

从水平评估到审核评估： 我国高校教学评估理论认知及实践探索

刘振天

摘要：从水平评估到审核评估，不仅是我国高校教学评估实践的新探索，更是教学评估理论的新认知。这种理论认知及实践探索，集中表现在四个方面：一是在评估理念上，从强调绩效问责走向注重协商对话；二是在评估作用上，从刚性强化走向柔性激励；三是在评估分类上，从单一标准走向多样标准；四是在评估重点上，从教学具体活动走向内部质量保证体系。

关键词：高等学校；教学工作；水平评估；审核评估；评估理论

我国高校教学评估工作始于20世纪90年代，历经25年，期间先后开展过合格评估、优秀评估、随机评估、水平评估、审核评估等几种不同形态的评估。但若从评估规模、效果及其影响来看，毫无疑问当以水平评估与审核评估最为典型、最具代表性。前者全称“普通高等学校本科教学工作水平评估”，习惯上也叫首轮水平评估，自2003年开始至2008年结束，6年间共评估589所高校。后者全称“普通高等学校本科教学工作审核评估”，实际上是第二轮评估，开始于2013年，将于2018年年底结束，630余所高校参加，包括了接受过首轮水平评估的所有高校。在笔者看来，审核评估与水平评估，既一脉相承，又有创新与发展。就此而论，审核评估之于水平评估，不是简单地换个名词或改一种称谓，实际上，在评估理念、标准、内容、方式、技术及组织等方面相应做了一系列调整 and 变化，表明我国对评估工作或评估制度在理论上有了新认识，在实践上有了新进步。审核评估全面总结和借鉴了水平评估的经验，深入反思并汲取了水平评估的教训，是我国高校教学评估实践的新探索，更是教学评估理论

认识水平的新提升。

一、教学评估理念：从强调绩效问责走向注重协商对话

教育评估或评价活动既古老又年轻。言之古老，是因为自教育产生时起就存在着评价。孔子就曾讲，听其言观其行而后知其人。苏格拉底也提到，未经反思和评判的生活没有意义。说它年轻，在于作为一门学问和一种制度，教育评估评价只是在20世纪初才得以确立，并且在很大程度上受益于工商业发展以及现代企业质量管理理论与实践的启示。在19世纪前的西方社会，高等教育的标准与质量多由大学内部掌握和控制，政府及社会很少关心和过问。随着高等教育规模不断扩张，入学人数不断增多，大学从过去教会和私人领域逐渐变成国家和社会的公共事业，高等教育也由远离社会一步步走到社会的中心，其事业的发展和扩充，越来越依赖于政府和社会的财力支持。20世纪70年代之后，西方国家先后进入高等教育大众化阶段，政府财政投入吃紧且呈减缩态势，相应地，

刘振天，中国人民大学教育学院教授。

个人教育成本负担增加。在此情况下，无论是政府还是社会与公众，普遍关注教育质量，要求大学对质量做出承诺，由此开启了政府和专门机构主导下的高等教育质量外部问责和绩效评价时代。西方主要国家纷纷制定高等教育质量标准，颁布质量法律，建立质量评价制度，成立质量评价机构，开发质量评估工具，开展质量评估认证活动。有些国家还开展双边或多边高等教育质量相互认证工作，区域性乃至国际性高等教育质量组织得以成立，掀起了大规模国际高等教育质量保障运动。所有这些，对改进高等学校工作、提高高等教育质量具有不可替代的作用。

在世界高等教育质量保障运动大背景下，20世纪80年代中期，我国高等教育质量保障工作亦提上议事日程。1985年，《中共中央关于教育体制改革的决定》首次提出，政府要对高等学校办学水平和教育质量进行评估。此后，教育部在工科高校数百个专业进行了教学评价试点。20世纪90年代，教育部将专业评估转入院校整体性教学评估。1994年，国家对改革开放后批准设立的高校本科教学工作开展合格评估；1996年，开始对那些申请进入国家“211工程”重点建设项目的高校本科教学工作开展优秀评估；1999年，对其他普通高校本科教学工作开展随机性水平评估。这三种评估工作均持续到2002年，总计评估254所高校，其中以接受合格评估的高校最多。2002年，教育部将三种评估合并为普通高等学校教学工作水平评估，即首轮教学水平评估，从2003年正式实施。虽然此前的三种评估被合并为一种评估，但其在评估思路、评估方案等方面并没有多少实质差别，可视为同类性质的评估模式。

