

# 研究生教育质量评估研究综述：理论演进、主体关系与实践进展

付小鹏<sup>1</sup>, 王 屹<sup>2</sup>

<sup>1</sup>重庆理工大学重庆知识产权学院, 重庆

<sup>2</sup>重庆城市管理职业学院数智财经学院, 重庆

收稿日期: 2026年9月26日; 录用日期: 2026年10月22日; 发布日期: 2026年10月29日

## 摘 要

研究生教育质量评估是研究生教育治理的关键环节, 也是高层次人才培养质量保障的制度基础。本文检索中外主要数据库, 围绕理论基础、评价主体、指标体系、方法工具与国际比较五个维度, 系统梳理该领域研究的理论演进与前沿议题。研究发现: 质量观已由学术导向与市场导向的分割走向共同利益视角的整合重构, 评价范式由投入-产出转向学习者发展; 评价主体走向多元共治, 但学生主体仍显缺位; 在纠正单一量化指标导向的政策背景下, 指标体系正走向综合化与多元化。本文认为, 已有研究在理论框架与方法工具上贡献显著, 但在学生主体视角、指标本土化与跨学科适用性方面仍存不足, 其深层根源在于现代大学治理的科层制逻辑、学术共同体的专业逻辑与人才培养的个性化逻辑之间的结构性变化。未来研究应聚焦数智技术与评价的深度融合、共同利益视角下的理论建构与增值评价的深化。

## 关键词

研究生教育质量, 质量评估, 质量保障, 增值评价, 共同利益

# A Review of Research on Graduate Education Quality Assessment: Theoretical Evolution, Stakeholder Relations and Practical Progress

Xiaopeng Fu<sup>1</sup>, Yu Wang<sup>2</sup>

<sup>1</sup>Chongqing Intellectual Property School, Chongqing University of Technology, Chongqing

<sup>2</sup>School of Digital Finance and Economics, Chongqing City Management University, Chongqing

Received: September 26, 2026; accepted: October 22, 2026; published: October 29, 2026

文章引用: 付小鹏, 王屹. 研究生教育质量评估研究综述: 理论演进、主体关系与实践进展[J]. 教育进展, 2026, 16(10): 2274-2290. DOI: 10.12677/ae.2026.16102347

## Abstract

Graduate education quality assessment is a key link in graduate education governance and an institutional foundation for quality assurance in high-level talent cultivation. Drawing on major Chinese and international databases, this paper systematically reviews the theoretical evolution and frontier issues of the field along five dimensions: theoretical foundations, evaluation stakeholders, indicator systems, methods and tools, and international comparison. The review finds that conceptions of quality have moved from the divergence between the academic orientation and the market orientation toward an integrated reconstruction from the perspective of the common good, and that the evaluation paradigm has shifted from input-output to learner development. Evaluation stakeholders have moved toward multi-stakeholder co-governance, yet students remain underrepresented. Against the policy background of correcting a single quantified-indicator orientation, indicator systems are becoming more comprehensive and diversified. The paper argues that although existing research has contributed significantly in theoretical frameworks and methodological tools, gaps remain in the student perspective, the localization of indicators and interdisciplinary applicability; these gaps are rooted in the structural tension among the bureaucratic logic of modern university governance, the professional logic of the academic community and the individualized logic of talent cultivation. Future research should focus on the deep integration of digital-intelligent technologies and evaluation, theoretical construction from the common-good perspective, and the advancement of value-added evaluation.

## Keywords

Graduate Education Quality, Quality Assessment, Quality Assurance, Value-Added Evaluation, The Common Good

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 时代背景与问题提出

研究生教育位居国民教育体系的顶端,是知识生产、人才培养与国家创新能力建设的关键枢纽。进入 21 世纪以来,知识生产模式正经历从模式一向模式二的深刻转型——知识生产不再囿于学科内部同质性的认知旨趣,而日益发生于跨学科、应用导向与多主体参与的异质性情境之中[1]。这一转型重构了研究生教育的功能定位:不仅要培养在既有学科框架内深度探索的学术传人,更要造就能在开放情境中整合多元知识、回应复杂社会问题的创新主体。就培养目标而言,研究生教育不再以学术再生产为唯一旨归,而须在学术训练、专业能力与社会责任之间寻求新的平衡;就质量标尺而言,单一的学术标准已不足以涵盖多元化的功能面向,如何界定并测度好的研究生教育由此成为悬而未决的问题;就评价权力而言,当利益相关者从学术共同体扩展到政府、市场与社会公众,质量判断的合法性来源也随之分散。研究生教育质量评估由一项技术性管理事务,上升为关乎高层次人才供给质量与国家创新体系效能的战略命题。

与此同时,我国研究生教育在规模上实现了跨越式扩张。据教育部统计,2023 年全国在学研究生总

规模达 388.29 万人[2], 我国已跻身世界研究生教育大国之列。规模的急剧扩张与质量的持续提升之间形成了日益突出的内生性矛盾——质量不高、创新能力不足被普遍视为制约研究生教育高质量发展的突出短板[3]。如何在规模扩张中守住质量底线、提升培养水平, 成为亟待回答的重大课题, 也使如何评价研究生教育质量这一问题获得空前的理论与实践张力。

教育评价在这一格局中处于枢纽地位。评价标准在很大程度上决定着办学导向: 有什么样的评价尺度, 就会形成什么样的办学取向。2020 年 9 月, 中共中央、国务院印发《深化新时代教育评价改革总体方案》<sup>1</sup>, 系统部署了改进结果评价、强化过程评价、探索增值评价、健全综合评价的改革方向, 并明确提出纠正存在人才评价中以单一量化指标为导向的倾向[4]。就研究生教育而言, 构建符合中国国情和教育发展规律的质量评估理论体系, 既是评价体系综合化重构的必然要求, 也是实现研究生教育高质量发展的关键举措。这构成本文的现实关切与理论抱负: 在梳理既有研究的基础上, 揭示研究生教育质量评估研究的理论演进逻辑、主体张力结构与范式转换趋向, 为本土化评估理论的建构提供学理参照。

## 1.2. 核心概念的学理界定

概念界定是评估研究的逻辑起点, 也是理论建构的前提性工作。围绕研究生教育质量及其评估, 学界形成了多层次、多维度的理解, 其间既有共识, 亦有分歧。

教育质量与研究生教育质量。教育质量是一个公认的多义乃至难以捉摸的概念。哈维与格林在《界定质量》这一被广泛引用的经典文献中, 将高等教育质量观归纳为五种相互关联的理解: 质量即卓越、质量即完善或一致性、质量即目标适配、质量即物有所值、质量即转化[5]。巴奈特进一步指出, 质量始终是价值负载的, 不同的利益相关者——政府、学生、教师、雇主与社会——对何为好持有不同的判断标准, 因此质量本质上是一个相对概念, 取决于由谁来界定、以何种基准来度量[6]。这为研究生教育质量评估的多元主体研究奠定了认识论根基。

就研究生教育质量而言, 国内学者的界定大体沿广义-狭义与静态-动态两条线索展开: 王亚杰等将其视为涵盖研究生教育所有主要功能与活动的多层概念[7]; 杨颖、陈学飞提出多样化、发展性、相对性、服务性是其基本特征[8]; 盛明科等则将其界定为研究生教育满足个人、社会及学科发展显性或隐性需求的一组特性或特征的总和[9]。《中国学位与研究生教育发展战略报告(2002—2010)》的表述——研究生教育是培养单位所培养的学生、创造的知识以及提供的服务满足现在和未来的学术需要、社会需要和学生个性发展需要的充分程度[10]——因兼顾学术价值、社会价值与个体价值, 成为较具共识性的界定。

