“本科医学教育”主要指毕业前医学教育或初始医学学位教育阶段的医学项目，不完全等同于我国高中后直接入学的本科医学教育；“研究生医学教育”主要指完成前期本科教育后进入医学院校接受医学博士培养的项目；“住院医师规范化培训”指毕业后医学教育阶段的住院医师培训项目。

100.00%；信息系统、电子学习档案、学习平台或学习档案系统在7个案例中有明确报告，覆盖率为87.50%。临床表现评估、基于工作场所的评估（workplace-based assessment，WBA）或临床现场观察出现在6个案例中，覆盖率为75.00%；阶段性进展评议或综合审议机制、反思写作或基于学习档案的自我分析分别出现在5个案例中，覆盖率均为62.50%。具体评估工具方面，客观结构化临床考试（objective structured clinical examination，OSCE）出现在4个案例中，覆盖率为50.00%；进展测试或纵向综合测试出现在3个案例中，覆盖率为37.50%；迷你临床演练评估（mini-clinical evaluation exercise，Mini-CEX）或其变体、病例讨论或标准化病例讨论分别出现在2个案例中，覆盖率均为25.00%。置信职业行为（entrustable professional activity，EPA）相关评估、情境判断测试、临床/个人问题识别与改进计划以及个人/团队准备度测试和团队应用任务主要出现在特定项目中，覆盖率均为12.50%。8个典型案例中评估工具与关键实施机制的出现情况见表2。

2.4 不同教育阶段程序性评价项目概述

2.4.1 本科医学教育项目

本科医学教育项目主要处于院校医学教育阶段，证据来源包括课程学习表现、临床技能训练、学习档案、反思任务和阶段性进展判断等。帝国理工学院临床与科学整合案例课程以病例学习和团队学习为主线，通过个人准备度测试、团队准备度测

表2 8个国际医学教育程序性评价典型案例中评估工具与实施机制的出现情况

类别	评估工具或实施机制	判定标准	出现案例数	覆盖率（%）	涉及案例
核心机制	持续反馈或导师反馈	文献明确报告课堂反馈、团队反馈、导师反馈、临床现场反馈、周期性反馈或反馈性评估机制	8	100.00	C1~C8
技术支持	信息系统、电子学习档案、学习平台或学习档案系统	文献明确报告在线课程平台、电子学习档案、学习档案系统、评估平台、数据记录系统、进展监测界面或类似技术支持	7	87.50	C1、C3~C8
临床评价	临床表现评估、WBA或临床现场观察	文献明确报告临床表现评估、WBA、临床观察、直接观察、临床表现记录或工作场所学习评估	6	75.00	C2、C3、C5、C6、C7、C8
决策机制	阶段性进展评议或综合审议	文献明确报告条件通过机制、学生进展委员会、胜任力进展委员会、阶段性或最终综合判断，或类似进展判断机制	5	62.50	C2、C3、C4、C5、C8
学习支持	反思写作、反思记录、学习计划或基于学习档案的自我分析	文献明确报告反思任务、反思记录、学习计划、自我评估、基于学习档案的自我分析，或学习档案被用于学生反思、导师讨论和学习改进支持	5	62.50	C3、C4、C5、C7、C8
评估工具	OSCE	文献明确报告客观结构化临床考试	4	50.00	C2、C4、C6、C7
评估工具	进展测试或纵向综合测试	文献明确报告进展测试、累积成就测试、纵向综合测试或类似纵向知识能力测试	3	37.50	C4、C5、C6
评估工具	Mini-CEX或Mini-CEX变体	文献明确报告Mini-CEX、micro-CEX或类似迷你临床观察评估工具	2	25.00	C3、C6
评估工具	病例讨论或标准化病例讨论	文献明确报告CBD、SCBD或类似病例讨论评估工具	2	25.00	C6、C8
临床评价	EPA相关评估	文献明确报告EPA、基于EPA的能力进展监测或EPA进展监测界面	1	12.50	C8
评估工具	SJT	文献明确报告情境判断测试	1	12.50	C6
学习支持	PIP-C/PIP-P	文献明确报告临床问题识别与改进计划或个人问题识别与改进计划	1	12.50	C8
团队式评估	iRAT/tRAT/tAPP	文献明确报告个人准备度测试、团队准备度测试或团队应用任务	1	12.50	C1