指导首轮水平评估（包括之前的评估）的基本理念与思想，就是人们耳熟能详的评估十六字方针：“以评促建、以评促改、评建结合、重在建设”。随着实践的深入，再增加“以评促管”，发展为评估二十字方针。不管是十六字还是二十字，都代表了那个时期人们对评估理念与功能的总结和概括，反映了人们对评估

本质、价值及目的的基本认识和理解。

透过首轮教学水平评估以及评估方针，我们不难发现，教学评估主要是由政府直接组织的，当然要体现政府的意志和意图，体现政府对高等教育质量的问责，评估是为政府对高等教育及其质量实施有效管理和监督服务的。通过评估，政府可以检查高校教学工作是否符合、贯彻和落实了党和国家有关教育方针政策及法律法规要求，检验高校教学效果是否达到了政府规定的质量标准，督促高校进一步加强教学建设、改进教学工作、提高教学质量。为实现上述目的，政府制定了教学评估指标体系，规定了评估标准和评估具体要求，其中不仅有量化指标，还要区分出评估结果等级序列，进而使高校教学与人才培养工作变得可观测、可衡量、可检查、可比较、可参照，进而与一定的奖惩政策制度衔接。

政府对高校教学质量进行问责评估本为分内之职，制定评估标准也是法律赋予的权力，但政府并不是质量问责和评估的唯一主体，否则就会否定和排斥其他组织或群体的价值存在，甚至在某种程度上造成评估本质扭曲、评估目的错位和评估功能变形。事实上，现代高等教育是一个巨大的利益共同体，政府、高校、社会、公众、媒体、家长、学生、校友、用户等诸多群体都有着各自的利益需要和价值诉求，也有着不同的评价标准，高校教学工作评价也是如此。由此，当前国际上出现了所谓的第四代评估理论。这种评估理论认为，先前评估所经历的三个发展阶段（学业测量、实情描述和效果判断）都是理性主义、实证主义、管理主义和功利主义的产物，它们均将评价对象看作被动存在者，评估成了一种权力对另一种权力的压迫、规约和限制。第四代评估理论宣称，评估不是为了检查和证明，不是为了单纯测量和判断，而是为不同利益主体提供平等交流和协商对话的平台；评估是交流对话的机制，是各利益主体相互间围绕着质量与改进所展开的建构性评价活动，评估本身不存在确定性的单一量化标准^[1]。第四代评估理论受到各国评估界的

高度重视，并以此为指导改进评估认证与质量保障工作，其中众所周知的学校主体、自我评估、持续改进、发展性评价、过程性评价等，就是这一理论的重要表现形式。

在我国，首轮教学水平评估之后，人们对评估经验教训、是非功过进行了认真总结和反思，认识到单纯政府问责、绩效检查和量化评估的局限性。在广泛听取各方面意见，借鉴国际先进理论与模式基础上，教育部对首轮水平评估进行了较大的调整和改造，其重点是合理吸收了第四代评估理论。审核评估这一新术语，某种意义上表达了第四代评估新理念。虽然审核评估并未根本改变也不可能更无必要消除以往评估的说明、检查、鉴定、激励和问责等特质，但它更有意识地注重和突出高校的主体性，强调政府与高校间的平等性，认可评估是不同利益方的沟通和交流，共同就人才培养与教学工作诊断协商、梳理思路、查找问题和谋求改进之道。审核评估明确提出“为国家负责、为学校服务、以学生发展为本位”，学校自我评估是全部评估的基础。审核评估取消了首轮水平评估的等级设置，突出评估的目的在于促进高校内涵建设和特色发展，不再使用统一的评估指标体系这种单一标尺来衡量所有的高校，而是立足高校自我目标定位，用高校自己设立的标准评价高校自身工作及其发展。在评估主体上，除政府评估外，积极鼓励专门机构、高校和社会参与评价，构建多主体、多形式、多渠道的评估制度，给高校发展空间，给公众和社会评价权利，形成多主体竞争与合作的格局。事实上，首轮评估后，第二轮审核评估中，政府除了制定评估标准和规划外，评估的具体组织已经由中央和地方性评估专门机构通过接受政府购买服务的形式获得了评估代理权，第三方评估建设取得了重要进展。此外，社会团体组织的专业认证、高校自主选择参与国际评估业已形成趋势，政府单一评估局面基本被打破，为多方对话和价值表达进而影响高等教育质量决策提供了广阔舞台。正因为如此，新一轮评估受到高校、专家及社会的普遍认可和积极评

