

doi:10.3963/j.issn.1672-8742.2015.01.01

审核评估:如何真正做到正确认识,确保平稳开展

袁益民

(江苏省教育评估院,江苏南京 210024)

摘要:目前,审核评估工作在试点基础上即将在全国全面展开。然而,在学界、院校甚至省级组织实施部门解读教育部的顶层设计方案和具体理解审核评估的理念、价值和方法时,仍然存在诸多误区。一方面,体现审核评估核心理念和价值的一些重要问题还没有真正做到正确认识,在为何评、谁来评、评什么等问题上还存在明显的误读。另一方面,对审核评估这样一种模式到底是什么套路和做法缺乏真正的了解,在应该做些什么、怎么去做、如何才能做好等方面还存在明显的曲解。因此,需从问题聚焦、角色定位、价值导向和方法使用等四个方面加强对审核评估的本体性研究,对审核评估所聚焦的内部质量保障这一核心问题、多元主体这一基本定位、赋权增能这一价值导向和综合系统这一方法等,真正地做到正确理解,以确保审核评估的平稳开展。

关键词:审核评估;评估理念;角色定位;价值导向;评估方法

中图分类号:G40-058.1 **文献标志码:**A **文章编号:**1672-8742(2015)01-0001-16

一、引言

《教育部关于普通高等学校本科教学评估工作的意见》和《教育部关于开展普通高等学校本科教学工作审核评估的通知》的发布,为我国普通高等学校本科教学工作审核评估的组织实施,提供了一个很好的顶层设计,方案较好地反映了国内外教育评估改革的大趋势,与国际评估发展的方向基本吻合。然而,在评估试点、业界讨论以及学术研讨过程中反映出来的一些重要问题需要予以特别关注和进一步明确。一方面,审核评估的核心理念、基本要素乃至整体思路的有些重要方面还没有很好地进入审核评估的工作视野,如伙伴基准比对与外部参照点使用、学业风险框架与主题评审、正式的预访程序、分层次审核结论标识及其评审细则、三角鉴定和痕迹跟踪等技术的使用、学生参与(担任评审员或提交书面意见书等)、中期跟踪、进展报告制度、最佳实践识别与传播等。另一方面,也有一些认识的误区和潜在的风险已经露头,如果处理不当,将在不同程度上影响到项目实施的效果,如认为审核评估“不分类型、不设

标准、不做结论”，认为评估报告只包括“值得肯定、需要改进和必须整改”等三个方面的意见而无需给出达标情况的描述和对于总体情况的综合判断，认为专家需要“把脉、诊断、开处方”，审核评估是“基于目标、基于自评、基于事实”等。应该看到，参与相关讨论的许多人的观点是不符合院校评估改革趋势的，也与国际通用的审核评估模式的基本定位相违背的。如关于审核评估“不分等级、不设标准、不做结论”的提法，就不符合国际审核评估模式本身的定位和多国开展该项评估的实际做法。有些学者认为，这“三不”是为了减轻评估压力和功利性倾向，但事实是，压力是评估这个行当与生俱来的，功利性问题本来就是在各种模式的评估中都要通过科学人文的设计去尽量规避，而不是不分类型、不设标准、不做结论就能解决的。不分类型就没有攀比之心、不设标准就没有功利倾向、不做结论就能没有评估压力的简单化思维方式，其实与上轮评估的有了统一标准和分等结论就能促进建设的思维在本质上是一样的。可见，如果这轮评估的具体运作过程中出现偏差，很可能就是在这些问题上。

其实，对于审核评估模式的选择是由评估目的所决定的，而选择这一关注内部质量保障的院校评审以后，许多关键的要素应该是相对比较确定的，与国际通行的普遍做法在基本面应该还是比较一致的。但是，目前这些要素和做法在我们的具体工作方案中还残缺不全，甚至有严重的曲解。而且，至今尚未有人对这一涉及审核评估未来走向的核心问题作出系统的梳理和澄清。如果对审核评估的理解一开始就出现大的偏差，且又不去及时地纠正，那么，尽管本轮评估已经有了一个比较好的顶层设计，但在具体的执行和操作过程中，仍然很容易出现不可预料的风险。从以往的经验来看，如果不能及时地识别和规避评估项目的潜在风险，项目将不能得到很好的实施，也就更不可能较好地服务于院校通过识别和规避学业风险切实保障教育教学质量的核心任务。

