·胜任力评价专题·

胜任力导向医学教育中的评价及其文化变革

何睿　齐心
北京大学第一医院整形烧伤科，北京 100034
通信作者：齐心，Email：randyq@126.com

【摘要】 胜任力导向医学教育（competency-based medical education，CBME）中的评价范式和内涵正经历着深刻的变革，程序性评价因其将学习与成果、评价与教学有机结合，成为CBME的标准评价范式。随之，评价文化需要变革，突破既有的认识误区，赋予评价更多的社会属性，纳入并完善项目评估，以实现评价体系的系统优化。本文旨在为我国CBME评价体系的建设提供理论参考与实践启示。

【关键词】 胜任力导向医学教育；　程序性评价；　评价范式；　医学教育评价；　项目评估
【中图分类号】 R-0；R-4
基金项目：2025 年度北京大学医学部教育教学改革与研究课题示范智慧课程建设专项（2025SFZD10）
DOI: 10.3760/cma.j.cn116021-20260414-02331

Assessment in competency-based medical education and its cultural transformation

He Rui, Qi Xin
Department of Plastic Surgery & Burns, Peking University First Hospital, Beijing 100034, China
Corresponding author: Qi Xin, Email: randyq@126.com

【Abstract】 The assessment paradigm and its connotation in competency-based medical education (CBME) are undergoing profound changes. Programmatic assessment has become the standard assessment paradigm for CBME because it organically integrates learning with outcomes and assessment with teaching. Accordingly, the culture of assessment must also evolve. It is essential to break free from existing misconceptions, endow assessment with more social attributes, and incorporate and improve program evaluation to achieve systematic optimization of the assessment system. This article aims to provide theoretical reference and practical insights for the development of CBME assessment systems in China.

【Key words】 Competency-based medical education;　Programmatic assessment;　Assessment paradigm;　Medical education assessment;　Program evaluation
Fund program: Special Project for Demonstration Smart Course Construction under the 2025 Education Teaching Reform and Research Program, Peking University Health Science Center (2025SFZD10)
DOI: 10.3760/cma.j.cn116021-20260414-02331

医学教育变革迈入胜任力导向的医学教育（competency-based medical education，CBME）时代，评价范式与内涵也随之发生巨变[1-2]。评价与学习的关系，从学习评价（assessment of learning），转变为以评促学（assessment for learning），最终实现学评融合（assessment as learning），并以程序性评价（programmatic assessment，PA）作为CBME的标准评价范式。深入理解这种评价范式与内涵的变革，需要冲破既有评价理念的桎梏，重新构建适合CBME的评价文化。

1　评价基础与范式

评价涉及教育学、心理学、人文科学，甚至统计学原理，如何将其科学、有效地应用于教学，教育学家进行了大量的探索。医学教育遵循其普遍规律，并根据专业特点，发展出独具特色的高质量评价标准、PA 和项目评估，由下至上、自上而下地践行 CBME。

1.1　评价模型

20 世纪下半叶，国际医学教育学界相继提出多种模型以指导评价。这些模型将复杂的理论概念可视化，全面提升了指导医师对评价的理解与应用。经典模型包括 Bloom 教育认知分类法[3]、Miller 金字塔[4]和 Kirkpatrick 评估模型[5]等。它们能够有效协助指导医师优化课程设计、教学指导和评价流程。其中，Bloom 教育认知分类法偏重于课程设计，Miller 金字塔专注于评价，Kirkpatrick 评估模型从结果角度解读评价，教育者需要根据教学需求、对象和阶段，合理选择评价模型。另外，这些模型还能在师资培训时，借助更形象化的方式，将复杂、概念化的评价方法和理念尽快根植于指导医师的脑海中，并加速其有效应用。

1.2　高质量评价标准

针对单次评价，高质量评价框架[6]包括：效度（validity）或者一致性（coherence）；可重复性（reproducibility）、信度（reliability）或者稳定性（consistency）；等效性（equivalence）；可行性（feasibility）；教育效能（educational effect）；催化作用（catalytic effect）；可接受性（acceptability）。