价。有高校形容，过去接受合格评估是为了生存，接受水平评估是为了声誉，而参加审核评估则是为了自我发展^[2]。有专家认为，参加审核评估最大感受就是高校的平常心 and 开放态，不回避矛盾、不隐瞒问题，学校愿意就教学工作中遇到的问题、困难和困惑对专家敞开心扉、平等交流，期待从专家那里受到启发，找到解决问题的办法。还有研究者指出，审核评估真正体现了教育评价的本质，让评价回归其本来面目和应有位置，它不再是独立于高校教学工作的外在强加物，而成为高校教学工作的内在构成，不再是简单地给高校教学工作下个结论或者贴上等级标签，而是为了帮助学校发展，各方从中都能受益^[3]。所以，审核评估较好地平衡了政府、高校和社会三者之间的关系，体现了不同主体协商对话的评估新理念。

二、教学评估作用：从刚性强化走向柔性激励

无论是水平评估还是审核评估，都是政府主导下的外部评估，其问责、说明、检查和督促的作用是既是正常合理的，又是正当合法的。长期大量的评估实践证明，作为督促和问责手段，评估本身也存在一个度的问题，掌握得好、运用恰当，评估就能发挥其积极正向作用；否则，将会适得其反和事与愿违。通常人们所谓的评估是把双刃剑，就是这个意思。

一般说来，评估者总是希望评估工作能够足以引起被评者的充分重视且对其工作产生实质影响，通过评估给工作带来新发展新变化，尤其是作为教学评估政策和评估方案的制定者与组织者，政府部门对评估工作的作用力往往持有更高的目标期待。在我国，评估被赋予特别的意义和价值。在政府看来，教学评估是高校工作的指挥棒和方向盘，是反映高校工作优劣的镜子和衡量教学质量的尺子，是助推高校教学的发动机，是撬动高校工作的强大杠杆，也是推行政策的重要手段，特别是当高等教育领域存在重科研轻教学、重学科建设轻人才培

养现象且这些现象和问题日趋严重却见不到有效解决迹象时，政府对评估就更为看重，期待通过教学评估扭转局面。从这点上讲，政府是加重评估强度、提高评估作用力的第一推手。政府提高评估强度和作用力的途径与办法很多，主要有设定评估标准、规定评估等级、量化评估指标、对评估结果进行排名和发布、评估结果与拨款或其他有关政策挂钩等。

与政府相比，在如何对待评估压力和强度问题上，高校的角色及心态可能要复杂得多，有时甚至非常矛盾。首先，高校是被评对象，必须接受政府或其代理机构的评估，如今，评估已经成为世界各国高等教育质量保障的一项长期制度安排，正如美国学者E·格威狄等人所指出的，“所有大学/学院的质量和效率需要定期的评估，特别是本科生的绩效评估应为重中之重”^[4]。然而，大学学术自由自治的传统与个性，又使其天然地表现出抵制和拒斥外部评估问责的倾向，时刻提防着外部评估认证对大学自身发展内在逻辑造成的威胁、伤害和破坏，尤其在一些自由主义文化浓厚的欧洲国家，很少有发自内心愿意主动接受评估的大学，即使接受了评估，对评估的非议和负面评价也从未间断过，在他们看来，评估会导致集权化、标准化、忽视大学个性化发展等问题^[5]。其次，高校教师的自由职业者特征常常散发出极强的各行其事的个体主义色彩，他们害怕也反感外部机构对自身的教学工作进行评价，教师的这种立场与角色又使得高校管理层面总是试图借助外在的力量对之加以刚性约束和规制，以便强化自身的权力和权威，落实或推行相应的政策与措施，提高行动与目标的一致性，达到工作的高效率和高质量。在这一点上，高校行政管理者又期望提高评估的强度和力度。评估和认证中存在的目标责任层层落实、任务层层加码和压力层层传导现象就是最有力的证明。在行政管理人员心目中，似乎评估的强度或力度与评估的实际效果呈现线性正相关，以为评估强度越大，人们就越重视，效果就越显著，反之，评估强度越小，就越受人轻视，效果也就越差。