注：同一案例中同一工具或机制多次出现时，仅按1次计；未在纳入文献中明确报告者不计入。各项工具或机制可在同一案例中重复出现，覆盖率不相加。OSCE，客观结构化临床考试；WBA，基于工作场所的评估；Mini-CEX，迷你临床演练评估；EPA，置信职业行为；SJT，情境判断测试；CBD，病例讨论；SCBD，标准化病例讨论；PIP-C，临床问题识别与改进计划；PIP-P，个人问题识别与改进计划；iRAT，个人准备度测试；tRAT，团队准备度测试；tAPP，团队应用任务。

试、团队应用任务、案例分析和团队式评估等方式收集学生学习表现[13]。奥塔哥大学在本科临床阶段推行程序性评价与条件通过改革，将基于标准的评估、OSCE、条件通过记录、专业行为或学习进展问题记录等纳入评价证据体系，并实行条件通过机制及补救支持[14]。马斯特里赫特大学六年制医学课程通过进展测试、学习档案、反馈分析和多源评估等积累纵向信息，并结合导师反馈、学习档案分析以及阶段性或最终综合判断支持学生发展[15]。多伦多大学医学博士项目通过小测验、掌握性练习、进展测试、OSCE、书面作业、反思、参与度和专业行为评价等方式收集学习表现，并通过 Learner Chart 电子学习档案系统、导师指导和学生进展委员会支持学习改进与整体性进展判断[16]。

2.4.2　研究生医学教育项目

研究生医学教育项目更强调临床准备度、专业行为和学习进展监测。墨尔本大学医学博士项目处于从传统评价向形成性评价转型的过程中，其评估体系包括进展测试及累积成就测试、Mini-CEX、OSCE、情境判断测试、WBA、标准化病例讨论和书面考试等，并通过个体化反馈报告、工作场所评估平台和学习进展监测界面支持学生能力发展[17]。圣母大学弗里曼特尔医学院医学博士项目将 OSCE、WBA、选择题或病例考试、基于评分量规的评估、反思任务和电子学习档案证据纳入整体评估体系，并通过基于学习数据的导师反馈和导师会议支持学生临床准备度发展[18]。

2.4.3　住院医师规范化培训项目

住院医师规范化培训项目的程序性评价主要体现为基于真实临床工作场景收集学习表现。麦克马斯特模块化评估项目（McMaster Modular Assessment Program，McMAP）围绕急诊医学核心胜任力建立模块化的评估机制，采用任务特异性 WBA、Mini-CEX 变体、现场观察记录和轮转表现数据等方式收集低利害评估证据，并通过学习档案、多源 WBA 数据，由胜任力进展委员会支持进展判断与补救决策[19]。爱尔兰全科医师学院以 EPA 为核心，构建了全国统一的基于工作场所的评估体系，采用 PIP-C/PIP-P、病例讨论、低利害 WBA 记录、反思记录和学习档案证据等评价资料，并通过 WBA 软件平台、EPA 进展监测界面和电子学习档案整合多源证据，由胜任力进展委员会作出高利害进展决策[20]。

2.5　程序性评价实施的跨案例比较

8 个典型案例均体现出程序性评价的共同要素，包括多次低利害评估、持续反馈以及纵向证据积累，但不同项目在证据来源、反馈组织方式、阶段性评议机制和信息系统支持方面存在差异，具体见表 3。在院校医学教育阶段，本科项目更多使用课堂学习表现、进展测试、学习档案、反思任务、临床技能评估和专业行为记录等作为证据；研究生医学教育项目则在此基础上更重视临床准备度、工作场所评估、个体化反馈报告和基于学习数据的导师指导。相比之下，住院医师规范化培训项目的证据主要来自真实临床工作场景中的 WBA、现场观察记录、学习档案和 EPA 相关证据。整体来看，多数项目借助信息系统或学习档案系统完成证据记录、反馈留痕和学习进展追踪，但正式的进展决策机制的成熟程度在不同项目之间存在差异。