高质量评价系统[7]则应具备连贯一致（coherent）、持续不断（continuous）、全面综合（comprehensive）、切实可行（feasible）、目标驱动（purposes driven）、可接受（acceptable）、透明公正（transparent and free from bias）。高质量评价体系将多项评价方式有机地聚焦于教学目标，全面培养学习者的核心胜任力。其将时间维度纳入评价体系，促进评价架构的立体化，推动了 PA 的应用。

1.3　程序性评价

传统评价体系难以将学习成果与临床结局相关联，PA 将学习与成果、评价和教学有机结合起来。CBME 中的学习框架由有限的学习成果组成，成果可以表现为核心胜任力、角色、职业行为，甚至技能，并以此为框架，形成学习—评价—指导的路径。在 PA 中，学习者与学习过程处于核心位置，真实临床任务促使学习者主动获取并构建知识体系，引导学习者适应并逐步建立自主学习能力；评价的目标是促进学习，而非单纯用于决策，评价成为连接学习和教学的纽带，每次评价必须提供学习者表现数据和有意义的反馈，并隐含着学习者将采用反馈指导学习的预期。PA 的整体信效度等测量学特征取决于纵向、多次评价的整体评估。在此框架下，纵向评价采用多源定期反馈机制，由指导医师、同伴及学习者本人通过描述性反馈共同评价学习者的行为表现。评价方法的安排与选择均需经过整体规划，确保 PA 传递给学习者的教育信息具有连贯性。因此，PA 也成为 CBME 的基本要素之一，从评价的视角，引导胜任力导向的教学与学习[8]。

1.4　项目评估

教育项目评估是“系统收集和分析与项目设计、实施及成果相关的信息，旨在监测和提升项目质量与效果”[9]，是对教育过程、方法及政策的系统性考察，支持项目从设计、实施、评估到修订的全流程持续优化。评估可由外部认证机构（例如美国毕业后医学教育认证委员会）强制要求进行年度评估，也可由机构自主开展内部审核，从项目内外部探究导致成果差异的根源，同时判断这些差异来源乃至项目成果本身具有的积极或消极意义。定期评估能够基于数据收集与趋势分析，揭示和（或）识别哪些因素促成了项目的“成功”，以及需要采取哪些措施来解决评估过程中发现的问题，确保教育目标的持续改进。当评估标准在不同项目或教育体系间统一实施时，还能够实现横向对比与优化提升。项目评估常采用的模型包括 Kirkpatrick 评估模型、逻辑模型、背景—投入—过程—产出（context-input-process-product，CIPP）模型、现实主义评估（realist evaluation）等，可用于不同类型的项目评估。从教育项目角度来看，这些模型旨在将数据收集与分析流程系统化，既可评估项目成效，又能对项目进行综合判断，并充分考虑项目的复杂性。选择模型时应重点考虑其适用性，评估不存在唯一

正确答案，不同模型往往能为关键问题提供差异化视角。例如，逻辑模型、CIPP 模型、Kirkpatrick 模型等多维度分析可用于项目成效评估，其中 CIPP 模型特别适合探究非预期成果，而逻辑模型和 Kirkpatrick 模型则专为衡量预期成果而设计；逻辑模型有助于厘清教育活动与预期成果间的关联，而 Kirkpatrick 模型则可对这些成果进行系统化梳理。评估者应根据其可用的人力与数据资源选择合适的模型[10]。项目评估以自上而下的视角，综合审视 CBME，在系统层面改进教学过程和成效。

2 评价文化的变革

医学教育改革，除了制度方面的变革之外，如何改变既有评价文化同样重要。教育者过去习惯将学习者知道多少等价于他们的能力强弱，但二者之间的关系在现代医学中越来越难以等价。一方面，有限的试题无法测量出“无限”的知识；另一方面，人工智能几乎能够随时调用现有已知知识，知识成为最容易获取的资源。教室里学习者的抬头率，转化为教师的“焦虑”，实实在在地展现出学习生态的变化，促使教育者反思既有教学和评价文化。教育者的角色，从“灌输者”“裁决者”，转变为“引领者”。评价不再单纯是检验学习成效的工具，还要成为学习的助力，加速实现以学习为中心的 CBME。