以高校首轮教学水平评估和目前正在展开的审核评估工作为例，通过比较不难看到，二者在作用强度和力度上恰好形成了明显对照。水平评估被分成优秀、良好、合格与不合格四个等级，每一评估等级所对应的标准和要求不同，等级越高，标准和要求相应也就越高。之所以进行水平评估，在于评估政策制定者和方案设计者不仅试图通过等级评估来客观描述和区别把握我国高等学校教学工作的“真实”水平，更在于刚性强化地方政府和高校对教学工作的重视和投入，进而实现有限资源得到非均衡性优化配置和奖惩的目的。应该说，水平评估在某种意义和程度上确实实现了政策制定者和实施者的初衷，在世纪之交高等教育大发展大扩招条件下，保证了高校教学基本条件、基本秩序和基本质量，也激发了高校之间在办学资源、声望与质量上的竞争。据对2007年198所参评高校的统计，水平评估给高校生均四项教学经费带来超过27%的增长率，而生均教学仪器设备值更是带来33%的增长额^[6]。另有调查表明，96%的高校领导认为水平评估作用显著^[7]，74%的高校教师认可水平评估的价值^[8]。然而，水平评估所设置的高低好差不同级别，不单纯表征不同高校的质量、水平、声誉、地位和影响，甚至主要显示着学校管理者包括其行政管理部门的管理水平、能力和政绩，因此，水平评估最终不由自主地演变成一场围绕等级、名声与地位的激烈角逐。高校对评估的重视，与其说是为了实现评估目的和提升教学质量本身，不如说直接为了等级结果、奖牌和称号，这势必导致评估变形和目标游移。

2008年水平评估全面结束后，政府、机构和高校各方面对水平评估进行了认真总结和深刻反思，认识到水平评估等级设置的优势与局限，提出了改进评估的基本对策，即取消等级，减轻评估对高校造成的过大压力和过重负担。待到审核评估方案确定后，水平评估等级已被彻底取消，代之以写实性评价，即说明和描述被评高校教学工作哪些方面值得肯定、哪些方面需要改进、哪些方面必须整改，评估压力及

评估成本随之大大降低。

不过，取消等级后的审核评估工作，似乎又走入另一极端、陷入另一种困境，评估由以往刚性过大、压力过重、强度高变成了柔性偏大、压力偏小和强度偏弱。水平评估中存在严重的功利欲、攀比心已经让位于审核评估中的平常心、正常态。审核评估回归了评估应有的面目，它更加强调高校在质量保障中的主体地位和多样化发展生态，学校评估中更加自主、自信和从容，真正把注意力放在学校内涵建设和特色发展方面，不再为外在等级结果所左右，审核评估不打扰学校正常秩序，但由于它注重问题导向，注重与学校平等交流和协商对话，学校放下了水平评估状态下的防御、戒备、博弈和讨巧心理，致力于通过审核评估发现问题、分析原因、寻求对策和提高质量。所以，审核评估得到许多高校和专家的充分肯定和积极评价，认为审核评估是尊重教学规律的评估，是促进学校发展的评估，是指向内涵建设的评估。但平常心之下学校对评估用心不多，正常态之下学校工作状态表现不佳，地方教育行政部门和高校对审核评估重视程度不够，审核评估带来的教学激励不大，教学改进实际效果不明显，甚至有图形式、走过场的迹象。所以，有政府管理者、高校领导和专家主张，要辩证对待和评价水平评估与审核评估，充分吸收各自长处，避免各自短处，适当增加评估压力和强度，以此提升评估影响力和作用力，更有效促进高校本科教学工作，提高教育教学质量。

水平评估到审核评估，评估的作用确实出现了从高到低、从大到小、从强到弱、从刚到柔的明显转换，相应地，人们对评估被重视程度和评估实际效力的判断也出现了变化。如何看待评估的强度和作用力以及由此引起的功用？到底是强度大好还是强度小好？评估作用强度与评估效果是否一定存在明显正相关或者说二者成正比例关系？笔者认为，这个问题比较复杂，不可一概而论。关键要看这种相关性是形式上的还是实质上的，是表象的还是内在的，是一时的还是长远的。无论作用强度大小，

外部评估问责都是重要的、不可或缺的，同时，它作为一种手段，必须服从和服务于教育教学自身目的及需要。外部评估问责所承载的功能不可能是无限的，其最终的效果和作用，进一步说，教学工作能否得到改进，教学质量能否得到提升，归根到底取决于学校自己。尤其是高等教育进入实现内涵式发展以及治理体系和治理能力现代化的新时代，政府职能将进一步向法治型和服务型转变，高校办学自主权日益扩大，外部评估作用刚性的空间会逐渐收缩，柔性空间逐渐扩充。由是观之，那种过于相信评估刚性作用的认识并不合时宜，适度柔性实为大势所趋。