质量评估、质量评价与质量保障的界分。三者既有联系又有区别。质量保障是涵盖标准设定、过程监测、结果评估与反馈改进的制度性安排, 其要义在于持续与责任; 质量评估是对质量状态作出事实判断与价值判断的核心环节; 质量评价在中文语境中常与评估互用, 但更突出价值判断的规范性维度。本文在多数场合将三者置于同一概念簇中讨论, 仅在国际比较部分沿用各国制度话语的本义。这一界分关涉评估的功能定位——究竟是将评估理解为外部问责的技术手段, 还是理解为质量改进的治理过程, 将直接决定评估的制度设计与价值取向。

研究生教育质量评估的内涵与边界。综合上述辨析, 本文将研究生教育质量评估界定为: 多元利益相关者依据一定的质量标准与评价尺度, 对研究生教育的输入、过程、产出及其社会影响作出系统的事实判断与价值判断, 并据此实施质量改进与问责的制度化活动。这一界定将评估锚定于质量改进与问责的双重功能, 将多元利益相关者置于主体位置, 将输入-过程-产出-影响作为对象链条, 为后文分析

<sup>1</sup>[https://www.gov.cn/zhengce/2020-10/13/content\\_5551032.htm](https://www.gov.cn/zhengce/2020-10/13/content_5551032.htm)

划定边界。

### 1.3. 综述范围、方法与分析框架

本文的文献检索遵循中外兼顾、近十年为主、兼顾经典的原则。中文文献主要检索中国知网、万方数据与维普期刊,英文文献主要检索 Web of Science、Scopus、ERIC 与 Google Scholar。核心关键词包括:中文研究生教育质量、质量评估、质量评价、评价指标体系、质量保障、增值评价,英文 graduate education quality、quality assessment、quality evaluation、evaluation indicators、quality assurance。时间范围以 2015~2025 年为主,兼顾哈维与格林、巴奈特等经典文献及中外权威机构报告。

综述的主题边界聚焦于评估而非泛化的研究生教育研究,即重点考察围绕研究生教育质量评估所形成的理论、主体、指标、方法与实践议题;人才培养模式、课程教学改革等与评估间接相关但不属于评估本体论域的议题,不在本文系统展开。

本文的分析框架预设三重理论视角:范式视角,用以刻画质量观与评价范式的演进逻辑;主体-关系视角,用以分析多元评价格局中的角色张力与协调机制;比较视角,用以审视国际经验与中国实践之间的张力。三重视角相互交织,共同支撑一个核心论题:研究生教育质量评估研究正经历从测量到治理、从单一到多元、从结果到发展的深层范式转换,而这一转换的理论自觉程度,决定其能否为中国研究生教育的高质量发展提供有效的知识供给。

## 2. 理论基础:评价范式的演进与分野

理论并非与实践并列的独立板块,而是评估活动的观念前提:质量观决定评什么,评价理论决定怎么评,治理理论决定谁来评。因此,下文所梳理的质量观演变与范式转换,为后文主体-指标-方法-比较四个实践维度提供解释框架;反过来,实践维度中暴露的张力也不断修正着理论自身的预设。理论、主体、指标与方法之间构成一种相互塑造而非单向决定的关系,这一关系构成全文的组织线索。

### 2.1. 评价理论的谱系梳理

研究生教育质量评估的理论根基,可追溯至三条相互交织又各有旨归的理论脉络,它们分别回答了评价的对象是什么、评价如何可能与评价由谁主导三个基础性问题。

其一,系统评价理论,回答对象是什么。研究生教育是一个涉及政府、社会、高校、导师与学生等多主体、多层次、多环节的复杂系统,其质量是多因素共同作用的结果。系统评价理论要求将评估对象视为开放系统,从输入、过程、输出的全链条而非单一环节把握质量生成。这一意涵直接催生了国内普遍采用的输入-过程-输出分析框架[7],并在此后延伸出发展维度,构成贯穿本领域的方法论底色。

其二,教育评价理论,回答评价如何可能。从泰勒的目标评价、斯塔弗尔比姆的 CIPP 模型,到古贝与林肯的第四代评价,教育评价理论经历了测量-描述-判断-建构的四代演进,评价重心从技术性测量逐步转向协商建构与价值多元[11]-[13]。这为引入利益相关者视角、强调多元主体协商提供了方法论依据。第四代评价对协商建构的强调,与质量即转化的质量观存在内在呼应——二者都将评价旨归指向主体的成长与发展。

其三,公共治理理论,回答由谁主导。随着高等教育从管理走向治理,评估被重新定位为多元利益相关者共同参与的治理工具。纽维提出的评估型国家概念,揭示了政府从直接控制转向遥控引导的治理逻辑,即国家不再直接实施评估,而是通过设定目标、标准与问责框架引导高校自我治理[14]。国内学者据此指出,研究生教育质量评价主体正从政府一家转向利益相关者共同问责的多元共治格局[15]。公共治理理论为评价主体的多元化提供了合法性论证,但也提出了一个更深层的问题:当多元主体各持其价值

立场时, 评价如何避免沦为利益博弈的竞技场, 而真正服务于质量的改进? 这一追问构成后文主体论部分的逻辑起点。

## 2.2. 质量观的理论分野

质量观的分野, 是理解研究生教育质量评估一切分歧的钥匙。文献中大体可以辨识出三种相互竞争又彼此渗透的质量观, 它们分别建基于不同的知识论预设与价值逻辑。

学术导向的质量观及其知识论预设。这一质量观以高深学问的探索、发展和传承为衡量标准, 强调学术共同体内部的同行认可, 将原创性知识贡献与学术声誉作为质量的核心标志[16]。其知识论预设在于: 知识的价值内在于知识本身, 质量的判断权应归属于深谙学科规范的学术共同体。这一质量观贴近研究生教育以学术为志业的内在逻辑, 但其局限亦十分明显——它容易将质量窄化为论文发表数量、期刊等级等可量化指标, 既忽视知识对经济社会发展的外部影响, 也易在声誉的马太效应中固化既有的等级结构。

市场导向的质量观及其效率逻辑。在市场与绩效逻辑主导下, 质量被理解为目标适配与物有所值, 即研究生教育满足劳动力市场与雇主需求的程度[5]。其贡献在于引入外部问责与效率维度, 但过度扩张则可能将研究生教育工具化为个体社会经济投资, 挤压学术自由与基础研究的空间, 甚至以就业率、起薪等短期市场信号遮蔽教育的长远育人价值。

共同利益视角的质量观及其超越性。联合国教科文组织在《反思教育: 向全球共同利益的理念转变》与《一起重新构想我们的未来》中, 系统阐发了教育作为共同利益的理念, 主张超越公共品与私人品的二元框架, 将教育视为须由社会共同治理、共同享有、具有关系性的共同利益[17][18]。玛金森等进一步论证, 高等教育作为共同利益, 意味着其评价须纳入研究对经济、社会、文化乃至可持续性发展的整体贡献[19]。这一视角深刻影响了多国博士生教育质量的评价改革——评价内容新增了研究对社会、经济和文化影响的维度, 从论文数、影响因子转向对知识生产社会效应的整体关照。共同利益质量观的价值在于, 它为评价体系的综合化重构提供了规范性的价值指引: 质量不再是对某一主体偏好的迎合, 而是对共同体整体福祉的增进。三种质量观并非简单的并列或替代关系, 而构成一个具有内在张力的谱系: 学术导向固守知识生产的内在标准, 市场导向回应知识应用的效率诉求, 共同利益视角则试图在更高层面将二者统摄于社会整体福祉之下。它也提醒我们: 任何单一质量观一旦走向极端, 都会以牺牲研究生教育的某种本质为代价。

