

学生学习成果评估的生态研究

——高等教育质量保障新路径

骆美 著

汕头大学出版社

湖北省教育科学规划课题“基于学生学习效能提升的生态化教学策略研究”（项目编号：2016GB013）成果
中央高校教育教学改革专项课题“教学水平评价改革与教学质量保障体系构建”（项目编号：YB2017035）阶段性成果

学生学习成果评估的生态研究

——高等教育质量保障新路径

骆美 著

汕头大学出版社

图书在版编目 (C I P) 数据

学生学习成果评估的生态研究: 高等教育质量保障
新路径 / 骆美著. -- 汕头: 汕头大学出版社, 2018.12
ISBN 978-7-5658-3790-6

I . ①学… II . ①骆… III . ①高等教育—教育质量—
研究—中国 IV . ① G649.21

中国版本图书馆 CIP 数据核字 (2019) 第 029105 号

学生学习成果评估的生态研究——高等教育质量保障新路径
XUESHENG XUEXI CHENGGUO PINGGU DE SHENGTAI YANJIU
——GAODENG JIAOYU ZHILIANG BAOZHANG XINLUJING

著 者: 骆 美

责任编辑: 邹 峰

责任技编: 黄东生

封面设计: 汤 丽

出版发行: 汕头大学出版社

广东省汕头市大学路 243 号汕头大学校园内 邮政编码: 515063

电 话: 0754-82904613

印 刷: 北京虎彩文化传播有限公司

开 本: 787mm × 1092mm 1/16

印 张: 12.25

字 数: 145 千字

版 次: 2018 年 12 月第 1 版

印 次: 2019 年 3 月第 1 次印刷

定 价 48.00 元

ISBN 978-7-5658-3790-6

版权所有, 翻版必究

如发现印装质量问题, 请与承印厂联系退换

目 录

第一章 导 论.....	1
一、我国高等教育质量保障概述	2
（一）高等教育质量保障的历程与阶段性进展	2
（二）高等教育质量保障的理论研究概况	4
二、研究学生学习成果评估的意义	5
（一）我国教育质量保障运动亟需方法与技术创新	5
（二）学生学习成果评估是国际相关领域的前沿热点	9
（三）为何选择从生态视角研究学生学习成果评估	13
三、研究设计	14
（一）研究目的	14
（二）研究问题	15
（三）研究目标	16
（四）研究思路	16
（五）研究方法	17

第二章 多源流政策分析框架及应用	19
一、主要政策分析理论概要	20
(一) 完全理性决策模型	22
(二) 有限理性决策模型	23
(三) 渐进式决策模型	24
(四) 垃圾桶式决策模型	24
二、多源流决策模型及其要素	27
(一) 问题源流	29
(二) 政策源流	31
(三) 政治源流	32
(四) 政策之窗	33
三、本书理论分析框架	35
(一) 对政策决策模型进行阐述的目的	35
(二) 本书分析框架——对“多源流决策模型”的调整	37
第三章 学生学习成果评估的生态源流（基础教育篇）	39
一、20 世纪 70 年代：教育公平诉求发生及发展	42
二、20 世纪 80 年代：从教育公平转向教育质量	49
三、20 世纪 90 年代至今：从教育投入转向教育产出	58
(一) 立法—行政体系之生态及努力	58
(二) 司法战线之生态及努力	61

第四章 学生学习成果评估的生态源流（高等教育篇）	65
一、政府和社会公众的问责压力	66
二、认证机构的作为与学术人的努力相辅相成	71
（一）外部问责力量的坚持	74
（二）质量认定机构的有力作为	74
（三）学习成果评估学术的蓬勃发展	80
第五章 学生学习成果评估的技术源流	90
一、什么是学生学习成果评估	91
（一）学生学习成果评估的内容	92
（二）学生学习成果评估的分类	95
二、学生学习成果评估的方法与工具	100
（一）评估基本方法	101
（二）具体评估方法（数据收集手段）	103
（三）学习成果评估分析方法	115
（四）学生学习成果评估工具	122
三、学生学习成果评估方法的技术效能	137
（一）技术效能的分析框架	138
（二）学习成果评估方法之技术效能分析	140
第六章 学生学习成果评估对我国高等教育质量保障的启示	147
一、高等教育真正的质量危机	147
二、美国对高等教育质量危机的应对	150