2.1 评价内涵的转变

CBME 中学习文化的变革深深影响了评价文化，需要指导医师与时俱进，重新审视既有观念，突破评价中的一些误区。

2.1.1 客观评价优于主观评价

客观评价基于事实、数据和可观测的标准，而主观评价常常基于个人的感知、经验、直觉和价值。CBME 中各项胜任力级别的锚定，大多只能依托主观判断。虽然管理者可以详细描述核心胜任力各个级别的典型表现，但不同评价者如何解读，以及不同学习者的行为表现能否契合这些标准，其判断过程相对复杂，甚至难以一一对应。主观评价实施前，常常需要开展评价者培训，这是统一评价尺度的关键步骤。即使如此，主观与客观的边界在临床医学中仍然相对模糊。即便对于多项选择题，以 A2 型题为例，在其客观化的过程中，如何将临床案例转化为“客观”题型，也需要不断克服出题者的主观因素，例如不同地区医师诊疗习惯和常规的差异等。另一个误区是客观等同于科学，似乎评价越客观越“高级”。不可否认，主观感受客观化的过程，敦促教育者系统思考、梳理其间不可或缺的关键要素，旨在有效地评价学习者。但是这种客观化在一定程度上也是简化过程，结构化后的评价维度固然能够达到较好的评价者间一致性，增加横向、纵向可比性，但是有可能“主动”舍弃了很多“内涵”和“机会”。如前所述，级别越高，就意味着其评价内涵越丰富、越复杂，越难以准确地进行结构化描述，随之而来的是，将主观评价客观化所需的成本效益比越低。另一方面，主观评价在这些复杂行为的评价方面，反倒有可能展现出更高的精准度，这完全依托于评价者的专业性与经验。因此，主观和客观评价在临床教学中并无优劣之分，其选择更多地需要评价者和设计者回归初心、紧扣评价对象，做出适当取舍。

2.1.2 量化评价高于质性评价

量化评价基于实证主义，侧重于甄别、选拔和预测。量化评价结果已经成为教育管理部门监测教学质量的重要手段以及评优、晋级的重要依据，其在判断学习者是否达到预期培训目标，并针对性选择补救措施方面的作用反倒退居其次。大部分量化评价以充分抽样为假设基础，评价的广度与深度能否覆盖预设的学习目标是量化评价首要回答的核心问题。单纯依赖成绩，可能出现相同成绩者之间能力千差万别的现象。由于评价的抽样偏倚，能力相近者的成绩也可能相差悬殊。成绩无法完整展现学习者之间的差异，它更多时候只能提示学习者哪些方面存在学习障碍，却难以呈现具体问题。一方面，评价的量化过程，促使评价者和学习者聚焦于评价标准，与此“无关”的行为表现“天然”会被摒弃或者忽略；另一方面，评价所涉及的部分内容即使非常重要，如果无法量化，则有可能在评价设计之初即被舍弃，这必然导致学习广度和深度的“缺陷”。质性评价，如同质性研究一样，基于解释主义/建构主义，重视过程、原因、体验与意义，透过现象“看”本质，记录并描述学习者的行为表现，

探寻支撑行为的深层原因。质性评价一般难以直接生成成绩，长期作为量化评价的补充。随着档案袋、学习日志等教学和评价方法的普及，质性评价的作用得以逐步彰显，使指导医师与学习者能够利用碎片化时间，跨越时空障碍，也能够避免面对面交流时的情感尴尬，实现双向交流。教学双方也可以定期回顾这些记录，发现学习者成长过程中的缺失，修订培训/学习计划，尤其有助于学习者自适应地主动调整，实现个性化学习。随着科学技术的发展，尤其是以大语言模型等为代表的人工智能方法在医学教育中的应用，有望进一步提高质性评价的效率与精准度，降低其应用难度，加速其普及。