## 2.3. 评价范式的转换逻辑

在质量观转型的牵引下, 研究生教育质量评估的范式发生了两次可辨识的转换, 其内在逻辑均指向对质量如何生成这一本体论问题的重新回答。

第一次转换, 是从投入-产出范式到学习者发展视角。传统评估聚焦生源素质、师资队伍、科研经费、学位论文等投入与产出指标, 隐含着优质投入必然产生优质产出的线性假设。但越来越多研究揭示, 这一假设忽视了质量真正生成于其中的教育过程与学习体验。阿斯汀的投入-环境-产出模型与学生参与理论[20]、帕斯卡雷拉与特伦齐尼对大学教育影响的系统综述[21], 为以学生发展为中心的评价提供了证据基础。受此影响, 博士生教育质量评价转向学习者发展视角, 注重考察教育情境要素是否得到积极调动、博士生在教育情境中的参与水平以及专业知识能力素养的培养结果[22]。国内学者据此构建了包含输入、过程、输出与发展四阶段的评价模型, 将学生发展作为评价主线[23]。

第二次转换, 是广义绩效评价理论框架的提出。针对传统绩效评价重结果、轻过程与重量化、轻质性的弊端, 林梦泉等提出广义绩效评价框架, 主张纳入广义产出和成效、投入产出等要素的动态漂移以

及教育效率比的非线性现象,以更真实地刻画教育质量生成的复杂性与情境性[24]。这一框架与探索增值评价的政策导向相互呼应,标志着研究生教育质量评估正从静态、线性的绩效测量走向动态、非线性的发展性评价。

两次转换共享同一深层逻辑:它们都在质疑质量可以脱离过程而被客观测量的预设,转而强调质量的生成性、情境性与发展性。其学理意义有三。其一,实现了评估本体论基础的位移——质量不再是教育系统固有的既成属性,而是在教育过程中持续生成的动态状态,这从根本上动摇了以静态截面测量把握质量的合理性。其二,重构了评估的认识论前提——当质量是生成的、情境的,任何脱离具体情境的普适指标都只能是对质量的抽象简化,评估因此必须从测量既定对象转向理解生成过程。其三,提升了评估的价值论自觉——如何评价不再是价值中立的技术问题,而是关涉什么值得被评价、由谁决定何为值得的价值问题。正是这一范式转换的理论自觉,构成了研究生教育质量评估研究区别于一般教育测量的独特品格,也为其与公共治理、公共利益等理论传统的对话开辟了空间。

## 2.4. 公共利益视角的学理深化:内涵、构成要素与协调机制

前述质量观谱系中,公共利益被反复援引为超越学术导向与市场导向的价值尺度,但其内涵、构成与运作机制在既有综述中多被一笔带过。本节试图弥补这一缺漏,并将其转化为可用于评价实践的学理工具。

(1) 内涵辨析。公共利益不同于经济学意义上的公共品:公共品以消费的非竞争性与非排他性为核心标准,着眼于供给属性;公共利益则是规范性概念,回答什么值得被共同追求与共同维护。联合国教科文组织指出,将教育视为公共利益,意味着教育的供给、治理与收益均应被视为社会共同的参与性承诺,而非可由市场或个人单独处置的私人物品[17]。玛金森进一步区分了公共利益的两个特性:知识增长与社会福祉具有非位置性,即一方受益不以他人受损为代价;优质教育机会则具有位置性,即竞争性稀缺[19]。这一双重性决定了公共利益视角既不排斥个体的合理追求,也不容许以个体竞争替代公共福祉。由此可提炼出三重规定:关系性,即质量生成于主体间的互动而非孤立产出;参与性,即质量标准的确定须经受影响者共同参与;共享性,即质量收益须在更广泛的社会范围内可及。

(2) 构成要素。结合既有研究与各国实践,公共利益视角下的研究生教育质量可解构为四个相互嵌套的要素:一是知识要素,即原创性知识的公共积累与开放共享,对应学术共同体的公共利益;二是社会要素,即研究对经济、社会、文化与可持续发展的贡献;三是个体要素,即研究生在能力、素养与人格上的增值与全面发展;四是程序要素,即多元主体的共同参与、信息公开与相互问责。四者并非并列清单,而是以关系性为纽带、以程序性为条件的嵌套结构:若无程序要素保障,前三个实体要素的判定权便可能被单一强势主体垄断;若缺失个体要素,知识与社会要素便失去承载主体,公共利益亦将退化为抽象的集体话语。

(3) 不同主体利益的协调机制。公共利益的实现难点在于:政府、高校、导师、学生、雇主与社会公众的诉求既相互依存又彼此冲突,且在价值意义上往往不可通约——学术原创性的价值无法与就业率提升折算为同一尺度。既有研究多以责任分担、自主互信作原则性表述,对具体协调路径则语焉不详。本文认为,公共利益视角处理利益冲突的方式不是寻求实质标准的折中,而是确立程序优先的协调逻辑,具体包含三种机制:一是协商机制,通过制度化的利益相关者议事平台共同设定标准,使质量标准在形成阶段即获得多元合法性;二是问责机制,以信息公开和相互问责替代单向的上级考核,使每一主体的行为均可被其他主体观测与评价,抑制任何一方将自身偏好伪装为公共利益;三是补偿机制,对具有非位置性却难以在市场中获得回报的价值(如基础研究、人文社会科学研究与对弱势群体的教育支持)予以倾斜性资源配置,以校正纯粹竞争性机制对公共利益的侵蚀。三种机制的组合使公共利益从价值目标转

化为可操作的治理安排, 也为其向评价指标的转化提供了程序基础。

### 3. 主体论: 多元评价格局的形塑与张力

主体维度是质量观在制度层面的直接投影。学术导向的质量观倾向于将评价权归属于学术共同体, 市场导向的质量观要求引入雇主与社会中介, 共同利益视角则要求所有受影响的利益相关者均获得参与机会。因此, 评价主体格局的形塑过程, 实质上是不同质量观争夺质量界定权的制度过程; 而主体格局一旦形成, 又会反过来决定指标由谁设定、数据由谁采集、结果向谁负责, 这正是本节与后文指标、方法两章之间的内在关联。

#### 3.1. 多元主体格局的历史演进

评价主体研究回答的是谁来评价这一制度性前提, 它关涉评估权力的分配与评价合法性的来源。我国研究生教育质量多元主体的形塑过程可分为三个阶段。第一阶段为监督管理、政府独一(1981~1993年), 评估主要作为政府行政监管手段, 学位授权审核与评估由政府统一组织, 评价主体高度一元化。第二阶段为打破垄断、评价实践迅速开展(1994~2009年), 随着高等教育管理体制改革与研究生教育规模扩张, 高校自评、社会评估相继兴起, 但政府仍居主导地位。第三阶段为多元参与、教育评价内涵化发展(2010年至今), 评估从外部问责工具转向内涵式发展的治理机制, 以培养单位为主体、政府宏观管理、专门机构实施、社会多元参与的格局逐步确立[3]。

从国际经验看, 研究生教育质量保障主体大多经历了从一元控制到多元管理的变革, 政府、高校与社会相互分工、相互协调, 共同参与质量保障[25]。这一变迁轨迹被普遍视为高等教育治理结构演进的普遍规律, 其动因在于: 质量的高度复杂性与利益相关者的多元性, 使任何单一主体都难以垄断质量判断的合法性。

#### 3.2. 各评价主体的角色张力

在多元格局中, 各主体的角色定位与功能边界发生了深刻调整, 同时也显露出内在的张力结构。

政府评估职能的转变。政府职能由行政型逐步转向服务型, 从直接的行政性评估转向通过法律法规、标准制定、政策调控与结果问责实现宏观监管。《教育部 国家发展改革委 财政部关于深化研究生教育改革的意见》《国务院学位委员会 教育部关于加强学位与研究生教育质量保证和监督体系建设的意见》等文件[26][27]明确重心下移: 质量保证的重心由政府转向学位授予单位, 政府角色由包办转向监管引导[3]。这一转变的张力在于: 政府放权与监管之间的边界如何划定——放权不足则复归行政化, 监管缺位则可能放任自流。