与此同时，在与院校同仁进行座谈和调研的过程中我们也发现，目前对于审核评估如何正确认识和平稳开展更是存在一些模糊理解，甚至可以感觉到其明显的茫然和一定的顾虑，这些都增加了未来项目实施过程中的不确定性。如有的认为，审核评估就是分类评估中的一种评估；而有的则认为，审核评估不再是强调院校分类的评估，甚至同类型同层次院校的情况也与被评院校无关了。有的认为，院校的目标定位而非计划路径才是审核评估的起点；而有的则认为，若干个“度”而非整个内部质量保障机制是关注的焦点，甚至内部质量保障机制就体现在那么几个“度”。有的认为，审核评估没有评估标准，只有审核范围；有的则认为审核范围就是审核评估的标准，甚至是最主要的基本标准。有的认为，每个学校自身的标准的达成情况是最重要的；有的则认为，一个省的工作重点和发展主题才是最要紧的，甚至把应用型人才培养定位等专题作为整轮评估的主题；有的认为，状态数据应该突出关键绩效指标数据，有的则认为状态数据越细越好、越全越好，甚至国家级数据也要如此。有的认为，在评估前、评估中和评估后，专家的进校活动应该常态化，应允许突访性考察；有的则认为，多进校就是多干扰学校，甚至境外普遍正式采用的预访程序也不得安排。有的认为，写实性评估报告除了肯定院校的优点，还要就需要改进和必须整改的方面开出处方；

有的则认为,评估工作不再需要等级的描述,甚至评估者内部也不需要赋分的过程和优劣的标识。有的认为,在强化院校自评、状态数据监测和发布质量报告的前提下,应该只向被评院校或在一定范围全文公开写实性评估报告;有的则认为,应向社会无条件全文公开报告内容,甚至事先不用征求被评院校的意见,也不要有任何保密条款。有的认为,评估不分等级后,学校会松懈了,师生的参与也难以保证了;有的则认为,如果在内部质量保障机制上动真格、下真功夫,评估服务的重心必然进一步下移。有的认为,现场访问期间要加大校外利害相关方满意度调查;有的则认为相关的调查可以留给学校自己平时去做,甚至现场访问过程中不需要关注满意度。有的认为,除了评后的信息传播和经验交流,应该增加后续评估程序以强化评后整改;有的则认为发布写实性报告后,评估工作应告一段落,甚至将内部改进的工作完全交给学校自己,无需再有进展报告。有的认为,评估机构应在“管办评分离”的原则下实现相对于行政主管部门和国家级评估机构的独立性;有的则认为,行政主管部门和国家级评估机构要进一步加强监管,甚至在一些方面仍包办代替。从方案整体上看,有的认为,审核评估作为全新的评估模式可以避免以往评估的所有弊端;有的则认为换汤不换药,传统评估的痕迹难以根本去除,甚至认为,到头来这无非就是另一轮评估而已,不会带来任何根本性改变,等等。

基于目前围绕该评估方案的现有研究,仍有必要加强对审核评估模式的本体性研究,对该项评估的问题聚焦、角色定位、价值导向和方法使用等四个基本项目构件作出系统的梳理和澄清,以真正做到正确认识,确保其平稳开展。其中,问题聚焦需要回答评估的最主要目的是什么、要解决哪个最核心的问题、最本质的评估性问题是怎样的等元问题;角色定位需要回答谁得益(为了谁)、谁来评(谁为主体)、谁参与(以何身份)、依据谁的标准、谁决定结果、结论告知谁、谁使用评估结果与结论等元问题;价值导向需要回答使用哪种或哪几种评估定位(问责、发展、学习等)、是强调自我发展还是强调外部帮助、以何种标准去评估、方法多些科学性还是多些人文性、是倡导独立性与客观性还是鼓励多元参与、结果与结论的利害度高些还是低些、评估信息的公开性大一些还是小一些、是用科学的语言来向大众突出其可信度还是用大众的语言来增加科学评估的使用性等元问题;方法使用则需要的问题聚焦、角色定位和价值导向确定后选择相应的方法,仍需要回答重质性分析还是重量化判断、重事实判断还是重价值引领、重过程还是重结果、重外部数据还是重校本数据、重专家还是重科学工具等元问题^[1]。如果这些元问题在方案制订和组织实施过程中能得到基本的回答,并得到院校的支持和配合,真正做到正确理解和平稳开展,就能够给本轮评估带来实质的改变,为学校教育教学质量的改进提供切实的保障。