三、教学评估分类：从单一标准走向多样标准

伴随高等教育大众化发展，高等教育的社会化程度也日益提高，其内部分化过程加速。换言之，高等教育大众化、社会化与多样化是密切相关、联为一体，因此才有所谓高等教育的多样发展和特色发展。要实现多样化、特色化，不仅需要高校办学自由自主、主动对接市场和社会需求，而且政府和社会也需要加强分类管理和分类评价。

高等学校分类是重要的教育理论问题和实践问题。中外学界有许多研究，取得了非常可观的成果。但严格说来，已有的研究及成果，各有优劣和长短，并且高校如何分类也是与本土经济、政治、文化与教育基础相适应的，并不存在一个统一的模式^[9]。最有代表性的美国卡内基教育基金会依照高等学校职能和学术层级，将高校分为研究型大学（内部又细分为研究型1类和2类）、博士类大学、硕士类大学、四年制文理学院以及社区学院等^[10]。其中，研究型大学和博士类大学，重视研究生教育和科研工作，而硕士型大学和文理学院主要侧重本科教学，社区学院则具有升学（本科）教育、职业教育及满足成人个性化发展教育的多种职能。这种分类与美国注重通识教育、研究生教

育规模大以及劳动力培训市场高度发达直接相关。与美国不同，欧洲高等教育分类更多从职能划分，主要分为学术型、应用技术型和综合型三类^[11]，其中，前两类是欧洲传统的学术性教育与职业性教育双轨制遗产。学术型教育地位高，实科与职业教育地位低。战后在民主化运动影响下，欧洲高等教育的平等化发展迅速，大面积出现学术与职业兼具的综合性中等学校和综合性高等学校。

在我国，高校分类是一个新课题，也是长期以来没有得到很好解决的实践问题。民国时期的高等学校，主要仿效西欧国家，分为综合大学和专门学院。前者侧重学理性或基础性教学与研究，后者则侧重某一方面的专门学问教学。中华人民共和国成立后，全面学习苏联经验，对高等学校进行根本改造，强调高等教育专业属性，除了保留少数文理性综合大学之外，绝大多数高校都是按照狭窄的专业领域甚至产品进行建设和管理，专业多而细。即使是综合大学，内部也存在着严格的专业划分并严格按照专业进行招生、教学、管理和就业工作。改革开放后，特别是新世纪以来，高等教育的大众化发展，市场经济体制的确立，高等教育多样化发展成为必然，高校如何分类和定位也成为突出问题。政府行政部门为了管理及资源配置等需要，组织力量加强了分类研究，社会评价机构为了高校质量、声誉等排名需要，在分类上也进行了许多有益的探索。有的按照职能划分，如研究型大学、教学型高校、服务型高校，进一步又细分为研究型、研究教学型、教学研究型、教学服务型、职业技能型等；还有的按照知识的性质进行划分，如学术型大学、应用型高校、技能型高校等。这些划分尽管各有道理和优长，但若将之应用到具体实践管理和评估工作中，问题和弊端就会暴露无疑甚至寸步难行。例如，严格意义上，不同分类之间的界限应该清晰，最好不要交叉重叠，这样才有利于归类、统计和行政管理，也有利于制定相应的政策。然而，现行的分类办法却难以应用到具体事务管理和政策，分类本身歧义就多，而实际管理中将哪

些高校划归哪一类别本身成本也会很高。谁来负责进行分类，用什么方式和办法分类；是高校自主归队，还是行政力量强行划分，抑或通过市场自发归类？这些都是难题。在现实政策与管理中，高校分类归类、分类管理或分类评价的背后，往往是权力和利益的分配与再分配，谁都不愿意在分类中失去权力和利益，因此，分类变成一种权力和利益的博弈。博弈的后果是成本的扩大和增加，无论是政府还是高校都在分类博弈中消耗宝贵的时间和精力。

如果分类的前提或者分类本身不合理不科学，教学评估合理性也会受到质疑。早在1994年教育部开展高校教学工作合格评估时，为了体现评估的科学性、针对性，当时把高校按照综合、理工、农林、医药、师范、政法等分为十个类别，分别制定不同的评估方案与标准进行评估，但实施一段时间后却发现，各科类院校在人才培养和教学工作上，其间的区分度并不很明显，其评估方案和标准过多反而人为地加重了评估的难度，增加了复杂性。所以，到2002年，十种不同的合格评估方案被废弃。不仅如此，合格评估与同时期的优秀评估及随机水平评估也被合并为一个方案，即后来统一的水平评估方案。