2.1.3　过程性评价等同于形成性评价

医学教育项目规模庞大且形式多样，医师的成长过程漫长崎岖，传统的终结性评价形式虽然能够在结业时有效甄别出“不合格者”，但是经过长时间培训之后，重新返修的风险和代价对学习者、教师，乃至机构和社会都难以承受。此外，终结性评价可能催生出不良的外部动机，助长“及格万岁”以及学习完全围绕着课程要求和及格标准的行为。医学教育者逐步关注培训“过程”，希望通过评价尽早发现问题，并予以不断修正，强过程以固结果，由此诞生出形成性评价和过程性评价的概念。二者之间既有相同之处，都是描摹学习过程，获取学习者信息；不同之处也显而易见，过程性评价更偏重于测量，旨在监督学习者阶段性学习成果达成情况，本质上是一种终结性评价，其更多的是从“教”出发，发起者和获益者往往是机构管理者和教师，而形成性评价则侧重于“形成”，即关注学习者的成长，重点在“学”，即以评价结果为指导，开展及时、有效的干预——反馈，实现以评促学，学习者从被动受教者，转变为学习的主导者。形成性评价是“教师和学生在教学过程中使用的评价过程，该过程提供反馈以调整正在进行的教和学，目的是提高学生对预期教学结果的成绩”[11]，它是一种持续性的评价与调整过程：明确学习目标、采用多样化评价方法、提供具备可操作性的反馈、根据收集数据调整教学策略。形成性评价需要回答 3 个基础问题：学习者在哪里、他们要去哪里以及如何到达那里。形成性特性促使其进行持续性评估，用以保持学生学习积极性、确保学业有效推进。通过设置“定期”的学习检查点，教育者能够及早发现学习者的薄弱环节并及时干预。这种教学方法不仅能够有效避免学习者长期累积学习困难，还能够促进能力的持续发展。置于 PA 之中，形成性评价不再是零散任务的堆砌，而是与课程目标相契合、激发学生参与度并提供可操作反馈的系统性设计。频繁开展低利害的形成性评价，可以通过提取效应（retrieval effect）增强学习者长期记忆，提高知识留存率，促使他们反思自身、开展刻意练习、培育成长型思维，并逐步掌控自身学习进程。形成性评价可以根据学习者的学习目标、阶段，选择不同的评价方式，例如笔试、口试、模拟评价和基于工作场所的评价等。形成性评价由此从一种评价方式蜕变为一种促进学习的方法，参照学习目标，教师与学习者共同比对、检验、改进、完善，循环往复，螺旋式提高学习者的能力和行为[12]。

2.1.4　混淆高利害评价和低利害评价

在教育测量与评价领域，高利害评价和低利害评价是根据评价结果所致后果的严重程度进行划分的。两者的核心区别在于“评价结果对学习者、教师或学校产生的直接利益关联程度”。PA 中未采用形成性评价和终结性评价的概念，而选用高利害评价和低利害评价进行描述，以利害程度为基准设计评价框架。单就利害程度的概念而言，国内教师接触不多，其出发点还是希望通过标定利害程度，打破大多数评价在教学双方心目中单纯的“测量”成见，以增加其教育学属性，即不去过度纠结“评”的准确与否，而更多地放眼于学习者“做”的好与不好，哪些好的方面需要固化，且通过评价使其不断内化，哪些不足之处需要逐步改进。低利害场景下获得的评价证据被用来提升学习，从而为学习者营造出支持性的学习环境与氛围。形成性评价、查房时的随机提问等即为低利害评价；出科考核、年度考核等可以归类于中等利害程度评价，虽然这些考核具备终结性评价的属性，但在结束后给予适当反馈，就使其同时具备了形成性特征；执业医师考核、结业考核等则属于高利害评价。利害程度的分类，属于一种态势感知，一目了然，从学理上便于教学双方理解，减少了教育学方面的认知负荷，有利于

降低PA的实施难度。

2.1.5 重视模拟评价，忽视基于工作场所的评价

随着科技发展，模拟技术取得长足进步，几乎能够实现将真实世界中的一切挪到实验室中，形成医学教育中的“楚门世界（The Truman Show）”。在模拟情境中指导学习者的刻意练习，引导学习者开展有目的、可控、安全的学习，成为保障医疗质量和患者安全的必由之路，尤其在外科手术教学中，科学仪器测量、记录手法、力度、运行轨迹，精准的评价、反馈、复盘，有效提高学习者的操作熟练程度，增强他们的自信心，也缩短了外科医师的成长曲线。随着标准化病人（standardized patient，SP）的出现，模拟对学习者的非技术性技能进行评价成为可能，并且越来越受到教育者的重视。由此，模拟也成为临床前学习的重要方式。但过度依赖科技可能加剧教育公平性的问题，同时需要警惕“去临床化”的趋势，试图完全采用科技方式取代临床实践无疑是饮鸩止渴。无论模拟多么“逼真”，所有过程都是在预设疾病脚本框架下的“表演”。因为存在无限“循环”的可能，教学双方怀着“大不了从头再来”的预设从容应对，这显然与临床的随机性、不可重复性、责任感背道而驰。将实验室搬到诊室、病房中的原位模拟，也仅仅是某种特殊临床情境下的“预演”。虽然基于工作场所的评价（workplace-based assessment，WBA）存在诸多测量学缺陷，但临床责任、紧迫感等无形压力下学习者的应对是无法模拟的。需要牢记，模拟环境下学习者习得的是“技能”，只有该技能在工作环境中得到验证后，方可称之为“胜任力”[13]。换言之，CBME的目标是让学习者能够胜任临床实践，由此WBA应该成为胜任力评价的“金标准”。虽然WBA不完美，但应该作为临床教学中的常态，为指导学习者的胜任力成长和高利害决策提供有力证据。