二、问题聚焦:关注内部质量的保障

审核评估的核心问题是院校内部质量保障机制(包括管理标准及质量工作等)的综合性、完整性、透明性和有效性,包括计划的可行、组织的健全、实施的严密、措施的

有力、活动的活跃、方法的可靠、信息的准确和结果的有用等,立足点是绩效表现的整体改进。围绕这一核心问题的评估性问题,是对于质量的状况和质量保障是否有效是怎么知道的,具体包括目标路径是否明确、执行措施是否落实、结果是否是目标路径和执行落实的后果、如何改进是否知道等四个方面,见图 1 所示。

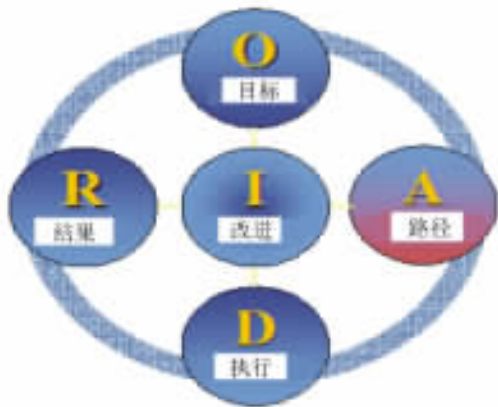


图 1 澳大利亚院校质量评审的 OADRI 模型^[2]

院校内部质量保障机制、能力及体系应该作为一个整体来看待,不能相互分离或者把它们与教育目标的制定、教学活动的过程和学生学习的结果割裂开来。内部质量保障的运作安排、组织水平和制度设计是相互关联的,以其所说、看其所做、评其所获也必须是一个连贯的过程。根据迈西(William F. Massy)的质量评审流程图,见图 2 所示,在投入、教与学的过程和学习结果三个要素中,正向的箭头刻画投入是如何激活教与学的过程的,教与学的过程随后又是如何产生出学习结果的。但是对于质量评审这样一种复核活动来说,反向的或者说“反馈”的箭头是最重要的。为了保障教育质量,如箭头 A 所示,教师必须持续地测量学习结果的质量,将其与目标加以比较,然后按照需要调整教与学的过程。同时,如箭头 B 所示,过程的自我反思以及同校内外最佳实践案例的比较,也能带来教与学的过程的调整。最后,如箭头 C 所示,过程的调整会促发所需投入的类型、数量和质量的变化的。离开反馈的过程,质量是注定要下降的,反馈是维持质量的基本保障。

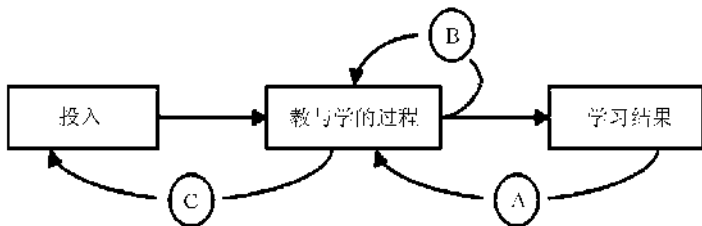


图 2 教育质量评审流程^[3]

境外的院校质量评审范围各有侧重,有的更多地关注质量工作的覆盖面,有的更多地关注质量活动的有效性,有的更多地关注质量的后续改进,但无一不是对于质量工作和质量保障体系的一个综合的、整体的架构,强调各项质量活动之间的关联性和