表 3　国际程序性评价典型案例的结构化比较

项目/分析维度	理念定位	典型评估工具与证据结构	反馈模式	学习支持机制	进程决策机制	信息系统/技术支撑	项目特征/规模
英国帝国理工学院 CSI 课程	团队学习与早期临床科学整合	iRAT、tRAT、tAPP、案例分析、团队式评估数据	课堂反馈、团队反馈与同伴讨论	团队协作学习、学生主动学习、案例引导的整合学习	课程内多次团队式评估累计评分与学习监测	数字化课程资源、电子化团队式评估、在线测试系统	大规模本科早期课程；强调早期整合案例与团队学习
新西兰奥塔哥大学程序性评价与条件通过改革项目	风险识别、补救支持与决策透明性	基于标准的评估、OSCE、条件通过记录、专业行为或学习进展问题记录	条件通过机制触发的结构化反馈与问题反馈	条件通过机制触发的补救学习、学习问题识别、学生事务负责人或相关教师介入与访谈支持	条件通过机制结合定期进展会议，基于多源证据进行综合判断	评估结果记录、问题追踪与文书化证据链	本科临床阶段程序性评价改革；重点在风险识别、补救和决策可辩护性

续表 3

项目/分析维度	理念定位	典型评估工具与证据结构	反馈模式	学习支持机制	进程决策机制	信息系统/技术支撑	项目特征/规模
荷兰马斯特里赫特大学医学课程	程序性评价经典理念下的纵向证据积累	进展测试、学习档案、反馈分析和多源评估证据	进展测试结果反馈、学习档案反馈与导师反馈	导师支持、学习档案分析、学习反思与自我调节支持	基于学习档案及多源证据的阶段性或最终综合判断	ProF 反馈系统、学习档案系统	六年制医学课程；突出进展测试嵌入程序性评价体系，并结合临床阶段学习档案和 WBA 经验
加拿大多伦多大学 MD 项目	纵向能力画像、导师支持与整体性决策	小测验、掌握性练习、进展测试、OSCE、书面作业、反思、参与度和专业行为评价	周期性导师指导反馈、定期沟通反馈和进展测试反馈	电子学习档案系统、个人学习计划、反思记录、导师支持与针对性补救	学生进展委员会基于多源证据进行整体性进展判断	Learner Chart 电子学习档案系统	加拿大大型四年制医学博士项目；强调数据整合、导师指导，以及低利害反馈与高利害综合决策相区分
加拿大 McMAP 急诊住院医师培训项目	模块化能力追踪与工作场所表现评估	任务特异性 WBA、Mini-CEX 变体、现场观察记录、轮转表现数据	临床现场即时反馈、工作场所反馈、结构化书面/口头反馈与汇总反馈	模块化学习目标、学习档案自我评估、教师反馈、反思促进与项目层面支持	胜任力进展委员会定期审查学习档案和多源 WBA 数据，支持进展判断与补救决策	专科学习档案、电子或纸质评估记录册、WBA 记录与汇总数据	急诊医学住院医师培训项目；具有模块化、任务特异性 WBA 结构
澳大利亚墨尔本大学 MD 项目	阶段式程序性评价改革与能力发展路径	进展测试及累积成就测试、Mini-CEX、OSCE、SJT、WBA、SCBD 和书面考试	进展测试反馈、SJT 和 OSCE 等个体化反馈报告、课程或临床学院层面反馈	基于反馈报告和学习进展监测界面的学习支持、能力发展反馈与学习进展监测	逐步探索基于聚合数据的进程回顾；正式进展决策仍为实施挑战	电子化评估系统、WBA 平台、评估平台与学习进展监测界面	大型研究型大学；研究生入学医学博士项目；处于从传统评价向程序性评价转型过程
澳大利亚圣母大学弗里曼特尔医学院 MD 项目	临床准备度、真实性实践与基于学习数据的导师指导	OSCE、WBA、MCQ 或病例考试、基于评分量规的评估、反思任务、电子学习档案证据	基于学习数据的导师反馈与导师会议反馈	反思性电子学习档案、导师辅导、PBL 支持	以导师会议和基于学习数据的进程支持为主	电子学习档案、Blackboard 学习管理系统、信息通信技术支持	中等规模四年制 MD 项目；约每年 100 名学生；以基于学习数据的导师指导作为程序性评价转型步骤
爱尔兰 ICGP 全国全科医学培训	学习导向 WBA 与全国统一 EPA 框架	EPA、PIP-C/PIP-P、CBD、低利害 WBA 记录、反思记录、学习档案证据	临床督导或培训者 WBA 反馈、教育督导周期性反馈与 EPA 反馈	导师与评估者角色分离、EPA 进展监测、自我调节学习和学习档案支持	胜任力进展委员会周期性审查定量与定性证据，并作出高利害进展决策	WBA 软件平台、EPA 进展监测界面与电子学习档案	国家级 GP 住院医师培训体系；全国统一实施；制度化程度较高