上述诸多专业名词在临床教学实践中的边界较为模糊，例如形成性评价在院校教育阶段多采用模拟形式，而在毕业后阶段尽可能选用WBA。这需要教育者设计评价体系时，系统地整合、管理各类评价方式与方法，将评价与学习深度融合（学评融合），并以最简单明了的方式告知指导医师针对什么情况采用哪种方式进行评价，旨在引导、促进学习者开展哪方面的学习（以评促学）。

2.2 赋予评价更多社会属性

CBME中的评价从指导医师的单向输出，演变成学习者与导师/教练之间的对话，指导医师帮助学习者不断构建出他们行为表现的意义，引导学习者讨论、反思。通过反馈，指导医师促使学习者认知到自身优势和不足之处，产生内源性学习动力，加速将建构出的“意义”向临床行为的转化。由此，CBME中的评价具备了更丰富的社会属性。PA的核心源于意义建构（meaning making）理念，通过长期、多次评价，PA能够提供大量关于学习者行为表现的信息，学习者则需要分析、整合这些信息，主动构建出这些信息的意义。这一过程需要得到指导医师的引导，使学习者从被动接受者转化为主动建构者，这天然赋予医学教育评价以社会属性；而指导医师的角色也随之从裁决者的上帝视角，转变为引导者或者教练的平行视角。同样，从某种意义上，评价的过程不只是利益相关者之间的交流，其本质是共享认知的过程，这更有助于教学双方建立互信关系、营造有益的反馈文化。

反馈也成为评价社会属性的具象体现。反馈蕴含着大量信息，其既是评价依据的载体，也是改进学习者行为的主要措施。在当下教育文化中，反馈成为“训诫”的代名词，忽略了定期反馈中学习者的获益。例如，周期性反馈行为“迫使”指导医师观察学习者的行为表现，并对比既往表现，以帮助学习者解读评价信息、进行自我分析，逐步培养学习者对自身学习过程和成长的自主权。对学习者而言，评价长期被视为高利害行为，日常评价中大概率面临负面反馈的风险，因此天然恐惧并抗拒这种评价与反馈。如何使学习者坦然面对反馈、不将负面反馈视为障碍而是学习机会，确实需要教育者思考。多数反馈模型选择以正向反馈作为起点，就是希望能够降低反馈的负面效应。在学习者成长过程中，负面评价和反馈无可避免。一方面，通过PA变革评价文化，降低日常评价的利害程度，在制度层面增加低利害评价的频次，不将其纳入正式的考核体系；另一方面，健全电子档案袋系统，鼓励学习者将负面体验记录在作品集中，让“不完美”成为常态，指导医师据此帮助他们设定新的学习目标、反

思改进、调整实践行为并积累新的证据，使其从中获益。

2.3　引入项目评估

医学教育长期采用线性思维，将医学教育分割成预科、基础和临床等不同阶段，不同阶段各自独立，虽经课程整合，但其仍然缺乏内在连贯性。传统的评价体系通过系统收集和解读信息来做出判断并提出改革建议，其核心仍聚焦于内部量化流程。教育项目评估通常以单一干预措施（如新型教学法）的成效作为衡量标准，且往往仅报告直接受影响的教学双方情况，而非从系统层面进行整体评估。结果导向的教育框架往往忽视教学的复杂性，转而采用针对可测量能力的碎片化方法。然而，医学教育项目受到诸多内外因素的影响，如项目参与者的特征、利益相关者或监管机构的影响、学科基础知识的不断变化、专业实践模式以及教育项目运作的环境等。项目评估需要从追求工具性和渐进性的线性变化，转向关注教育者、学习者和患者行为的多层次改变、医疗保健结果的改善以及群体态度的转变。