整合一致,如表1所示。尽管这些院校质量评审的范围略有不同,对于质量保障机制或质量工作的理解也略有差异,但是,在聚焦质量保障机制或质量工作这一点上是完全一致的。

表1 多国院校质量评审的标准对照

| 国别 | 质保机制与程序 | 质保的参与 | 评价与信息管理 | 目标 | 学习机会 | 复核与后续 |
|-----|---|--|--|--------------------|-----------------------------------|---|
| 丹麦 | 质量工作的策略与程序;质量工作的覆盖面和组织 | 内部利害相关人的参与;外部利害相关人的参与 | 公共信息;信息系统与数据收集;信息与数据的使用;对学生的评价 | 质量目标 | 学习资源与学生支持 | 专业和学位的批准、监测与定期复核 |
| 瑞典 | 质量保障过程的执行策略;院校所有活动中质量工作整合一致 | 与利害相关人的合作;教职员参与质量强化与保障 | 外部专业关系 | 学业领导力 | 国际化;性别平等 | 评估与后续活动 |
| 芬兰 | 质量保障的综合性;质量保障体系与管理指导的相互作用;文档,包括质量保障政策的制订和程序、责任人及责任的定义 | 教职员、学生和外部利害相关人参与质量保障;质量保障信息对外部利害相关人的适切性和可获得性 | 质量保障体系生成信息的使用;质量保障信息适切性和可获得性 | 质量保障体系的目标、结构和内部一致性 | | 质量保障体系的监测、评估与持续发展;质量保障程序及结构的效率 |
| 挪威 | 教育质量工作如何成为院校战略性工作不可或缺的部分;质量工作如何贡献于资源管理和院校优先领域(人力,基础设施,服务);如何与各层指导和管理相连接 | 质量工作如何日常组织,测量保证广泛参与,有工作不同阶段明确定义的职责和权威的分配 | 如何系统提供对于质量工作目标达成情况的信息和评价的分析;院校如何找回和处理为对所有学习单位作出令人满意评价所必须的数据和评估信息,信息如何在高层得到积累 | 院校质量工作的目标是如何定义的 | 体系如何保证聚焦于整个学习环境和学生积极参与质量及整个学习环境工作 | 给董事会年度质量报告如何给出教育质量的一致总体评价和持续质量工作计划和测量的概览;使用质量工作结果发展学习质量 |
| 加拿大 | 质量保障机制的有效性 | 质量保障机制的实施 | 学生成就评估 | 战略性规划;成功的规划 | 专业 | 保证质量保障机制有效性的机制的复核与更新 |
| 英国 | 院校质量强化的路径 | 院校对研究生的安排 | 公开的信息 | 院校对学业标准的管理 | 院校对学习机会的管理 | 合作方面的安排 |

目前,围绕审核评估的关键或重点问题的基本提法,包括突出内涵建设,突出特色发展,强化办学定位,强化人才培养中心定位,强化内部质量保障体系建设,以及所谓“四个度”、“五个度”等。这些无疑均是我国高等教育内涵建设和质量提升过程中的最基本的问题。但是,无论是从国际的普遍实践来看,还是从我国强化内部质量保障以支撑现代学校制度完善的紧迫任务来看,院校内部质量保障机制应该是审核评估的核心问题。

鉴于某一项评估总有其适用范围及局限性,唯有专注于相对集中的范围而非面面俱到,才能在相关领域有切实的、更多的收获。当然,任何评估在其所涉及的目标之外会有更多的副产品,但是这些产品只是副产品,往往不是刻意而为就会自然产生的,倒往往是聚焦核心问题之后无心插柳柳成荫的结果。作为本身关注内部质量保障的院校质量评审的审核评估,无疑将有利于强化办学定位和人才培养中心地位,并促进学校内涵建设和特色发展。所以,审核评估的关注点应始终聚焦院校内部质量保障机制这一核心问题。参与内外部质量保障的各方利害相关人应就此达成共识。其实,教育部相关文件也已明确,“审核评估核心是对学校人才培养目标与培养效果的实现状况进行评价”,也就是对内部质量保障的有效性进行评价。从广义上说,现有审核评估的范围本身就是对内部质量保障机制进行考查的基本维度,而其中的“质量保障”审核项目,就是对狭义的内部质量保障机制专门进行考查的特定维度。