三、对我国现阶段高等教育质量保障的反思与建议	154
(一) 我国是否有进行学生学习成果评估的必要	155
(二) 我国的生态条件是否适合引进学生学习成果评估 ...	160
参考文献.....	165
附录 1 嫫茜学习成果评估三步法的核心要点	178
附录 2 学生学习成果评估操作九项原则	180
附录 3 中部认证委员会关于学生学业成长档案夹的评估量规 ...	184
附录 4 NSSE 问卷例题 (2017 年)	187

第一章 导 论

20 世纪 90 年代末,中国高等教育开始从精英化进入大众化阶段。国家通过老校潜力挖掘,新建地方本科院校,大力发展民办教育、独立学院,发挥成人教育、远程教育作用等措施,在短时间内迅速调整了高等教育的布局结构,使高等教育资源分布在地域空间以及办学层次上更加丰富、平衡,也大大满足了人民群众对高等教育的迫切需求,在促进我国政治、经济、文化的发展上具有划时代的历史意义。2002 年我国大学在校生人数达到 1512 万,毛入学率达到 15.3%,提前 3 年完成了《面向 21 世纪教育振兴行动计划》(1998)提出的高等教育大众化第一阶段目标^①。据教育部《全国教育事业统计公报》显示,2016 年,大学毛入学率已达到 42.7%,在校生人数达到 3700 万。我国高等教育事业正在由大众化向普及化过渡。

国际上首先实现高等教育大众化的是美国,于 20 世纪 40 年代达到

① 关于高等教育发展阶段划分,美国学者马丁·特罗的系列研究认为:当一国高等教育毛入学率低于 15% 时,为精英化阶段;当毛入学率介于 15% ~ 50% 时,为大众化阶段;当毛入学率超过 50% 时,则进入普及化阶段。这一结论也是目前国际教育界通用的划分标准。

15% 的毛入学率，然后是欧洲。我国用不到 20 年的时间走过了发达国家数十年甚至上百年的从精英教育到大众化教育的发展阶段，尽管成就举世瞩目，但“跑得过快”的副作用也凸显出来。部分高校的物质条件、队伍建设、制度构架等没有达到承担大规模教育任务的要求，是不得不承认的事实。

一、我国高等教育质量保障概述^①

（一）高等教育质量保障的历程与阶段性进展

在这种情况下，各界要求提升高等教育质量，进行内涵式发展的呼声此起彼伏。教育质量保障由此进入最为重要的历史时期，其力度之大，范围之广，用“运动”概括实不为过。

事实上，我国具有教育质量保障性质的行政管理实践早在 20 世纪 80—90 年代就已开始。1985 年，《中共中央关于教育体制改革的决定》提出，教育管理部门要“定期对高等学校的办学水平进行评估”。而 1990 年，当时的国家教育委员会通过颁布《普通高等学校教育评估暂行规定》，进一步对 1985 年发文提及的评估性质、目的、任务、指导思想和基本形式等做出了明确规定。同年，8 个部委、6 个省市的近 500 所学校接受了国家教委的评估，内容包括办学水平评估、专业评估和课程评估等，标志着我国高等教育质量保障进入起步阶段（张建新，2014）。

① 高等教育质量保障涵盖本科和研究生教育两个层次。鉴于域外学生学习成果评估的理论研究与实践主要针对本科教育，因此本书中高等教育评估与质量保障的研究对象范围仅限于本科层面。

1993年,中共中央、国务院颁布的《中国教育改革和发展纲要》指出,“建立各级各类教育的质量标准和评估指标体系”,各地教育部门要把检查评估学校教育质量作为一项经常性的任务。从1994年起,国家教委开始实施普通高等学校本科教学工作水平评估。截至2001年,教育部共完成对254所学校的评估,其中,在192所新建本科院校实施了“合格评估”,在16所一线大学实施了“优秀评估”,抽取介于二者之间的46所大学实施了“随机性水平评估”。同时期,江苏、上海、云南等地建立了多家教育评估机构,“武书连大学评价排行榜”以及类似自发性民间评价活动也呈现出日益活跃的姿态。

经过10年的摸索,2003年,教育部在《2003—2007年教育振兴行动计划》中,将“完善高等学校教学质量评估与保障机制”作为一项制度加以确立,执行周期为5年。2004年,高等教育教学评估中心成立;同年,首轮“普通高校教学工作水平评估”工作启动,教育质量保障开始步入规范化、制度化的繁荣时期。首轮评估动用了6000人次进校听课1.5万余门,审阅毕业论文2.9万份,座谈2500多场次(刘振天,2012)。截至2008年,全国共计1300多所普通高校接受了评估,其中,本科院校589所(占当时普通本科院校的近80%),“优秀”率为72%，“良好”率为24%，“不合格”率为0%。