注：CSI，临床与科学整合案例；iRAT，个人准备度测试；tRAT，团队准备度测试；tAPP，团队应用任务；OSCE，客观结构化临床考试；MD，医学博士；WBA，基于工作场所的评估；Mini-CEX，迷你临床演练评估；SJT，情境判断测试；SCBD，标准化病例讨论；MCQ，选择题；PBL，问题导向学习；EPA，置信职业行为；PIP-C，临床问题识别与改进计划；PIP-P，个人问题识别与改进计划；CBD，病例讨论；GP，全科医学。

3 讨论

3.1 程序性评价具有共同核心要素，但实施方式受培养阶段和情境影响

本研究结果显示，8 个典型案例均包含持续反馈或导师反馈，7 个案例明确报告信息系统、电子学习档案、学习平台或学习档案系统，提示反馈和纵向证据记录是程序性评价较稳定的核心要素。同时，临床表现评估、WBA 或临床现场观察出现在 6 个案例中，阶段性进展评议或综合审议机制出现在 5 个案例中，说明不同项目在临床情境证据使用和进展判断机制成熟度方面仍存在差异。具体而言，本科阶段更重视医学知识的整合、自我反思和早期能力的形成，研究生阶段则更强调临床准备以及从学生向医生的角色过渡，住院医师培训阶段则更加关注学习者在真实工作场景中的表现。因此，程序性评价在不同阶段实施时，所依赖的证据来源、反馈方式和重点均呈现出不同的特点。

程序性评价的这种情境化实施特征不仅体现在培养阶段差异上，制度环境与现实资源条件也同样会影响其实施方式。本研究纳入的典型案例在教师资源、信息平台、课程结构和评估方式等方面存在差异，这些因素均会影响程序性评价的实施。例如，8 个案例均涉及持续反馈或导师反馈机制，说

明反馈是程序性评价的重要组成部分。然而,这种反馈机制的建立对师资配置提出了较高的要求。在我国本科医学教育情境下,教学方式往往仍以“一对多”为主,较难实现持续的“一对一”反馈。本研究中阶段性进展评议或综合审议机制出现在5个案例中,说明集体审议或阶段性综合判断是部分成熟程序性评价项目的重要安排,但并非所有项目均已建立正式、稳定的进展委员会机制。而胜任力进展委员会通常由经验丰富的教师组成,负责对学生在学习过程中累积的阶段性低利害评价信息进行评议[3]。从实际运行负担看,学生在学习过程中会持续积累大量纵向评估数据。在委员会成员本就相对有限的情况下,如何在处理大量纵向评估数据的同时保证评议质量和判断可信度,仍需进一步讨论。更有研究指出,胜任力进展委员会在一定程度上是胜任力评价过程中大量评估数据与准确评议结果之间协调的产物,也体现了现实条件下的妥协[3]。在我国的实施场景下,这类做法更需谨慎推进。

此外,我国部分医学教育场景中仍存在师生比偏高、临床导师时间有限、信息平台分散、评估素养不均等现实因素,这也限制了程序性评价的实施。在这种情况下,若简单引入双导师制、胜任力进展委员会等安排,而缺乏岗位分工、时间保障和配套支持,这些机制就可能从原本的支持性设计转变为一线教学人员的额外负担。在我国医学教育背景下,如果脱离培养阶段和实施环境的具体情况,程序性评价可能流于形式。如果只是在形式上将各种评估工具叠加,而缺乏纵向证据整合、导师支持和集体审议,往往难以兼顾学习支持与阶段性判断。当前,我国医学教育虽在政策层面逐步转向胜任力导向,但评价体系仍以阶段性考试和终结性考试为主。纵向证据积累、学习档案、工作场所评估和集体审议等程序性评价要素,整体上仍处于初步探索阶段。因此,我国在推进程序性评价时,不宜简单照搬国外模式,而应结合本土培养体系、资源条件和临床教学实际进行情境化设计,同时避免把本应用于支持学习和改进决策的创新机制异化为新的教学管理压力。