三、角色定位:发挥多元主体的作用

在角色定位方面,如何发挥多元主体的作用,让各方利害相关人参与到评估工作中来,是评估工作得以平稳开展的关键。其实,教育部相关文件已经对多元参与和多元评价提出了明确的要求。麦克贝斯和麦克格林(John MacBeath and Archie McGlynn, 2002)提出的评估使用者的重要性排序调查表,可以为我们理解如何发挥各方利害相关人的作用提供了参照工具,如表2所示。

表2 评估对评估使用者(服务对象)的重要性程度排序^{[4][11]}

| 使用者(服务对象) | 根据重要程度从1排到7 |
|-----------|-------------|
| 为了家长 | |
| 为了主管部门 | |
| 为了学校督导人员 | |
| 为了学生 | |
| 为了教师 | |
| 为了当地社区 | |
| 为了学校领导 | |

注:为了适应中文的使用习惯,对原文有个别非本质性的调整。

当然,利害相关人在评估中的利益和权利大小本身是各不相同的。相关研究认为,如图3所示,评估对于作为参与者的评估对象来说是高利益、低权利的;对于作为玩家的评估者来说是高利益、高权利的;对于作为背景设置者的政府或立法者等来说是低利益、高权利的;而对于作为旁观群体的公众和社会媒体等来说是低利益、低权利的。一项精心设计的有效的评估应该在各方利害相关人之间建立良好的平衡。目前的天平是从作为参与者的学校、师生、社区、家长向作为背景设置者的政府主管部门及作为玩家的评估人员过度倾斜的。

| | |
|-----------------|------------------|
| 高利益 低权利 参与者 | 高利益 低权利 玩家 |
| 旁观群体 低利益 低权利 | 背景设置者 低利益 低权利 |

图3 利害相关人的利益与权利关系比较

根据帕顿(Michael Quinn Patton,2009)的研究,如图4所示,系统的复杂性与技术的确定性和社会的一致性之间成反比的关系。由于教育系统的繁杂性,对于教育质量的技术方面的确定性和达成社会共识的难度均是比较大的,因此在项目实施中,院校、评估机构、社会、政府的互动关系就显得极为重要。一方面,我们要确保作为质量主体的参评学校与作为评估专业服务提供者的评估人员之间的良性互动关系的建立,充分发挥院校自身的质量主体作用及评估机构的专业服务的功能;另一方面,我们最终要落实多元主体的作用。

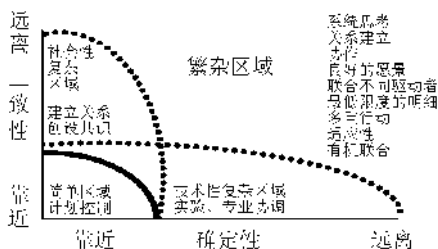


图4 系统复杂程度对于社会一致性及技术确定性的影响

“约哈里窗”给我们的启示是:只有院校的主动告知及公开与同行的询问及反馈结合起来,我们才能对学校教育的质量状况有更多的了解,如图5所示。院校层面的质量问题太繁杂、太微妙了,院校走进“顺从模式”从而扣留、扭曲或者无法提供证据的机会太大了。我们需要进一步完善现代学校制度,落实现代学校教育的管理分工,明确政府、学校和社会的“铁三角”关系。政府主要管好教育公平,学校作为办学主体办出质量,社会负责相关保障(包括法律保障、舆论监督和社会服务等),实施管办评分离,进一步强化评估工作的独立性、专业性和权威性,评估工作本身应围绕以学校为主体、以质量为核心、以服务为导向的宗旨进行改革,着力提高教育评估的专业服务水平,为咨询政府决策和支撑学校自主发展作出贡献^[5]。



图 5 “约哈里窗”(The Johari Window, Luft & Ingham)