高校自评的重构及其自主性困境。学位授予单位被确立为质量保障的第一责任主体, 其内部质量保证体系(内部评估、常态监测、自我诊断)的地位空前凸显[3]。高校自评从被动的迎评走向主动的以评促建, 质量文化、常态监测与持续改进机制成为重构的核心内容。然而, 其自主性始终面临双重困境: 一方面, 评估标准与程序高度依附于政府设定的外部框架, 自主空间有限; 另一方面, 自评在利益驱动下易流于自我美化的形式主义, 缺乏真正的问题诊断功能。

社会评估的专业化发展及其合法性建构。社会中介机构、行业部门、专业学会、第三方评估机构及新闻媒体等社会力量, 通过认证、排名、质量报告等专业化活动参与评估, 既回应社会问责需求, 也弥补政府评估的科层化局限。第三方评估被视为实现管理、评价、办学三者分离、促进数据共享与诊断改进的重要力量[3]。但其合法性建构仍面临难题: 专业性、独立性与公信力尚未得到充分确证, 评估标准的科学性与评估结果的可信度亦受质疑。

学生主体参与评价的权利与缺位。学生是研究生教育的直接受益者与质量生成的共同参与者,理论上应是评价的核心主体之一。然而大量研究揭示,当前多元主体评价形塑过程中存在学生主体边缘化与缺位的现象:学生不仅在评价指标设计、评价过程实施、评价结果反馈等环节缺乏话语权,其学习体验与发展诉求亦难以充分进入评价议程[15][28]。这一缺位被视为制约评价全面性与客观性的关键短板,也是以学生发展为中心的评价范式落地的主要障碍。从学理上看,学生主体的缺位并非偶然,而是评价权力结构失衡的必然表征——当评价被主要设计为自上而下的问责工具时,处于权力末梢的学生便天然地沦为被评价者而非评价者。

### 3.3. 利益主体关系的协调机制

多元主体的引入必然带来利益诉求的冲突与协调问题。研究的核心关切在于政府、社会与高校三者关系的再平衡:政府主导可能挤压高校自主与社会参与;市场主导可能偏离学术本真;高校自治则可能弱化外部问责。如何构建责任分担、自主互信的主体关系,成为评价制度设计的难点。这一问题在理论上即前文第2.4节所讨论的共同利益协调机制在主体层面的具体呈现:协商机制对应质量标准设定的环节,问责机制对应结果使用的环节,补偿机制对应资源配置的环节。

国外经验提供了可资借鉴的参照系。美国以市场主导型的社会模式著称,其研究生教育质量评价以认证制度为基础,依托行业协会与专业学会实施,形成政府不直接介入、以社会中介为主体的格局[29]。英国建立了以高等教育质量保障署(QAA)为核心的统一外部质量保障体系,通过资格框架、学科基准与定期评估实施监督,并以《研究生研究体验调查》等大规模学生调查倾听学生声音。荷兰则形成国家系统模式,以荷兰-弗兰德认证组织(NVAO)为核心,对学位项目进行六年一周期的外部认证与同行评议,并辅以《标准评估协议》对研究机构实施周期性评估。三种模式的共同点在于:以专业、独立的专门机构为评估实施主体,以同行评议为基本方法,以信息公开与问责为约束机制,以学生体验调查为重要补充。这为协调我国政府、高校、社会三者关系提供了制度参照,但其移植须直面前提条件:域外模式的有效性端赖于成熟的法治传统、专业化的评估机构与独立的社会问责文化,而这些条件在我国尚处于发育之中。

## 4. 指标体系论:从单一到多元的建构逻辑

如果说主体维度回答谁来评,指标维度则回答评什么。指标是质量观的技术化表达:任何一种质量观若不能转化为可观测、可比较的指标,便无法进入实际操作;反之,指标一经确立,又会通过资源配置与声誉机制塑造培养单位的行为,从而在事实上重新定义什么是好的研究生教育。这一观念-指标-行为的循环,是理解指标体系之争的关键。

### 4.1. 指标体系构建的方法论基础

指标体系是评估活动的技术核心,其构建质量直接决定评估的科学性与公信力。既有研究在指标设计原则上形成了基本共识:导向性与代表性原则(指标应体现评估目的与价值导向,选取典型、有代表性的要素)、客观性与有效性原则(指标应独立、可测、有效,保证信度与效度)、可测性与易行性原则(指标应便于测量与获取,兼顾成本与执行可行性)[30]。同时,分类分层、差异化设计日益受到强调——针对学术学位与专业学位、不同学科门类设置差异化指标,以破解一把尺子难以引导分类发展的难题[9]。

在指标权重确定上,层次分析法与模糊综合评价方法应用最广。前者通过构建目标层-准则层-指标层的递阶结构,结合专家判断确定权重,适用于多目标、多准则的复杂决策;后者借助模糊数学处理评价中的不确定性与主观判断问题。二者常结合使用,以兼顾权重的科学性与评价的容错性。然而,此

类方法的技术精致性并不能掩盖其内在局限：权重的确定高度依赖专家主观判断，而专家结构一旦偏向某一利益群体，便可能使科学的权重沦为特定价值立场的量化包装。这一局限提示我们，指标体系的建构本质上是一个价值选择问题，而非纯粹的技术问题。

#### 4.2. 核心评价指标的多维建构

既有文献识别出的评价指标维度可归纳为若干共识性类别。在国际文献中，教学因素、管理因素、影响评估被概括为研究生项目学术质量评估的三大核心类别；导师-学生关系、学校资产、个人学习状态、学术成就、精神状态等维度亦被纳入考量。巴基斯坦高等教育委员会依据《研究生教育政策(2023)》与项目效能与提升审查标准，于2025年启动全国性研究生项目质量审查，聚焦治理与管理、监督质量、研究诚信与伦理、学生支持与申诉、研究质量与文化、项目基础设施、学习与评估、外联与专业整合等八大维度[31]-[33]，为发展中国家研究生质量评估提供了较为系统的制度样本。

在国内，博士生教育质量评价要素被概括为三个层面：教育机会层面的公正招生选拔机制、教育过程与结果方面的综合素养培养、知识生产及其后续应用和传播。这一框架贯通了机会公平-过程质量-结果影响的完整链条，与共同利益质量观形成呼应。值得注意的是，随着学习者发展视角的引入，导师指导质量、导学关系、课程教学体验、学术交往、奖助支持等过程性、体验性指标的权重显著上升，打破了以论文、奖项为主的结果性指标的一统格局。这一变化的意义不止于技术层面的指标增删，而在于评价重心的本体论位移——从教育提供了什么转向学生经历了什么、发展了什么。

#### 4.3. 评价体系综合化重构的实践进展

单一量化指标导向的评价模式，是当前中国研究生教育评价体系改革所面对的核心问题。传统评价过度依赖论文数量、期刊等级、获奖层次、项目级别与学历头衔等外部可计量标志，导致重结果轻过程、重量化轻质性、重科研轻育人的偏差，诱发了学术功利化、创新能力弱化等问题。从学理上看，传统单一评价模式的理论困境在于：它以可量化的代理指标替代了难以量化的质量本身，从而陷入以指标测度质量、又以指标异化质量的循环。

对此，改革的基本方向是纠正以量代质的偏差，并同步重建实质性的质量标准。一方面，突破传统单一评价模式：部分高校率先开展评价改革试点，不再以论文发表数量作为博士生学位授予的硬性门槛，代之以创新成果的实质贡献评价，将学术评价权收归学术共同体，并以代表作、同行评议为核心机制。另一方面，建立综合化、多元化评价机制：将基础研究、应用研究、成果转化等多元成果形态纳入评价，坚持分类评价、过程评价与增值评价，以确保评价的公平性、准确性与客观性[34]。从制度逻辑看，这一改革的深层指向并非否定量化指标本身，而是纠正以量代质的异化，使评价回归育人本质与质量贡献本位。其未竟之问在于：在旧的量化标准被削弱之后，新建立的标准能否真正度量创新贡献，而非以新的外部标志取代旧的外部标志。