3.2　信息系统是程序性评价从分散评估走向连续整合的关键条件

本研究结果显示,信息系统、电子学习档案、学习平台或学习档案系统在7个典型案例中有明确报告,覆盖率为87.50%。这一结果提示,程序性评价的实施并非仅仅是评价方式或评价次数的调整,还需要有能够承载大量纵向评估数据的信息系统或学习档案系统的支持。在院校医学教育中,信息系统主要表现为课程学习平台、电子学习档案、反馈系统或学习进展监测界面;在住院医师规范化培训中,则更多表现为WBA记录系统、学习档案和EPA进展监测界面。具体而言,信息系统主要用于储存个体化评议数据、导师反馈、学习进展以及阶段性回顾等多来源信息。相较于传统评价方式,这种依托信息平台开展的连续、多维度记录,不同于仅提交一次形成性测试报告并将其计入总成绩的做法。它更强调对多个来源数据的持续记录与利用,有助于将多次低利害评议所形成的信息转化为可跟踪、可解释的连续证据链,而不仅仅是简单增加测试次数。

从现有案例来看,持续反馈或导师反馈在8个案例中均有体现,反思写作、反思记录、学习计划或基于学习档案的自我分析出现在5个案例中,阶段性进展评议或综合审议机制也出现在5个案例中。这些机制通常需要信息系统或学习档案系统提供记录、整合和追踪支持,其作用主要体现在以下几个方面。首先,相较于纸质记录,电子化的信息存储方式能够实现不同来源信息的记录和呈现,更加直观地反映学习者的综合情况。其次,这种记录方式有助于保留学习者在培养过程中的发展轨迹,使阶段性评议不再仅依赖教师印象或单次测评结果,而是能够建立在可追溯的纵向证据基础之上。再次,信息化记录系统还有助于教师和学习者及时识别能力发展的薄弱环节,从而开展更具针对性的支持与反馈。值得注意的是,已有研究提到可以将评估信息接入人工智能,用于辅助处理评价资料、识别学习进展,为学习者提供个性化建议并形成针对性反馈[5]。

近年来,我国持续推进教育数字化改革,也在医学教育领域不断加强平台建设和教学数据管理,这为程序性评价所需的信息化支撑提供了一定基础。然而,需要看到,程序性评价所要求的信息系统并不只是一般意义上的教学管理平台,而是能够用于纵向评价、反馈留痕和综合判断的系统。当前我国不少医学教育场景中的信息平台仍以课程管

理为主，不同教学环节、不同课程模块之间的数据往往未实现有效互通，导致学习过程数据难以形成连续的证据链。此外，虽然目前不少医学课程已将网络平台用于过程性学习记录或课程考核，但整体上仍存在平台繁多、质量参差不齐、缺乏统一规范等问题，这也会在一定程度上限制程序性评价的实施效果。因此，在我国推进程序性评价时，信息系统的建设应作为评价改革的重要组成部分。当信息系统真正服务于评价证据的积累、整合与使用时，程序性评价才可能从分散的评估转变为具有连续性和支持性的评价体系。具体而言，可先从统一评价数据字段入手，将评估工具、评价者、能力维度、反馈内容、学生反思、改进计划和后续追踪结果纳入同一电子学习档案，以减少不同课程、轮转和平台之间的数据割裂。

3.3 程序性评价在支持学习与阶段性决策方面具有应用潜力，但需控制低利害评估的累积负担

本研究结果显示，8 个典型案例均体现出持续反馈、纵向证据积累和多次低利害评估的特征，说明低利害评估是程序性评价区别于传统终结性评价的重要基础。程序性评价能够整合原本分散的表现记录，形成连续的学习证据链，为阶段性进展判断提供更丰富的依据。与仅依赖少数高利害考试相比，这种评价方式更有利于识别学习者能力发展的动态变化，并为反馈、补救和进展决策提供支持。然而，低利害评估并不必然带来促学效果。学习者如何感知一项评估的利害程度，不仅取决于制度上是否将其定义为低利害，还与评估结果是否影响后续决策、是否存在改进机会、反馈是否具体可用，以及学习者是否理解评价证据的使用方式密切相关[6-7]。

因此，低利害评估的潜在风险不只是评估任务数量增加所带来的工作负担，还包括学习者对评估任务的高压感知和学习策略的改变。当低利害评估数量过多、任务过于碎片化、反馈质量不足，或者学习者不清楚这些证据如何被后续使用时，原本用于促进学习的评估活动仍可能被感知为持续性的考核压力。在这种情况下，学生可能将低利害评估视为需要不断应对的任务，而不是用于发现问题、获得反馈和调整学习策略的机会。部分研究也指出，评估数量的增加不仅会加重教师和学习者的工作负担，还可能使评估变得耗时、低效，甚至出现形式主义的倾向。因此，低利害评估的价值不在于简单增加评价频次，而在于通过多次低压力、高质量的反馈，帮助学习者实现持续改进。