我们需要调动各相关方面的积极性和创造性,以更好地达到我们所设计的评估目标。不仅要充分发挥广大师生在参与质量保障过程中的主体作用,而且要发挥省级行政部门的统筹作用和评估机构的专业服务功能,还要特别注意发挥同行专家、行业专家及用人单位等多方面的作用。审核评估应关注利害相关人的能力建设,通过参与评估来提升其介入质量保障的自觉性和能力水平,使其成为质量的守护和变革的使者。

四、价值导向:强化赋权增能的功能

对于内部质量保障而言,机制建立是起点,体系完善是归属,能力建设是核心。作为关注内部质量保障的院校评审模式,审核评估应特别关注学校内部的质量保障方面的能力建设问题。

首先,当我们的学校具备基本合格的条件以后,学校需要的是在更少外部干扰情况下找到自我定位,其办学自主权和质量的拥有权得到尊重,使其能够自主错位发展、特色创新发展,而政府和社会应该通过服务为其赋权增能。

其次,内部质量保障的过程也是院校内部知识管理的过程。根据德尔菲集团(Delphi Group)的调查,如图 6 所示,法人团体中的知识有 42% 存在于职员大脑,其余分别存在于电子知识库(12%)、电子文档(20%)和纸质文档(26%)。作为知识分子集聚的高等学校,广大师生及管理者的的大脑是相关知识的主要“存放处”。对于这样的知识的管理不能靠简单的信息获取手段来进行,而是要通过人的自觉意愿并连接人的能力才能实现。

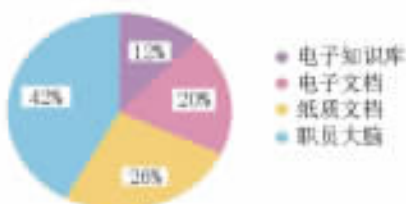


图 6 法人团体中显性知识和隐性知识的比例

同时,对于麦克贝斯和麦克格林(John MacBeath and Archie McGlynn, 2002)提

出的评估具有问责、知识和发展三个视角的理论,如表3所示,我认为后两者本质上均属于学习的视角,都与知识的积累以及组织性学习(Organizational Learning)有关。

表3 评估的三个视角^{[4]13}

| | 问责视角 | 知识视角 | 发展视角 |
|------|----------------------|-------------------------|-------------------------|
| 评估目的 | 提供绩效表现、有效性和是否物有所值的数据 | 生成领导力、风气、学与教等方面哪些相关的新见解 | 强化学校的自我改进能力 |
| 服务对象 | 公众和家長 | 学校管理者和教师 | 老师、学生、家长和领导 |
| 参与人员 | 学校管理者 | 教师、学生、学校管理者 | 教师、学生、家长、后援人员、学校管理者 |
| 内外关系 | 由自评数据支撑的终结性外部评估 | 主要通过自评的诊断性质的评估 | 主要通过外部有判断力朋友支持的自评的形成性评估 |

注:为了适应中文使用习惯,对原文有个别非本质性的调整。

因此,最本质的评估定位其实就是问责和学习这样两种。这两种定位在不同价值追求、管理目标、关注焦点和主导机构下,对于评估内容、方式和动力来源等都存在完全不同的路径依赖,也将产生完全不同的评建效应,如表4所示。当然,不能把问责的定位与学习的定位完全对立起来,问责和学习都是为了带来改进。不过,前者是促发改进,而后者是诱发改进;前者带有更多的压力和强制性,而后者更加偏爱同行之间的平等交流和坦诚对话。既然你是来问责我的,那么我们之间就有一种天然的冲突关系,我自觉找问题就是主动找麻烦,所以暴露问题不仅不是我的职责,而且是我的忌讳。

表4 2种评估定位的不同路径依赖

| 评估定位 | 追求价值 | 管理目标 | 关注焦点 | 主导机构 | 重点内容 | 背景条件 | 工作方针 | 路径选择 | 使用方式 | 动力来源 | 效果影响 |
|------|------|------|------|------|------|------|------|------|------|------|------|
| 问责定位 | 服从 | 执行 | 当前 | 政府 | 外在 | 不用 | 以评 | 施加 | 逼你 | 外部 | 恐惧 |
| 学习定位 | 自主 | 提升 | 长远 | 评估 | 内涵 | 作为 | 以评 | 提供 | 赋权 | 自身 | 渴望 |
| 发展定位 | 发展 | 能力 | 发展 | 双方 | 建设 | 前提 | 助建 | 支持 | 增能 | 意愿 | 学习 |