2011年,教育部颁布了《关于普通高等学校本科教学评估工作的意见》,确定了以学校自我评估为基础,以院校评估、专业认证及评估、国际评估和教学基本状态数据监控为主要内容的评估制度,简称“五位一体”本科教学评估制度。该制度是我国现阶段以及将来一段时间内开展评估工作的主要路径。首先应用这一制度的是2014年启动的“普通高等学校本科教

学工作审核评估”。与水平评估不同，审核评估不设国家统一标准，主要是考察被评估对象是否达到了自身设定的人才培养目标，审核结论不分等级，仅形成写实性审核报告。根据《普通高等学校本科教学工作审核评估实施办法》，凡参加普通高等学校本科教学工作水平评估获得“合格”及以上结论的高校均应参加审核评估，评估时间为2014—2018年。

（二）高等教育质量保障的理论研究概况

我国高等教育质量保障实践的展开离不开理论界的不断探索。20世纪80年代中期，也即国际教育质量保障运动在美国蓬勃生长的节点，《比较教育研究》开始刊发研究西方高等教育认证、鉴定和评估的相关论文和译文（林杰，2015）。1996年，以“高等教育评估与质量保障”为主题的国际学术研讨会在北京召开。从20世纪90年代后期至今，在国内学术期刊上刊发的与高等教育质量评估和质量保障相关的论文达上千篇，出版专著无数。学者们的焦点主要汇聚于三个子领域：第一，高等教育质量的内涵；第二，高等教育质量保障体系的构成；第三，高等教育质量保障的方法与技术。即，解决关于高等教育质量及其质量保障“是什么”和“怎么做”的问题。

上述某些研究对我国高等教育质量保障的推进具有积极的指导意义。中早期代表中，陈谟开主编的《高等教育评价概论》，陈玉琨著的《中国高等教育评价论》，陈玉琨、代蕊华、杨晓江、田圣炳合著的《高等教育质量保障体系概论》，王致和主编的《高等学校教育评估》等著作，在介绍和引进国外经验的同时，结合中国的实际情况，论述教育评估与质量保障的本质、目的、指导思想、基本原则、模式和标准、评估的程序以及政

府、社会、学校在评估中的地位作用和相互关系等，为我国开展高等教育评估和质量保障的本土化工作打下了坚实的理论基础。吴岩主编的《构建中国特色高等教育质量保障体系》和《国际高等教育质量保障体系新视野》则可视为在近段时期的集大成者。《构建中国特色高等教育质量保障体系》运用文献研究和历史分析法，记录了中国高等教育以评估为抓手进行质量保障的进程，并较为详尽地阐释了“五位一体”本科教学评估制度的具体内涵。而《国际高等教育质量保障体系新视野》则以横断研究的方式，对现阶段域外保障运动的基本理论、体系构建、方法技术、组织实施、国际项目等进行了较为全面的梳理与介绍。对于初入高等教育评估与质量保障领域的学子以及研究者而言，这两本著作具有较高的参考价值。

二、研究学生学习成果评估的意义

（一）我国教育质量保障运动亟需方法与技术创新^①

在应对高等教育质量危机方面，我国管理界的反应并不慢。“合格评估”“优秀评估”“随机水平评估”“本科教学工作水平评估”“本科教学工作审核评估”等系列评估活动的开展，不仅覆盖范围广，而且多年来一直坚持不懈。“以评促建、以评促改、以评促管、评建结合、重在建设”，教学评估促使高校进一步厘清了办学思路，明确了办学定位，强化了高校教学工作的中心地位，显著改善了高校的办学条件（吴岩，2014）。

^① 本书所涉及“技术”一词，并非为一般意义的“信息化技术”的简称，而是特指统计技术或测量技术，具体含义将在本章主要概念部分做阐释。

在充分肯定教学评估取得成绩的同时,必须清醒地认识到,作为新生事物,其还存在许多问题和不足。有学者认为当前教学评估的主要不足在于评估标准多样性不够,分类指导的作用未得到充分发挥;有学者认为由于缺少常态性的质量监测手段,过分依赖于专家短期进校阶段的评估考察,会造成教学评估带有一定程度的“运动式”特点。