有研究比较了不同评估模式下学生的考试表现，发现传统高利害终结性考试往往取得更高的分数和通过率，而将考试作为非高利害的程序性评估组成部分时，学生成绩略有下降[11]。荷兰一项多中心研究显示，在传统模式中终结性评价的平均成绩显著高于纯形成性评价（84.30% 与 67.50%），而当相同考试置于程序性评价框架且不计入高利害时，学生得分相对略低（约 81.00% 与 84.00%）[11]。这一结果提示，评价利害程度会影响学生的外部动机、备考投入和短期考试表现。当考试不再直接决定通过、排名或晋级时，学生可能减少针对单次考试的应试性准备，从而导致纸面成绩下降。然而，这一现象不宜简单理解为程序性评价降低了培养质量。程序性评价关注的不只是单次考试表现，而是学习者在较长时间内的能力发展、反馈利用、自我调节学习和综合胜任力形成。因此，低利害情境下的成绩下降更可能反映评价激励结构和学习策略的变化，而不能直接等同于学习能力或临床胜任力下降。

上述现象也提示，程序性评价并不意味着降低评估的严肃性。单次评估可以保持低利害，但关键节点上的晋级或结业判断，仍应依据纵向积累的多源证据和综合审议结果作出。换言之，程序性评价的严肃性不应体现在每一次评价都具有高压力，而应体现在评价标准明确、证据来源充分、反馈过程可追踪和关键决策可辩护。对于我国医学教育而言，推进程序性评价时尤其需要避免将低利害评估简单理解为“增加过程性考核次数”。更可行的做法是明确每类评价的功能定位，区分用于反馈改进的低利害证据和用于阶段性判断的综合证据；同时提高反馈质量，减少重复性、碎片化和低价值评价任务，并通过导师支持、补救机制和心理安全环境帮助学习者正确理解评价结果。只有在控制评价负担的基础上，实现低利害反馈与高质量综合判断之间的衔接，程序性评价才能更好地发挥其支持学习和改进决策的作用。

3.4 中国医学教育推广程序性评价宜采取分阶段、系统化推进路径

本研究结果显示，国际典型案例的运行普遍依

赖持续反馈、纵向证据积累、信息系统或学习档案支持以及一定形式的阶段性进展判断。其中，持续反馈或导师反馈在8个案例中均有体现，信息系统、电子学习档案、学习平台或学习档案系统在7个案例中有明确报告，临床表现评估、WBA或临床现场观察出现在6个案例中，阶段性进展评议或综合审议机制出现在5个案例中。上述结果提示，程序性评价的本土化推进不应仅从增加单项评估工具入手，而应同步考虑反馈机制、数据平台、学习支持、临床表现证据收集和阶段性判断机制的配套建设。结合国际案例比较及我国现有探索，程序性评价在我国并非完全缺乏实践的基础。近年来，已有不少研究在本科课程教学和临床教学等场景中引入程序性评价，通过混合评价、反思任务和多源反馈引导学生积极参与学习、深入反思并有效利用反馈进行改进[8-10]。但总体来看，现有的探索仍具有一定局限，主要集中于单门课程或局部教学环节中的某一部分实验性实施，且实施范围较小，在纵向证据整合、阶段性综合判断及决策机制的建设方面仍处于探索阶段。因此，我国推广程序性评价不宜直接进行大范围整体移植，而应采取分阶段、可评估、可逐步扩展的推进路径。

推广初期宜在低风险、可控制的教学场景中开展小范围试点，如核心课程、临床见习模块、实习轮转科室或住院医师单个专业基地。试点阶段的重点不宜聚焦于一次性增加大量评价工具，而应验证低利害评估、反馈记录、学习档案和阶段性回顾能否在现有师资与平台条件下稳定运行。同时，应关注评价完成率、教师工作量、学生接受度和反馈质量，为后续推广提供依据。