水平评估方案大大简化了教学评估工作，也提高了评估效率，降低了评估成本。但新的问题随之出现了。水平评估标准过于单一，没有考虑到不同高校之间的基础和差异，不仅造成高校相互攀比结果，而且导致高校办学模式同质化、划一化，消解了高校特色办学和自主发展，忽视高校不同利益群体多样化诉求。在水平评估即将结束的2007年，教育部成立新一轮评估课题研制组，无论课题组还是教育理论界，不约而同地提出分类评估的要求和设想。有学者提出按照层次分类，如“985工程”大学（39所）、设有研究生院的大学（56所）、其他“211工程”大学、博士学位授予权大学、硕士学位授予权大学、地方本科高校、新建本科高校等^[12]。也有学者主张，要按照研究型（创新型）、教学型、应用型对高校进行分类^[13]，但问题在于，这些

分类设想和研究成果难以操作，不仅造成教学评估方案研制与转化困难，更造成高校在归类归属上的争议与困境，谁也不愿意被归入低一个层次。

分类问题在学理上可以进行探索，但在实践中难以真正得到落实。如何既进行分类，又不使高校在分类中因担心歧视而导致选择困境？经过认真研究和讨论，审核评估在这一问题上做出有益的尝试，取得了重要突破，那就是不把分类作为预设的前提，分类不是预先由政府或评估机构提出，分类方案也不预先由政府或评估机构制定，然后要求学校作出选择。审核评估提出一个非常好的动议，即“用自己的尺子量自己”，把分类这一难题留给学校，让学校根据自身实际情况确定办学定位和培养目标，评估者只是根据学校提出的定位与目标，审度其定位是否符合学校实际与社会需要，是否有足够的资源支持，运行过程与教学效果是否达到了预期定位要求，说的、做的、做的结果是否能够相互印证，是否得到同行与社会认可。这便是审核评估提出的“五个度”标准。

审核评估较为有效地解决了高校分类问题。当然，这种解决是相对的，它是权宜之计，还是理应如此，确实需要人们继续进行深入细致的研究。

四、教学评估重点：从教学具体活动走向内部质量保证体系

第四代评估理论出现之前，西方教育评估即已经历了三个阶段，即测量阶段、描述阶段和判定阶段。测量也好，描述或判定也好，就其实质而言，共同特点在于把教育活动及其过程本身作为评价的对象和内容，通过评价证明教育活动过程效率与效果，用以改进教育教学活动，提升教育教学质量。

传统教育评估的理论总体上基于管理主义、科学主义和实证主义，总是相信可以通过那些所谓价值中立的科学客观的评估标准或评估工具来认识和把握教育教学活动过程及结果。这

一点，我们可从现实生活中各式各样的大学排行榜和绩效评价指标体系得出结论。就高校本科教学工作评估看，也没有脱离如此的理念和模式。至于将高校具体的教学工作和教学环节作为评价对象与评价内容，更是我国计划经济体制下所形成的政府注重过程管理和微观事务管理的产物。水平评估及其之前的所有教学评估形式，无不以具体教学工作为其对象，即这种评估评的是学校教学工作过程、环节、内容是否符合规定，是否达标，是否有效。每一种评估都有一整套具体明确的指标体系，不仅设置若干个一级指标，并且每一个一级指标又被划分为若干个二级指标，同样，二级指标下面又设置若干个具体观测点。一般说来，指标体系几乎涵盖或容纳了高校办学目标定位、培养模式、师资队伍、教学条件、专业设置、课程与教学、教学管理、质量与效果、就业等教学工作所有方面和过程。为了能够客观说明、展现与比较学校内部或不同高校间教学工作各方面环节质量与成效，评估方案又根据不同方面、不同环节在教学工作及在影响质量过程中的地位与所起的作用，赋予不同指标以不同的权重。评估政策制定者和实施者通常认为，指标越全面、具体、细致和深入，越能够通过数据或行为加以说明和描述，对教学工作及其效果和质量的衡量就越确切、越容易把握和控制，进而制定相应的针对性的政策与措施。以水平评估为例，该评估方案一级指标7个，二级指标19个（其中高权重的重要或核心指标11个），观测点44个，被评高校要预先根据这套指标体系和标准逐一对照本校教学工作，写出自评报告，说明每一方面和过程实际情况及其效果，尤其是要自我判断，自我做出优秀、良好、合格或不合格等级评价。然后，评估专家根据学校自评报告、自评等级和进校实地考察情况，给出评估结论建议和相应的评估等级。这个等级可能与学校自评等级一致，也可能有较大出入，或者说评估专家组可能认可学校自我评价，也可能更修正学校的自我评价。