上述问题仅仅是我国现阶段教育质量保障问题的表象,造成其实践不足背后深层次的原因在于现有教学评估的方法与技术难以协助实现常态化质量保障的最终目的。以“本科教学工作水平评估”为观察点,窥斑见豹。据高等教育教学评估中心记载,水平评估中采用的评估方法与技术主要如下:

(1) 审阅材料。专家采取浏览和精读相结合的方式,根据分工,审阅学校提供的评估材料和支撑材料,必要时调阅有关原始材料。

(2) 考察教学基本设施。考察的范围包括学校的基础设施、公共基础课以及主干专业基础课实验室、重点实验室、主要实习基地、图书馆、网络中心、多媒体教室、体育设施等,目的是在短时间内对学校的育人环境、办学条件有一个宏观的了解和整体的印象。

(3) 随机听课。该方法是评估专家了解课堂教学、实验教学情况,对教师、学生、课堂、教学内容、教学方法、教风、学风等情况进行感性了解的主要手段。专家根据自身学术背景,选听不同类型、不同教师的课程,听课填写专家听课评价表。

(4) 调阅毕业设计或毕业论文和试卷。采用学校自荐和随机抽查相结合的形式。专家审阅毕业论文和试卷后,填写相应的评价表。

(5) 座谈会。根据不同目的召开不同类型的座谈会,如管理人员座

谈会，老中青专任教师座谈会，高年级（低年级）学生座谈会，教辅人员座谈会等。

（6）走访教学单位和职能部门。专家根据自身学科背景和负责的评估项目分工，走访教学单位和主要职能部门，走访时间一般控制在1个小时以内。走访时除了座谈讨论外，一般还需参观教学设施、实验室和查阅资料。

（7）学生的调查和测试。该方法为了解学校的教学状态、学生的学习质量和基本素质的主要手段，方式一般为问卷调查或座谈，从实际执行情况来看，后者居多，而前者多用于院校的自我评估。

以上七大评估方法，对评估对象的教学工作覆盖还是比较全面的，但是评估技术却相对滞后，导致评估工作本身的效率和质量不尽如设计预期，体现在以下两个方面：

第一，在评估指标的选择与构建上，重心倾向于师生比、基础设施、软硬件设备、资金投入、课程内容和教学形式等硬件和外在指标。外在条件型指标不仅检查工作量大，而且已经被域外大量研究证明与教育质量并无显著、直接的正相关关系。正如彭森明（2008）在《将学生学习成果纳入大学评鉴指标项目之必要性与可行性》一文中比喻，“这就好比我们只在评价厨师（教师）的资历，以及做菜的材料与工具（学校设备与课程），而未能实际品尝做出来的饭菜”。因此，从测量学的技术标准之一——效度来看，我国现阶段教学评估在质量保障方面的评价的有效性有待提升^①。

① 高等教育教学评估中心清楚地认识到了这一点。尽管学界将“本科教学工作水平评估”等系列评估活动视为我国高等教育质量外部保障机制的代表，但是高等教育教学评估中心始终以“教学工作评估”命名之，而非“教育质量评估”。

第二，与教育质量较为相关的指标的数据采集多采用随机听课、抽查作品、抽样访谈等方式，考虑到专家组规模（一般为7~9人）和进校时间（一般不超过7天），相对于大多数高校的学生规模、教师规模以及开课数量，不难判断专家组的多个现场调查均不能达到具有代表性的抽样标准。此外，听课、查试卷、访谈、问卷调查等评估手段均以专家（或者学生）的主观质性（或量化）评价为最终结果，人际评价标准差异较大，评价结果影响因素众多。从测量学的另一技术标准——信度来看，我国现阶段教学评估在质量保障方面的评价的可靠性也并不高。

教学评估在指标选择效度方面不达标，会在一定程度上造成参评高校形式主义泛滥。突击基础设施修缮、软硬件设备更新，教学大纲、教案、试卷大清理，一反常态的课堂纪律要求，扬长避短的自评报告……防不胜防的形式主义，正在侵蚀教学评估的公信力（杨惠明，2015）。教学评估在数据信度方面的不达标，则进一步恶化形式主义作风的动机机制。由于评估结果大量充斥着主观型数据，缺乏客观公正标尺的统一度量，教学评估结果一直无法像一级学科评估^①那样形成常模参照式（Norm-reference）全国排位，而只能采用标准参照式（Criteria-reference）等级结论（优秀、合格和不合格），但前者恰巧是社会和市场所追求的信息。再加上“教学工作水平评估”到目前为止没有一所学校因此而“不合格”，甚至72%参评高校达到了“优秀”等级，其“形势一片大好”的总体结论对于“以脚投票”督促高校提升教育质量的逻辑选项已失去意义。因此，在我国出现了一个信息流通偏差，以媒体为主要介质，大众往往基于一级学科评估结