在教育这样一个繁杂系统中,针对质量工作这样一个极其复杂的评估对象,封闭和隐瞒自然显得更为简单和安全。问责和改进的压力往往会带来被评者的消极应对,而指出结果的缺陷常常也并不能带来过程的变革,甚至一些短期的好的结果反倒会掩盖许多长期的不可持续的问题。因此,如何通过强化学习功能为质量的主体赋权增能才是问题的核心。作为教师这样一个依靠自觉性开展工作的群体,让他知道改进的需要就已经足够了,让他知道如何改进则是更高的要求,要考虑他的资源条件以及能力和意愿。因为,改进总会涉及教育的某种变革,涉及关于人、为了人的变革,而对于教育变革实践者的评价是个十分复杂的问题,如图7所示。教育质量的提

升需要耐心,更需要尊重。学校需要检查,更需要信任,缺少检查容易助长自以为是,缺少信任则会造成自欺欺人。问责比学习的定位更加容易带来抵制和欺骗的行为或者顺从和应付的文化,更加容易造成犬儒主义的愤世嫉俗或玩世不恭。总之,学习之路并不会比问责之路更为漫长,却可以带来质量主体更为积极主动的、更有意义的实质性的改进。说到底,内部质量保障的自我完善本身总是能够比对于这一机制的评估带来更多的关于改进的信息,只不过前者更加关注实质性的改进建议和实际的改进效果,而后者更多地聚焦院校内部的改进意愿和改进措施。

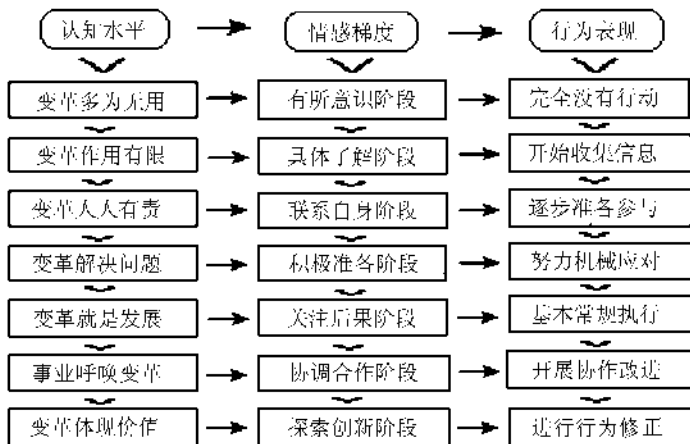


图 7 教育变革实践者的评估模型

审核评估以学习定位支持院校的能力建设,它“基于标准、基于计划、基于证据”,为参评院校诊断把脉、赋权增能,是一个同行驱动、背景参照、基于证据的复核过程。目前学界普遍认同的“基于目标、基于自评、基于事实”的特点并非审核评估的自有特点,而是对于其他评估也同样适用的对所有评估的基本要求。审核评估的基本特点倒是应包括“基于标准、基于计划、基于证据”。对于内外部的质量标准、院校自身的质量活动计划安排的强调以及学校对于质量工作的自我搜证和举证、利害关系人的佐证和旁证、外部同行专家的取证和查证等,是审核评估的诸多亮点。

审核评估不设标准的说法是不符实情的,现在的审核范围就是一种标准。各国对院校评审的标准有多种称呼,如: criteria, standards, auditing targets, areas, aspects, factors 等。2005 欧洲质量保障网络 ENQA 的质量保障标准及准则中明确规定:“作为外部质量保障活动的结果所作出的任何正式决定,应当基于明确的、公开发表的、连续引用的标准。”各国的院校质量评审均依据各种各样的内外部评审标准,如英国、澳大利亚便依据多种标准。英国高等教育质量准则(the UK Quality Code for Higher Education,在 2012-2013 年前称为 Academic Infrastructure),规定了英国高等教育必须达到的“预期”(expectations),其关注的核心问题包括:每个学生得到公正、平等的对待;学生有机会参与设计自己的学习体验;与学生学习相关的事,要及时主动告知学生;与学生学习相关的所有政策和流程要明确、透明;对学术标准和质量的战略监管位于教育机构学术管理最高层次;所有政策和过程得到常规而有效的监