水平评估虽然追求全面性、客观性和过程

性，以此来说明学校教学整体工作以及具体方面或环节的工作状况，以便更有效地发挥评估作用，然而水平评估也存在非常明显的问题和缺欠。一是指标体系过于单一，没有考虑到不同学校多样性发展的实际需要，导致评估区分度不高，忽视和制约了学校特色化办学；二是等级标准过于刚性，过于重视教学条件等硬件，学校出现了相互攀比和虚假评估现象，未能有效达到以评促建促改的目标；三是水平评估重视评价学校具体的教学工作过程及其效果，相当程度上限制了学校办学及教育教学的自主性和积极性，学校被动跟着评估指标体系运转，以期赢得理想的评估结果。在某种程度上可以说，水平评估中倡导的以评促建变成以评代建，以评促改变为以评代改，以评促管变为以评代管。

与水平评估相比，审核评估在评估内容和重点方面作出了重大改变。尽管在结构形式与内容上二者看似相近，但实质上却有着明显的变化。一是将水平评估指标体系改称为审核评估范围，一级指标改称为审核项目，二级指标改称为评估要素，观测点改称为审核要点。名称不同，实际蕴含着审核评估不再像水平评估那样过于强调指标体系的刚性规范，而是将评估指标更多的作为一种框架性、方向性和引导性要求，其弹性和柔性空间增大，以有效加强评估的针对性，充分发挥高校主体性和创造性，适应高校大众化和多样化发展需要。二是审核评估的重点不再仅评价具体教学工作及环节，尽管它仍然是评估的主要内容，但审核评估更重视强调评价学校内部质量保证体系建设。审核评估明确提出，要强化高校合理定位、强化教学工作中心地位和强化质量保证体系建设，这是水平评估未能顾及到的。在审核评估看来，只有学校自身建立健全了内部质量保证制度和体系，其教学和人才培养质量才是有保证的。所以，审核评估重点不在于评价学校某项具体工作做得如何，关键是看学校内部是否建立了有效的质量保证体系。

实际上，审核评估正式把质量保证体系作为独立的评估项目（相当于水平评估的一级指

标），项目之下设置质量标准、质量组织、质量监控、质量信息、质量反馈和质量改进等审核要素，所有这些构成了完整的高校内部质量保证体系，进而引导并统领着高校质量建设实践。将内部质量保证体系作为审核评估重点，体现了质量保障中学校的主体性，质量标准的系统性，资源配置的有效性，培养活动的针对性，质量措施与改进的持续性，尤其是体现了质量保障工作中不同利益相关者的参与性、互动性和建设性。

有研究表明，外部评估是否有效，是否能长期发挥作用，重要的前提就是与高校内部质量保证对接，或者说外部质量保障的目的在于积极推动和促进学校内部质量保证^[14]。当代西方国家大学具有悠久的自由自治传统，对外部的质量保障持批判态度居多，但尽管如此，欧美国家的外部质量保障却坚持上百年，尤以近三四十年为盛，其原因恰恰在于西方外部质量保障活动的目标和重点指向于高校内部质量保证制度和机制，也就是说内部质量保证体系是外部保障和评价的主要目标。美国六大区域认证中心实施的高等学校质量认证，英国高等教育质量保障署(QAA)开展的高等学校质量审核，澳大利亚国家高等教育质量保障局实施的质量评估，均以高校内部质量保证为重要内容。当前我国实施的审核评估即是将直接评价具体教学工作转移到评价高校内部质量保证体系建设，尽管因为我国高校内部质量保证体系建设无论在理论上还是实践上发展都还不健全不充分不平衡，使得这种转变尚不全面也不彻底，但无疑已经是非常大的进步，在引导高校内部质量保证体系建设方面发挥了十分重要的作用。可以预期，随着评估不断深入，高校质量保证体系建设将提升到新的高度。正如有学者指出，“未来评估的重点是审查质量保障体系的适宜性和充分性，核验其运行的有效性。质量保障体系由质量方针、质量目标、质量标准、质量要素、程序文件、质量监控、信息反馈及持续改进等部分构成”^[15]。

（下转第 25 页）

content_9740.htm, (2015-05-13) [2018-03-08].