在试点运行基础上，应逐步建立程序性评价的最小数据集和电子学习档案。可优先统一记录学习者基本信息、培养阶段、能力维度、评估工具、评价者、反馈内容、学生反思、改进计划和后续追踪结果等核心字段，减少不同课程、不同轮转和不同平台之间的信息割裂。在此基础上，将课程测验、OSCE、Mini-CEX、工作场所评估、临床观察记录、专业行为评价、反思写作和导师反馈等证据纳入同一学习档案，使电子学习档案能够呈现学习者在不同阶段和能力维度上的发展轨迹，而不仅仅是材料上传或成绩汇总的平台。

电子学习档案的建设还应与导师反馈和补救支持机制相衔接。对于本科阶段，可由课程负责人或学业导师定期查看学习档案并提供反馈；对于临床实习和住院医师规范化培训阶段，可由临床导师结合工作场所表现开展即时反馈和阶段性反馈。对于连续出现同类问题的学习者，应形成明确的补救路径，包括个体化学习计划、针对性训练、再次观察和后续复评，避免低利害评估停留在记录层面。

在试点机制相对成熟后，可逐步建立阶段性综合评议机制，并根据试点效果决定推广范围。阶段性评议可设置在学期末、轮转结束或年度考核节点，由课程负责人、导师、临床教师和教学管理人员共同参与。评议重点不应是简单计算平均分，而应综合多个数据点判断学习者是否达到阶段性能力要求、是否需要补救支持、是否可以进入下一培养阶段。对于住院医师规范化培训，可在条件成熟时探索胜任力进展委员会或类似集体审议机制，但应同步明确教师职责、评议流程、时间保障和质量控制要求。后续推广则应依据评价完成率、反馈及时性、学生反馈利用率、教师工作负担、补救措施执行率和学生能力改进情况等指标逐步开展，避免在支撑条件不足时将程序性评价异化为新的教学管理负担。

3.5　研究的局限性

本研究仍存在一定局限。首先，本研究主要根据已发表文献和公开资料进行比较，因此，部分案例的实施细节、内部运行方式，以及难以在公开资料中充分呈现的支持条件，可能未被完整反映。其次，本研究虽然按照项目成熟度、文献支持情况、代表性、情境差异和资料可获得性等标准筛选了典型案例，但典型案例遴选仍不可避免地带有研究者的主观判断，可能存在一定选择偏倚。同时，本研究可能遗漏了部分仍处于发展阶段、尚未充分发表或公开信息较少的实践项目。此外，本研究主要关注不同项目的实施方式和关键机制。因此，本研究的结果更适合帮助理解程序性评价的分析框架和实践特点，而不宜直接比较不同项目的实施效果，也不能据此作出因果判断。最后，关于程序性评价在中国医学教育情境中的实施路径及实际效果，目前相关研究仍较有限，现有公开研究的数量也限制了对我国程序性评价整体实践情况的判断。未来，研究者还需要在不同培养阶段和不同制度背景下开展更多实证研究，进一步检验这些分析框架是否适

用于中国的医学教育情境。

4 结语

本研究通过跨国多案例比较发现，程序性评价并不是单纯改变评估工具或增加评估次数，而是一种以多次低利害评估、持续反馈、纵向证据积累、信息系统支持以及关键节点综合判断为特征的整体性评价框架。不同国家和不同培养阶段在具体实施方式上存在明显差异，但这些案例都表明，程序性评价的有效实施离不开与培养目标相匹配的制度安排、教师支持和信息系统支撑。对我国而言，程序性评价具有一定的借鉴价值，但不宜脱离本土培养体系、资源条件和教学实际进行照搬。我国未来可在已有课程改革和局部试点的基础上，围绕不同培养阶段的实际需求，循序渐进推进程序性评价探索，并通过更多本土实证研究不断完善其实施机制与应用路径。

利益冲突 所有作者声明不存在利益冲突

作者贡献声明 王文涛：文献检索、案例分析、论文撰写；鲁子越：文献复核、资料核对、论文修改；潘俊颐：案例整理、数据统计；沈健：研究指导、内容审核、论文修改；李响：研究构思、质量控制、论文终审

参考文献

[1] van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes [J]. Med Educ, 2005, 39(3): 309-317. DOI: 10.1111/j.1365-2929.2005.02094.x.

[2] van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose [J]. Med Teach, 2012, 34(3): 205-214. DOI: 10.3109/0142159X.2012.652239.

[3] 齐心. 医学教育评价的演进及各阶段特征[J]. 中华医学教育杂志, 2025, 45(5): 332-337. DOI: 10.3760/cma.j.cn115259-20240906-00930.