#### 4.4. 共同利益视角下评价指标框架的示例性建构

前文第2.4节将共同利益质量观解构为知识、社会、个体与程序四个要素，并提出了协商、问责与补偿三种协调机制。为使这一视角具备落地能力，本节尝试将其转化为一个包含具体可操作指标的示例性评价框架。框架的建构遵循四条原则：要素对应原则，即指标体系须覆盖四个构成要素，不得遗漏程序要素；过程与结果并重原则，即除成果性指标外，纳入学习者发展、导学互动等过程性证据；分类分层原则，即同一维度下区分学术学位与专业学位、基础学科与应用学科的可比范围；可操作原则，即每一项指标均须有明确的数据来源与采集口径，避免以抽象概念充当指标。依此原则，框架可归纳为表1。

**Table 1.** An illustrative indicator framework for graduate education quality assessment from the common-good perspective  
**表 1.** 共同利益视角下研究生教育质量评价的示例性指标框架

| 维度                     | 观测要点               | 可落地的指标                                          |
|------------------------|--------------------|-------------------------------------------------|
| 知识贡献<br>(学术共同利益)       | 原创性知识的公共积累与开放共享    | 代表作同行评议等级；开放获取成果占比；跨机构、跨学科合作成果占比；学位论文创新性评价结果    |
| 社会影响<br>(经济、社会与文化共同利益) | 研究对经济社会文化及可持续发展的贡献 | 成果被政策文件、行业标准采纳数；产学研合作项目数与转化收益；社会服务与知识普及工作量      |
| 个体发展<br>(学习者共同利益)      | 研究生能力与素养的增值        | 入学－毕业能力素养增值分数(增值评价)；导师指导满意度；课程与学术交往体验；学位按期完成率   |
| 治理与程序<br>(关系性共同利益)     | 多元主体共同参与与信息公开      | 利益相关者参与指标制定的制度化程度；评估报告公开率；申诉与反馈闭环率；自评与外部评议结论一致性 |
| 伦理与包容<br>(规范性共同利益)     | 科研诚信和选拔公平          | 学术不端查处与整改完成率；奖助覆盖率与资助公平指数；来自弱势群体研究生的比例及毕业率      |

注：本表为学理示例，用于说明共同利益质量观向可操作指标转化的路径，不构成对具体培养单位或学位点的评价结论。表中各项指标须依据学科与学位类型分类分层设定基准值。

上述框架在方法论上有三点意义。其一，它把共同利益从价值宣示转化为可观测的指标集合，回应了共同利益视角缺乏操作性的批评。其二，它以程序要素与补偿机制为制度支点，试图矫正单纯依赖结果性指标的激励偏差——奖助覆盖与资助公平类、评估报告公开类指标，其功能并非度量产出，而是约束评价过程本身。其三，它与增值评价方法形成互补：个体要素中的增值分数需借助增值评价模型估计，而增值评价亦须依托此类框架，方能避免被简化为单一分数排名。

5. 方法论：评价工具的开发与应用

指标确定之后，方法问题便凸显出来。方法并非中立的容器：层次分析法与模糊综合评价隐含专家共识优先的价值预设，大数据监测隐含可数据化者优先的价值预设，增值评价则隐含净增长优先的价值预设。因此，方法的选择本身就是价值选择的一部分；方法与指标之间构成相互约束的关系——指标体系规定了方法的适用边界，方法的技术特性又反向筛选了哪些指标能够被真正测量。这一认识构成本节讨论方法问题的基本立场。

5.1. 评价方法的类型学分析

研究生教育质量评估的方法体系，可沿三条轴线加以类型化。其一，自我评价法与外部同行评价法的结合：自我评价是培养单位基于自身目标对质量状态进行诊断的起点，外部同行评价则以专家独立判断提供校准与问责。二者结合构成现代质量保障的基本架构——荷兰的认证、英国的院校审计、巴基斯坦的项目审查，均遵循自评报告－外部专家评议－评估结论－反馈改进的通用流程[35]。自我评价的发展性与外部评价的问责性之间的张力与平衡，是方法研究持续关注的议题。其二，定量方法与定性方法的整合：定量方法提供可比较、可问责的客观证据，定性方法捕捉难以量化的质量维度。英国研究卓越框架的实践表明，同行评议应居于主导地位，文献计量仅作辅助，且须对定量指标的使用作出解释，以防滥用[36]。这种定性主导、定量辅助的整合思路，对纠正国内评价中过度依赖量化指标的倾向具有借鉴意义。其三，比较教育方法的应用：跨国比较被广泛用于评价模型的借鉴与本土化检验，其关键在于识别制度情境差异，避免机械移植。

## 5.2. 评价工具的开发与检验

评价工具的科学化是评估效度的重要保障。在问卷工具层面, 预试问卷的编制与检验形成了较为成熟的技术流程, 包括项目分析、效度检验与信度分析; 大规模结构化调查成为倾听学生声音、服务诊断改进的重要工具[23]。

在国际层面, CALOHEE 模型是较具代表性的创新。该模型基于 Tuning 教育结构项目, 以透明可靠指标在学习成果层面评估质量与相关性, 构建了从副学士到博士(第三周期)的资格参考框架与分学科学习成果框架, 使跨国诊断性评估成为可能[37][38]。其方法论启示在于: 将质量评估的锚点从投入与声誉转向可观测的学习成果, 并以共同的资格框架支撑跨国比较。这一转向对中国评估改革的镜鉴在于, 它提供了一种超越单一论文导向的成果观——以明确界定的学习成果作为质量的核心证据。

## 5.3. 数字化赋能评价的前沿探索

数智技术正在深刻重塑研究生教育质量评估的方法生态, 形成三个可辨识的新方向。其一, 基于大数据的学科评价跟踪体系: 依托研究生教育信息数据库与数据治理, 构建全员、全程、全方位的常态监测体系, 生成研究生、导师与学位点画像, 实现从经验判断向数据驱动的转型[24]。其二, 数据密集型评价新机制: 评价日益依赖多源异构数据的采集、融合与挖掘, 以接近研究生培养的真实状态, 并支撑动态监测、实时反馈与预警干预, 推动评价从事后评价向事中干预、事前预警延伸。其三, 人智混合评价新模式: 在强调人工智能赋能评价科学性、精准性的同时, 警惕算法偏见与技术替代人的倾向, 主张以专家判断与机器智能的协同保证评价的公平、透明与包容[28]。值得注意的是, 数智技术的引入在提升评价效率的同时, 也潜藏着技术决定论的风险——若数据之客观遮蔽了价值之判断, 评价反而可能异化为更精致的工具理性。

## 5.4. 增值评价的统计模型与应用挑战

《深化新时代教育评价改革总体方案》明确提出探索增值评价[4], 使增值评价成为研究生教育评估研究的前沿议题。然而在既有综述中, 增值评价多以政策语汇出现, 其在教育统计学上的模型构造与应用条件尚未得到充分讨论。本节尝试补充这一缺环。

(1) 基本内涵。增值评价以净增长为核心, 试图估计学生在特定教育经历中获得的、可归因于该经历的增量, 而非其既有的绝对水平。其关键特征在于参照点的转换: 不再以同伴横截面比较为基准, 而以学习者自身的起点为基准, 考察其在一定时期内的进步幅度。这使其在价值取向上与学习者发展范式高度契合, 也使其成为判断培养单位教育贡献而非生源优势的潜在工具。