[3] 徐晓飞, 丁效华. 面向可持续竞争力的新工科人才培养模式改革探索 [J]. 中国大学教学, 2017 (6): 6-10.

[4] 亨利·埃茨科威兹. 三螺旋创新模式 [M]. 陈劲, 译. 北京: 清华大学出版社, 2016.

[5] 陶铭, 曲超, 等. “新工科”背景下物联网工程专业人才培养方案探索 [J]. 现代计算机 (专业版), 2017 (21): 39-43.

[6] Jianfeng B, Hu L, Li Y, et al. The progress of CDIO engineering education reform in several China universities: A review [J]. Procedia-Social and Behavioral Sciences, 2013, 93: 381-385.

[7] Gemma Tejedor, Jordi Segalàs, Martí Rosas-Casals. Transdisciplinarity in higher education for sustainability: How discourses are approached in engineering education [J]. Journal of Cleaner

Production, 2018(175):29-37.

[8] Nieuwma D. Analyzing Context by Design: Engineering Education Reform via Social-Technical Integration [J]. International Perspectives on Engineering Education, 2015(20): 415-434.

[9] Potkonjak V, Gardner M, Callaghan V, et al. Virtual laboratories for education in science, technology, and engineering: A review [J]. Computers & Education, 2016, 95: 309-327.

[10] 古天龙, 魏银霞. 以新工科理念推动地方高校建设一流本科教育 [J]. 中国大学教学, 2018 (2): 32-35.

[11] 周绪红. 中国工程教育人才培养模式改革创新的现状与展望——在 2015 国际工程教育论坛上的专题报告 [J]. 高等工程教育研究, 2016 (1): 1-4.

[责任编辑: 夏鲁惠]

(上接第 11 页)

从水平评估到审核评估, 我国在高校教学评估理论、评估制度和评估实践等方面进行了有益探索, 取得了可喜进展, 当然, 这种进展是相对的、有限的, 具有鲜明的阶段性特征, 但由此也说明, 高校教学评估理论发展与评估实践探索不会有止境。

参考文献:

[1] 刘康宁. “第四代”评估对我国高等教育外部质量保障的启示 [J]. 国家教育行政学院学报, 2010 (9).

[2] 李志义, 等. 如何正确认识本科教学审核评估 [J]. 中国大学教学, 2012 (10).

[3] 刘振天. 回归教学生活: 我国新一轮高校本科教学评估制度设计及其范式变革 [J]. 清华大学教育研究, 2013 (6).

[4] E·格威狄, 等. 高等教育中的质量与问责 [M]. 毛亚庆, 等, 译. 北京: 北京师范大学出版社, 2008.

[5] 张应强, 等. 高等教育质量保障: 反思、批判与变革 [J]. 教育研究, 2014 (5).

[6] 吴岩, 等. 构建中国特色高等教育质量保障体系 [M]. 北京: 教育科学出版社, 2014: 51.

[7] 李延保, 等. 普通高校本科教学工作水平评估调查综述 [J]. 中国大学教学, 2006 (6).

[8] 高耀明, 等. 本科教学工作水平评估对高校教学工作影响的调查研究 [J]. 高等教育研究, 2006 (11).

[9] 王保华. 关于高校分类评估的若干思考 [J]. 中国教育政策评论, 2012 (11).

[10] 陈厚丰. 中国高等教育分类研究现状述评 [J]. 大学教育科学, 2010 (2).

[11] 潘懋元, 等. 从高等教育分类看我国特色型大学发展 [J]. 中国高等教育, 2010 (3).

[12] 钟秉林, 等. 坚持分类指导 制定分类标准 实施分类评估——新一轮本科教学评估基本问题探析 (二) [J]. 中国高等教育, 2009 (6).

[13] 李立国, 等. 建立以人才培养为基础的高等教育分类体系 [J]. 教育研究, 2018 (3).

[14] 郑冕. 高校内部质量保障: 框架与措施 [J]. 中国高教研究, 2016 (9).

[15] 张安富. 本科教学工作审核评估的再认识及持续改进 [J]. 高教发展与评估, 2018 (3).

[责任编辑: 周 杨]