[4] 齐心, 闫辉, 齐建光, 等. 程序性评价在胜任力导向医学教育中的作用[J]. 医学与社会, 2023, 36(6): 140-144. DOI: 10.13723/j.yxysh.2023.06.025.

[5] 齐心. 程序性评价的历史演进及其在临床教学实践中的应用[J]. 中华医学教育杂志, 2025, 45(10): 735-739. DOI: 10.3760/cma.j.cn115259-20250626-00715.

[6] Schut S, Maggio LA, Heeneman S, et al. Where the rubber meets the road: an integrative review of programmatic assessment in health care professions education [J]. Perspect Med Educ, 2021, 10(1): 6-13. DOI: 10.1007/s40037-020-00625-w.

[7] Torre D, Schuwirth L, Van der Vleuten C, et al. An international study on the implementation of programmatic assessment: understanding challenges and exploring solutions [J]. Med Teach, 2022, 44(8): 928-937. DOI: 10.1080/0142159X.2022.2083487.

[8] 王荣, 何玉海, 史璇, 等. 疫情背景下程序性评价在呼吸科本科生临床教学中的应用[J]. 中国高等医学教育, 2025(6): 101-102, 107. DOI: 10.3969/j.issn.1002-1701.2025.06.046.

[9] 翟佳燚, 陆媛, 石建军, 等. 程序性评价在全科临床诊疗思维课程评价中的应用[J]. 中国全科医学, 2025, 28(4): 420-426. DOI: 10.12114/j.issn.1007-9572.2023.0743.

[10] 翟佳燚, 陆媛, 潘莹, 等. 全科临床诊疗思维课程引进程序性评价课程体验的质性研究：基于建构主义视角[J]. 中国全科医学, 2025, 28(25): 3193-3199. DOI: 10.12114/j.issn.1007-9572.2023.0933.

[11] Donker EM, van Rosse F, Janssen BJA, et al. The impact of summative, formative or programmatic assessment on the Dutch National Pharmacotherapy assessment: a retrospective multicentre study [J]. Eur J Pharmacol, 2025, 989: 177267. DOI: 10.1016/j.ejphar.2025.177267.

[12] Guth TA, Wolfe RM, Martinez O, et al. Assessment of clinical reasoning in undergraduate medical education: a pragmatic approach to programmatic assessment [J]. Acad Med, 2024, 99(8): 912-921. DOI: 10.1097/acm.0000000000005665.

[13] James M, Baptista AMT, Barnabas D, et al. Collaborative case-based learning with programmatic team-based assessment: a novel methodology for developing advanced skills in early-years medical students [J]. BMC Med Educ, 2022, 22(1): 81. DOI: 10.1186/s12909-022-03111-5.

[14] Wilkinson TJ, Tweed MJ, Egan TG, et al. Joining the dots: conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress [J]. BMC Med Educ, 2011, 11(1): 29. DOI: 10.1186/1472-6920-11-29.

[15] Heeneman S, Schut S, Donkers J, et al. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment [J]. Med Teach, 2017, 39(1): 44-52. DOI: 10.1080/0142159X.2016.1230183.

[16] Tait GR, Kulasegaram KM. Assessment for learning: the university of Toronto temerty faculty of medicine M.D. program experience [J]. Educ Sci, 2022, 12(4): 249. DOI: 10.3390/educsci12040249.

[17] Ryan A, Judd T. From traditional to programmatic assessment in three (not so) easy steps [J]. Educ Sci, 2022, 12(7): 487. DOI: 10.3390/educsci12070487.

[18] Bate F, Fyfe S, Griffiths D, et al. A preliminary evaluation of data-informed mentoring at an Australian medical school [J]. Asia Pac Scholar, 2021, 6(1): 60-69. DOI: 10.29060/TAPS.2021-6-1/OA2239.

[19] Chan T, Sherbino J. The McMaster modular assessment program (McMAP): a theoretically grounded work-based assessment system for an emergency medicine residency program [J]. Acad Med, 2015, 90(7): 900-905. DOI: 10.1097/ACM.0000000000000707.

[20] Hanley K, McSwiney E, Malley BO, et al. How we built workplace based assessment-for-learning in Irish GP training [J]. Perspect Med Educ, 2025, 14(1): 411-422. DOI: 10.5334/pme.1428.

（收稿日期：2026-02-10）

（本文编辑：兰明娟）