(2) 模型谱系。在教育统计学中, 增值估计已发展出若干成熟的模型族, 其差异主要在于对测量误差与数据嵌套结构的处理方式。其一, 协变量调整模型(又称回归残差法): 以先期成绩、生源特征等协变量预测后续表现, 以实际值与预测值之差作为增值估计, 技术门槛较低, 但对协变量遗漏较为敏感[39]。其二, 增益分数模型: 直接以后测成绩减去前测成绩, 操作直观, 但测量误差在差分过程中被放大, 估计信度较低。其三, 多层线性模型(HLM): 将学生、导师、学科、院校设为嵌套层级, 以随机截距与随机斜率刻画不同层级对增值的贡献, 是目前增值估计最主流的框架[40]。其四, 交叉分类随机效应模型: 针对研究生同时归属多个情境(学科、导师、课题、项目)的非纯层级结构, 可同时估计多个情境的效应, 避免强制分层的偏误。其五, 学生成长百分位模型(SGP): 以历史同辈的条件分布为参照, 给出学生在相似起点群体中的相对成长位置, 兼具绝对增长与相对位置的双重信息[41]。其六, 潜变量增长模型: 通过多个时点的观测构建潜在增长轨迹, 可在一定程度上缓解观测变量信度衰减的问题[40]。

(3) 应用挑战。增值评价在研究生教育中的推行面临统计与制度层面的多重挑战。第一, 测量工具问

题：研究生教育缺少统一、纵向可比的标准化测验，能力增长只能依赖自陈量表或成果评级观测，其纵向可比性与测量等值性难以保证。第二，估计稳定性问题：增值分数年度波动较大，单一期次的估计不足以支撑高风险决策，其估计精度与模型选择高度相关[42]。第三，选择偏误问题：学生进入特定项目的自选择过程，容易使增值被误读为培养单位的因果效应。第四，参照系的不可通约问题：不同学科、学位类型的天然增长幅度差异显著，跨学科直接比较缺乏合理性。第五，数据缺失问题：纵向追踪中的失访往往系统性偏向特定群体，进而扭曲增值估计。第六，天花板与地板效应：在高起点或低起点群体中，可观测的增值空间被人为压缩。第七，伦理与制度风险：若将增值直接用于排名或问责，可能诱起重可测、轻不可测的新一轮指标化博弈，以更容易提升的维度替代真正重要的育人维度。

(4) 应用定位。基于上述挑战，增值评价在研究生教育中的应用宜采取三条策略：一是定位为诊断与改进工具，而非排名与问责工具，以避免激励扭曲；二是采用增值加多主体证据的组合，将增值分数与同行评议、学习体验调查结果联合解释，而非单独作为结论；三是明确其适用的比较范围，原则上仅在同一学科、同一学位类型内部进行比较。唯有统计条件与制度条件同时具备，增值评价才能实现其以学习者发展为中心的初衷。

## 6. 比较论：国际经验与中国实践

前几章的分析均以中国语境为隐含背景，而中国研究生教育评估所面对的问题，很大程度上是在一个高度压缩的时间跨度内，同时遭遇了先行国家在数十年中陆续面对的问题。因此，国际比较的意义不在于移植具体制度，而在于识别何种制度设计回应了何种问题。本章先描述主要模式，再归纳共同趋势，进而回到中国语境，最后总结各国经验对中国具体问题的启示。

### 6.1. 主要发达国家的评价模式

英国：从科研评估到研究卓越框架的演进。英国研究卓越框架(REF)的前身为科研评估(RAE)。REF 2014 将评估维度调整为科研成果(65%)、科研影响(20%)、科研环境(15%)，首次将研究影响——研究成果对经济、社会、文化、公共政策与生活质量产生的可验证变化——纳入核心指标，并建立四星级质量评级[36]。REF 2029 拟进一步重构评估维度，凸显对研究文化与整体贡献的重视。其演进集中体现了研究影响定义的扩展——从学术影响力扩展至社会影响力，为研究生教育评估提供了结果导向、影响导向的参照。此外，英国《研究生研究体验调查》通过大规模学生调查，将博士生学习体验纳入质量保障，体现了对学生主体声音的重视。

欧洲：CALOHEE 项目的跨国诊断评估。如前所述，CALOHEE 通过构建跨国的资格参考框架与学习成果框架，在超越国别的层面诊断学位项目质量，代表了欧盟在博洛尼亚进程框架下推动跨国质量比较的前沿探索[37]。

美国：认证与市场主导模式。美国以区域性与专业性认证为质量底线，以市场声誉与学科排名为质量信号，政府不直接实施评估，形成认证保底线、市场定声誉的格局[29]。

巴基斯坦：HEC 的全国性项目审查。作为发展中国家的制度样本，巴基斯坦高等教育委员会于 2025 年启动全国性研究生项目质量审查，依据《研究生教育政策(2023)》与项目效能与提升审查标准，聚焦治理与管理、监督质量、研究诚信与伦理、学生支持与申诉、研究质量与文化、项目基础设施、学习与评估、外联与专业整合等维度[31]-[33]，为后发国家在资源约束下构建系统性研究生质量评估提供了经验。

### 6.2. 国际评价实践的共同趋势

综合各国实践，可以辨识出三条共同趋势。其一，从输入导向到结果导向的转变：评价重心由生源、

师资、经费等投入条件, 转向学习成果、研究产出与社会影响等结果与成效指标, REF 的影响维度、CALOHEE 的学习成果框架均是其体现。其二, 从单一主体到多元主体的转变: 政府、专门机构、高校、社会中介与学生的多元参与成为普遍制度安排, 学生调查、利益相关者协商日益纳入评价流程。其三, 从国内标准到跨国比较的拓展: 在高等教育国际化背景下, 资格框架、学习成果框架的跨国对接使质量评估从封闭的国内标准走向可比较的跨国基准。

### 6.3. 中国评价实践的特殊语境与问题

中国研究生教育质量评估在借鉴国际经验的同时, 形成了鲜明的本土特色, 也面临特定的实践张力。就特色而言, 中国模式兼具公共政策导向与绩效主义的双重底色: 一方面, 评估服务于国家人才战略的公共目标, 政府通过学科评估、学位授权审核、论文抽检等实施有力的宏观调控; 另一方面, 绩效主义与量化导向又深刻渗透于院校与个人的评价实践。双重底色之间并非和谐共存, 而潜藏着深刻的张力: 公共政策导向要求评估服务于育人的根本目标与国家战略的人才需要, 绩效导向却倾向于以可量化的产出指标来测度与激励; 当二者在以何种标准衡量质量上发生抵牾时, 评估便容易在价值导向与绩效导向之间摇摆, 甚至以绩效的可计算性替代价值的不可通约性。这一张力构成本土评估困境的深层根源, 也提示中国评估理论的建构必须正视并回应公共政策导向与绩效导向的结构性矛盾。

就问题而言, 学界普遍指出的症结包括: 其一, 评价实践中的形式化问题——迎评、包装、指标性应付与集中式评估并存, 评估的诊断改进功能弱化; 其二, 本土化评价理论的建构不足——大量研究移植西方评价理论与方法, 对中国研究生教育的制度情境、文化传统与治理结构关照不够, 尚未形成具有解释力与指导力的本土评价理论; 其三, 学生主体缺位、过程性评价薄弱等前述问题在此语境下被进一步放大。因此, 建构既接轨国际前沿又立足中国语境的本土化评估理论, 成为学术共同体的普遍呼声。

### 6.4. 国际经验对中国评价改革的启示

若将上述国际经验与中国当前的突出问题相对照, 可以提炼出两类具有针对性的启示: 一类回应学生主体缺位, 一类回应指标功利化。

(1) 针对学生主体缺位的启示。其一, 将学生声音制度化为常规评价证据。英国《研究生研究体验调查》的价值不仅在于采集体验数据, 更在于其周期性、独立性与公开性, 使调查结果成为质量保障流程中的常规输入; 其启示在于, 中国的学生体验调查应避免成为一次性、突击式的信息收集, 而应建立稳定的调查制度, 并使结果进入质量报告与反馈闭环。其二, 把学生从被调查者提升为共同评价者。欧洲高等教育区《质量保障准则与指南》明确要求学生在内部与外部质量保障的各环节均应有代表参与[43], 荷兰的同行评议亦吸纳学生委员; 相比之下, 中国的学生参与多停留在被调查这一最低层级。可考虑设置被调查者-共同评价者-共同决策者的三级参与阶梯, 并在指标设定与结果审议环节率先引入学生代表。其三, 以制度化的申诉与反馈机制保障学生表达的实际效力。仅有表达渠道而不设回应义务, 学生参与仍会流于形式。