项目评估与质量保障在其他行业和国外医学教育中已得到广泛应用，并形成了很多可供借鉴的模型用于临床教学，但在我国毕业后医学教育中，尤其是临床医学教育项目中，相关探索与实践仍然缺乏，教育者需要认知教育项目的复杂性，选择并应用适当的项目评估模型，积极开展评估，螺旋式改进项目，“自上而下”提升学习者的胜任力。

3　结语

CBME的评价应该是多层次的：宏观层面，国家管理机构规划出CBME的方针和政策，以指导行业协会制定相关胜任力指标，据此设计并实施CBME方案；微观层面，依据具体的胜任力指标，选择相应的评价工具与方法，并按照培训要求，设计基于置信职业行为的里程碑评价体系，采用PA模式，系统、综合地评价、指导学习者成长；中观层面，监管机构和培训机构分别从外部和内部对项目进行审核、评估，确保基地和机构能够全面、有效并持续优化CBME体系。我国医学教育者需要不懈努力，从制度和文化层面，不断完善我国胜任力评价体系。

利益冲突　所有作者声明不存在利益冲突

作者贡献声明　何睿：文献查阅、文章起草；齐心：提出思路、对文章的知识性内容作批评性审阅及修改

参考文献

[1] 齐心. 置信职业行为的演进、内涵与实践[J]. 中华医学教育探索杂志, 2024, 23(9): 1153-1160. DOI: 10.3760/cma.j.cn116021-20240201-01830.

[2] 齐心. 置信职业行为：通往胜任力导向医学教育之路：基于《医学教育中的置信职业行为与置信决策》的解析[J]. 中华医学教育探索杂志, 2025, 24(6): 721-727. DOI: 10.3760/cma.j.cn116021-20250310-02054.

[3] Krathwohl DR. A revision of Bloom's taxonomy: an overview [J]. Theory Pract, 2002, 41(4): 212-218. DOI: 10.1207/s15430421tip4104_2.

[4] Miller GE. The assessment of clinical skills/competence/performance [J]. Acad Med, 1990, 65(9 Suppl): S63-S67. DOI: 10.1097/00001888-199009000-00045.

[5] Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's four levels of training evaluation [M]. Alexandira VA: ATD Press, 2016.

[6] Norcini J, Anderson B, Bollela V, et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference [J]. Med Teach, 2011, 33(3): 206-214. DOI: 10.3109/0142159X.2011.551559.

[7] Norcini J, Anderson MB, Bollela V, et al. 2018 consensus framework for good assessment [J]. Med Teach, 2018, 40(11): 1102-1109. DOI: 10.1080/0142159X.2018.1500016.

[8] 齐建光. 胜任力导向医学教育实施要素与实践思考[J]. 中华医学教育探索杂志, 2025, 24(4): 433-439. DOI: 10.3760/cma.j.cn116021-20250122-02020.

[9] ACGME International. Glossary of terms [EB/OL]. [2026-04-17]. https://www.acgme-i.org/Accreditation-Process/Glossary/.

[10] Blanchard RD, McDaniel KE, Engle DL. Five ways to get a grip on applying a program evaluation model in health professions education academies [J]. Can Med Ed J, 2024, 15(5): 142-145. DOI: 10.36834/cmej.77995.

[11] 曹飞. 形成性评价前沿动向及借鉴[J]. 湖南师范大学教育科学学报, 2024, 23(2): 87-96, 116. DOI: 10.19503/j.cnki.1671-6124.2024.02.011.

[12] 齐建光. 探讨形成性评价与住院医师自主学习能力的培养[J]. 中国毕业后医学教育, 2020, 4(1): 33-36. DOI: 10.3969/j.issn.2096-4293.2020.01.009.

[13] Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78 [J]. Med Teach, 2013, 35(6): e1197-e1210. DOI: 10.3109/0142159X.2013.788789.

（收稿日期：2026-04-14）

（本文编辑：兰明娟）