① 一级学科评估是高等教育学科学术实力的主要评估机制。

果判断高校本科教育质量高低，这是一个对评估结果应用的误导。雪上加霜的是，当我国高等教育跨入追求教育质量提升的实质性、“内涵式”发展时代，国家领导人和教育部倡导高等教育须回归“教书育人”本分的时候，教学评估结果仍未与任何国家教育财政拨款和教育资源的分配产生直接关联（张建新，2014）。在如火如荼的“双一流”建设及其一流学科评估基石面前，本科教学工作评估在高校眼里已经基本演变为质量保障的“鸡肋”。

我国教学评估亟需方法和技术的创新，助力其摆脱信、效度低下的“输入条件”类指标以及主观型数据的限制。没有新方法和技术的介入，“五位一体”教学评估制度将面临“换汤不换药”的改革困境，“自我评估”“院校评估”“教学基本状态数据常态监测”等核心构成仍容易遭受形式主义的侵蚀。尽管“专业认证”和“国际评估”的抗形式化能力要强一些，但毕竟不是每个学科、每个大学都有条件实施的。

（二）学生学习成果评估是国际相关领域的前沿热点

自质量保障运动于20世纪中叶在各国相继兴起后，评价指标如何合理、准确地反映教育质量的内涵一直是国际高等教育界面临的共通性难题。随着高等教育大众化、普及化的持续推进，全球经济一体化影响的不断渗入，一些世界主要国家（比如，美国、英国、法国、日本、澳大利亚等）均构建了基于自身文化传统、政治体制以及大学自治实际的高等教育质量保障体系及其评估模式。总的来讲，学界倾向于将迄今出现的多个质量保障机制划分为外、内二维结构。外部质量保障一般包括外部机构对高校的认证、审计、评估；内部质量保障主要包括高校的自我认证、自我评估和质量管理。

此外，经济的市场化和全球化也催生了多个大学排行^①。严格来讲，大学排行榜不属于高等教育保障的组成部分，也不属于外部机构对大学的评估（林杰，2015）。但是，鉴于其在促进高校提高办学质量、积累学术声誉方面有着极大的影响力，学界在研究教育质量保障时，往往将其纳入讨论范畴。

尽管内、外保障机制与模式已经比较丰富多样，但大多数机制并没有实现令人满意的对教育质量的数据诠释。以读者更为熟知的大学排行榜为例。比如上海交通大学的“世界大学学术排名（ARWU）”，对“教育质量”的评价仅设置一个指标，即诺贝尔奖和菲尔茨奖校友的数量，赋权 10%。而 QS（Quacquarelli Symonds）^② 系统中教学类指标权重比 ARWU 高一些，为 40%。其中，学生就业评价或雇佣者的评价占 10%，师生比占 20%，国际教师占总教师数的比例占 5%，国际学生占学生总数的比例占 5%。相比较而言，QS 对教育质量的解读，虽不一定能准确反映大学教学的输出端质量，但至少还在教学相关范围。而 ARWU 以极少数校友的科研成就衡量大学的普遍教育质量就显得极其狭隘了。当然，类似于 ARWU 这样解读教育

① 王英杰（2008）认为，经济的市场化和全球化直接催生了大学排行。学生、家长以及其他利益相关者在经济市场化的过程中，正在从高等教育消极的“接受者”转化为主动积极的“消费者”。大学教育因而从他们的角度变为“消费产品”，从而有了名牌与非名牌之分，有排名靠前与靠后的差异，于是大学排行就成了他们的“消费”指南。而在经济全球化中，高等教育成为跨境贸易的重要构成部分，道理等同，既然是贸易，就会出现商品“质量”排行，大学全球排行榜就应运而生了。据 UNESCO2013 年的调查，全球范围共有国家大学排名 50 余个，世界大学排名 10 个。其中，由政府机构作为主办方进行的有 12 家，独立机构主持的有 14 家，媒体作为主办方的 15 家（Marope、Wells、Hazeorn，2013）。

② QS 是一家跨国教育及学术资讯信息公司。美国《美国新闻与世界报道》的“世界最好大学”以及英国《泰晤士报》“世界大学排名”均先后与该公司合作。