(2) 针对指标功利化的启示。其一, 以质性判断约束量化指标。REF 确立的同行评议主导、文献计量辅助原则, 以及要求对定量指标的使用作出强制性解释的做法[36], 为抑制指标滥用提供了可操作的制度设计; 凡使用量化指标之处均须说明其局限。其二, 以学习成果替代产出数量作为评价锚点。CALOHEE 以明确界定的学习成果作为质量证据, 为论文之外的质量提供了正当的呈现形式[37]; 可通过分学科的学习成果框架, 把难以量化的育人成效转化为可讨论、可比较的证据。其三, 在资源约束条件下以程序规范弥补资源不足。巴基斯坦 HEC 通过标准化的项目审查流程实现系统性质量信息的持续积累[31]-[33], 说明后发情境下程序先行是一条可行路径。

综合来看, 国际经验对中国改革的启示可归结为一条原则性判断: 制度移植的前提是制度条件, 而制度条件的成熟需要制度实践的累积。因此可行的路径是程序先行——优先实现评价程序的规范化、公开化与参与化, 再在持续的实践反馈中逐步调整实质标准。这一判断亦与前文强调的程序优先协调逻辑相互印证。

## 7. 结论: 研究述评与未来展望

### 7.1. 已有研究的学术贡献

综观既有文献, 研究生教育质量评估研究在三个层面取得了显著进展。其一, 理论框架的丰富与深化: 从质量观的多维界定(学术、市场、共同利益)到评价范式的两次转换, 研究的理论自觉不断增强, 为评估实践提供了日益精细的分析工具。其二, 方法工具的多元化发展: 层次分析法、模糊综合评价、问卷工具开发、大数据监测、学习成果框架、增值评价模型等方法的引入, 使评估从单一的定性判断走向定量与定性融合、静态与动态结合的多元方法体系。其三, 国际比较视野的拓展: 对英、美、荷、欧及巴基斯坦等国模式的系统比较, 为中国评估改革提供了丰富的制度参照, 也推动了结果导向、多元主体、跨国比较等共识的形成。

### 7.2. 现有研究的理论困境

在肯定贡献的同时, 既有研究仍存在若干理论困境。其一, 学生主体视角的缺失: 尽管理论上承认学生是核心利益相关者, 但以学生为中心、以学习与发展成果为锚点的实证研究仍相对薄弱, 学生声音在评估制度设计中的制度化表达不足。其二, 评价指标的本土化不足: 指标体系构建多移植国际框架, 对中国研究生教育的特殊语境(如导师制、学位类型结构、育人要求)关照不够, 本土适切性与可操作性有待提升。其三, 跨学科、跨文化的适用性问题: 现有指标与工具多基于单一学科或西方情境开发, 其在跨学科研究、跨文化情境中的信度与效度缺乏充分检验。更深层地看, 三重困境共享一个理论根源: 研究生教育质量评估研究尚未完成从方法移植到理论原创的跨越, 其知识生产仍以域外理论为参照系, 缺乏基于中国实践的理论建构自觉。

### 7.3. 理论解释框架: 三重逻辑的综合

既有研究多将上述困境归因于技术手段不足或数据基础薄弱, 本文认为, 这一归因未能触及根本。研究生教育质量评估之所以长期在测不准与改不动之间徘徊, 其深层根源在于它同时受到三种制度逻辑的牵引, 而三者之间缺乏有效的转换与耦合机制。

第一种是科层制逻辑。作为现代大型组织, 大学必须依靠统一的标准、可控的程序与可追溯的责任完成治理与资源配置。这一逻辑使评价成为可管理的行政技术, 同时也内生地偏好标准化与可计算化的指标。指标功利化并非某个机构的偏好, 而是科层制逻辑在评价领域的投影[44]。

第二种是学术共同体的专业逻辑。它以知识生产的内在规范为基础, 强调同行认可与不可替代性, 主张质量判断权归属于专业共同体。专业逻辑拒绝以外部标准替代内在标准, 因而与科层制的可计算性要求存在天然张力; 当专业判断被量化标准挤压时, 学术评价便让位于指标达标[44] [45]。

第三种是人才培养的个性化逻辑。教育过程的本质是个体化的: 每名研究生的知识基础、研究兴趣与成长节奏各不相同, 其质量表现为差异化的增值。个性化逻辑要求评价承认发展的多样性与不可比性, 这恰恰与科层制逻辑的可比性预设直接冲突。

三种逻辑的冲突具有结构性: 科层制逻辑追求可比较, 专业逻辑追求可判断, 个性化逻辑追求可发展。任何以单一逻辑统摄其余两者的方案, 都会以牺牲某一维度为代价——以科层制逻辑统摄, 评价将

退化为指标考核, 育人过程被简化为数据填报; 以专业逻辑统摄, 评价可能封闭于同行共识, 丧失公共问责; 以个性化逻辑统摄, 虽尊重个体差异, 却难以形成可供公共决策使用的可比信息。由此可给出一个更具统合性的解释: 前文所述困境均是三重逻辑失衡的不同表征——学生主体缺位, 是科层制逻辑遮蔽了个性化逻辑, 使学习者从评价主体退化为数据来源; 指标功利化, 是科层制逻辑挤压了专业逻辑, 使学术判断让位于指标达标; 本土化不足, 则是三种逻辑在中国制度情境中的组合方式尚未被理论化。

据此, 未来的理论建构不应再纠缠于哪种逻辑更为正确的规范性争论, 而应转向研究三种逻辑的转换条件与耦合机制。至少有三个问题值得优先探讨: 其一, 情境化分工, 即何种议题适用何种逻辑——学位授予的学术判断以专业逻辑为主导, 资源配置的效率考量以科层制逻辑为主导, 学习者发展的评价以个性化逻辑为主导; 其二, 程序性耦合, 即如何设计使三种逻辑相互约束而非相互替代的评价程序, 如以专业评议校正量化偏差、以学生参与约束科层决策、以公开问责约束专业封闭; 其三, 证据性转译, 即如何将不可通约的价值转化为可对话的证据形式, 如以学习成果框架转译个性化发展, 以成本效益分析转译公共投入的正当性。这一框架把评估困境从技术改进问题重新界定为逻辑协调问题, 为本领域提供了超越归纳困境的理论支点。

#### 7.4. 未来研究的可能进路

面向未来, 研究生教育质量评估研究可在以下方向深化。其一, 数智技术与教育评价的深度融合: 数智技术已成为研究生教育高质量发展的新抓手, 也是评价改革的战略突破口, 未来研究应着力探索大数据、人工智能赋能评价的科学路径与伦理边界, 构建人智混合的智能评价范式。其二, 共同利益视角下的评价理论建构: 在评价体系综合化重构的关键节点, 以共同利益质量观为价值根基, 建构兼顾学术价值、社会影响与个体发展的本土化评价理论。其三, 增值评价与学习者发展评价的深化: 增值评价在研究生教育中的理论内涵、统计模型、指标体系与实现路径仍有待系统研究。其四, 跨国比较研究的拓展与深化: 跨国比较应在方法上更加严谨, 深入制度机理与文化逻辑的深层分析。

综言之, 研究生教育质量评估研究正处于理论转型与方法革新的关键期。构建既立足中国国情、又回应共同利益价值关切、并善用数智技术的评价体系, 既是学术研究的重要课题, 也是高层次人才质量保障的制度基石。这一任务的完成, 有赖于学术共同体在理论原创、方法创新与本土关怀之间的持续平衡与不懈求索。

#### 基金项目

本文受到重庆理工大学研究生教育改革研究项目(项目号: gzljg2023310): 基于 OBE 理念的专业学位研究生培养质量评估及提升路径研究——以重庆理工大学为例、重庆理工大学本科教育教学改革研究项目(项目号: 2023YB152)、人口老龄化背景下重庆技术技能人才有效供给的路径研究(22SKGH567)、乡村振兴背景下重庆市职业教育与乡村工匠培育的耦合发展及路径研究(K24YG3130394)、职业学校三全育人质量提升研究(2022ZJXH431005)等项目的支持。

#### 参考文献

- [1] 迈克尔·吉本斯, 等. 知识生产的新模式: 当代社会科学与研究的动力学[M]. 陈洪捷, 沈文钦, 等, 译. 北京: 北京大学出版社, 2011.
- [2] 2023 年全国教育事业发展统计公报[EB/OL]. [http://www.moe.gov.cn/jyb\\_sjzl/sjzl\\_fztjgb/202410/t20241024\\_1159002.html](http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202410/t20241024_1159002.html), 2024-10-24.
- [3] 王战军. 构建研究生教育质量保障体系——理念、框架、内容[J]. 研究生教育研究, 2015(1): 1-5.
- [4] 中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL].

- [https://www.gov.cn/zhengce/2020-10/13/content\\_5551032.htm](https://www.gov.cn/zhengce/2020-10/13/content_5551032.htm), 2020-10-13.
- [5] Harvey, L. and Green, D. (1993) Defining Quality. *Assessment & Evaluation in Higher Education*, **18**, 9-34. <https://doi.org/10.1080/0260293930180102>
  - [6] Barnett, R. (1992) Improving Higher Education: Total Quality Care. Open University Press, 45-52.
  - [7] 王亚杰, 王沛民. 研究生教育质量: 宏观层面及其发展战略的思考[J]. 高等教育研究, 2004(5): 58-63.
  - [8] 杨颀, 陈学飞. 研究生教育质量: 内涵与探索[M]. 上海: 上海交通大学出版社, 2007.
  - [9] 盛明科, 唐检云. 研究生教育质量评价指标体系设计的框架[J]. 学位与研究生教育, 2007(7): 13-17.
  - [10] 《中国学位与研究生教育发展战略报告》编写组. 中国学位与研究生教育发展战略报告(2002-2010) (征求意见稿) [J]. 学位与研究生教育, 2002(6): 5.
  - [11] Tyler, R.W. (1949) Basic Principles of Curriculum and Instruction. University of Chicago Press.
  - [12] Stufflebeam, D.L. and Shinkfield, A.J. (2007) Evaluation Theory, Models, and Applications. Jossey-Bass.
  - [13] Guba, E.G. and Lincoln, Y.S. (1989) Fourth Generation Evaluation. N Sage.
  - [14] Neave, G. (1998) The Evaluative State Reconsidered. *European Journal of Education*, **33**, 265-284.
  - [15] 陈叶军. 研究生教育: 从管理到治理[N]. 中国社会科学报, 2014-03-28(A02).
  - [16] 弗莱克斯纳. 现代大学论——美英德大学研究[M]. 徐辉, 陈晓菲, 译. 杭州: 浙江教育出版社, 2001.
  - [17] UNESCO (2015) Rethinking Education: Towards a Global Common Good? UNESCO.
  - [18] UNESCO (2021) Reimagining Our Futures Together: A New Social Contract for Education. UNESCO.
  - [19] Marginson, S. (2016) Higher Education and the Common Good. Melbourne University Publishing.
  - [20] Astin, A.W. (1993) What Matters in College? Four Critical Years Revisited. Jossey-Bass.
  - [21] Pascarella, E.T. and Terenzini, P.T. (2005) How College Affects Students: A Third Decade of Research (Vol. 2). Jossey-Bass.
  - [22] 罗英姿, 陈小满, 张佳乐. 博士生教育质量评价模型构建与实证检验——基于学生发展相关理论[J]. 高等工程教育研究, 2019(2): 171-177.
  - [23] 许丹东, 沈文钦, 翟月, 陈洪捷. 中国博士生的培养现状与问题——基于 2021 年全国博士毕业生离校反馈调查的分析[J]. 学位与研究生教育, 2022(5): 73-80.
  - [24] 林梦泉, 等. 研究生教育水平与绩效评价理论探索及应用[M]. 北京: 高等教育出版社, 2023.
  - [25] 王战军. 构建质量保障体系提高研究生教育质量[J]. 学位与研究生教育, 2011(1): 3-6, 82.
  - [26] 教育部 国家发展改革委 财政部关于深化研究生教育改革的意见[EB/OL]. [http://www.moe.gov.cn/srcsite/A22/s7065/201304/t20130419\\_154118.html](http://www.moe.gov.cn/srcsite/A22/s7065/201304/t20130419_154118.html), 2013-03-29.
  - [27] 国务院学位委员会 教育部关于加强学位与研究生教育质量保证和监督体系建设的意见[EB/OL]. [http://www.moe.gov.cn/srcsite/A22/s7065/201402/t20140212\\_165554.html](http://www.moe.gov.cn/srcsite/A22/s7065/201402/t20140212_165554.html), 2014-01-29.
  - [28] 张薇薇. 研究生教育质量评价的范式变革与实现路径[J]. 教育创新与实践, 2025, 1(9): 53-56.
  - [29] 潘武玲. 美国研究生教育质量评价中的三种主要力量[J]. 现代教育科学(高教研究), 2006(5): 38-40.
  - [30] 王战军, 李明磊. 中国高水平研究生院评价: 特性、指标和方法[J]. 清华大学教育研究, 2021, 42(2): 33-40+68.
  - [31] Higher Education Commission (2023) Graduate Education Policy 2023. HEC.
  - [32] Higher Education Commission (2023) Pakistan Standards and Guidelines for Quality Assurance in Higher Education (PSG-2023). HEC.
  - [33] Higher Education Commission (2025) HEC Launches Nationwide Quality Review of MS/MPhil & PhD Programmes.
  - [34] 教育部 国家发展改革委 财政部关于加快新时代研究生教育发展的意见[EB/OL]. [http://www.moe.gov.cn/srcsite/A22/s7065/202009/t20200921\\_489271.html](http://www.moe.gov.cn/srcsite/A22/s7065/202009/t20200921_489271.html), 2020-09-04.
  - [35] Brennan, J. and Shah, T. (2000) Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change. Open University Press.
  - [36] Stern, N. (2016) Building on Success and Learning from Experience: An Independent Review of the Research Excellence Framework. Department for Business, Energy & Industrial Strategy.
  - [37] Wagenaar, R. (2024) Towards a Robust Approach for Evidencing Quality in Higher Education Learning: The New CALOHEE Model. *Journal of International Cooperation in Education*, **26**, 67-84. <https://doi.org/10.1108/jice-11-2023-0038>

- 
- [38] González, J. and Wagenaar, R. (2003) Tuning Educational Structures in Europe: Final Report—Pilot Project Phase 1. University of Deusto.
  - [39] Sanders, W.L. and Horn, S.P. (1994) The Tennessee Value-Added Assessment System (TVAAS): Mixed-Model Methodology in Educational Assessment. *Journal of Personnel Evaluation in Education*, **8**, 299-311. <https://doi.org/10.1007/bf00973726>
  - [40] Raudenbush, S.W. and Bryk, A.S. (2002) Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Edition, Sage Publications.
  - [41] Betebenner, D. (2009) Norm- and Criterion-Referenced Student Growth. *Educational Measurement: Issues and Practice*, **28**, 42-51. <https://doi.org/10.1111/j.1745-3992.2009.00161.x>
  - [42] McCaffrey, D.F., Lockwood, J.R., Koretz, D.M. and Hamilton, L.S. (2003) Evaluating Value-Added Models for Teacher Accountability. RAND Corporation.
  - [43] EURASHE (2015) Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). Brussels.
  - [44] Clark, B.R. (1983) The Higher Education System: Academic Organization in Cross-National Perspective. University of California Press.
  - [45] Mintzberg, H. (1979) The Structuring of Organizations: A Synthesis of the Research. Prentice-Hall.