督、评估和改进;为保持学术标准和学习条件,要有足够的、合适的外部力量帮助和监督;职员发展受到重视,从而使他们能够更好为学生学习提供服务。英国的质量保障署(QAA)在院校评审中使用的参照标准,包括高等教育资格框架(Framework for HE Qualifications)、学科基准陈述(Subject Benchmark Statements)、QAA的行为守则(The QAA Code of Practice)和专业规范(明细与规格, Programme Specifications)。澳大利亚的第三级教育质量与标准局(TEQSA)的前身澳大利亚大学质量保障署(AUQA)除了参照院校的目标,还设置了简称为QAFs的质量评审要素(The Quality Audit Factors)标准,其中每一个要素下面包含范围和聚焦点,然后再分若干方面设置复核标准(Criteria for Review)。同时,AUQA还明确了包括国家高等教育批准程序议定书、院校得以认可所依据的任何法令或法规中的目标规定、2003年联邦高教支持法和相关法规以及包括其他相关的立法在内的各种外部参照点(External Reference Points)。

“基于计划、基于证据”也是审核评估的显著特点。其实,教育部相关文件中的主体性、目标性、多样性、发展性和实证性等五项基本原则,就是在传达两个非常清晰的指导思想:一是审核评估对于作为质量主体的学校的已经计划好的质量目标,应特别关注、特别尊重,使其成为评估的出发点、基本导向和最终的依归,这样才能体现不同学校的多样性的定位,确保学校的自主、多元、特色、创新发展;二是评估者应与评估对象结成学校发展中的伙伴关系,通过学校自身的举证和外部评估者的查证,帮助学校共同收集关于质量工作的证据,为学校提供专业性的服务,支持学校强化内部质量保障的机制。所以,主体性应该解读为学校自身可以自我制订自己的自主发展的质量计划;目标性就是把学校自己已经计划好的质量目标作为评估的起点和基本导向;多样性就是只有尊重了学校自身的自主发展的质量计划、坚持了学校自己已经计划好的质量目标以后,学校才能有体现多样性定位的自主、多元、特色、创新发展;发展性就是评估者与评估对象把质量问题搬上桌面来一起讨论,为学校的质量计划及质量目标服务;实证性就是指出评估工作是基于证据的事实描述、价值判断和同行交流,是专业性的、规范性的工作,应遵循有多少证据说多少话的原则。

五、方法使用:突出综合系统的方法

在方法使用方面,要解决如何评的问题,除了问卷调查、访谈、文档复核、观察、专题组和个案研究等常见技术工具的使用外,审核评估应采取分校立项、分类参照、分项评审、分层举证的实施办法,特别重视三角定鉴和痕迹跟踪等方法的使用。对于教育质量这样的繁杂问题,我们需要从多维度认识质量的实际状况,以多方面证据认定同一事实,包括使用多种评估方法得出一幅完整的学校图景,如图8所示。

各种评估方法具有各自不同的优势和劣势,均有一定的诊断功能,也有其固有的局限性。审核评估中的有些方法就类似于医疗诊断,适用于特定的诊断用途,如表5所示。



图8 使用多种方法拼出一幅完整的图景

表5 审核评估的“望、闻、问、切”

| 评估手段 | 粗略地近似的诊断手段 |
|----------|------------|
| 状态数据分析报告 | 称体重 |
| 自评报告审问报告 | 胸透 |
| 问题单、问卷调查 | 把脉 |
| 学校相关文档复核 | 问病史 |
| 校园氛围课堂观察 | 察颜观色 |
| 一定范围深度访谈 | B超 |
| 抽查试卷(论文) | 抽验血 |
| 专题小组研究讨论 | 会诊 |

在方法使用方面,审核评估要特别注意通过不同参照系的使用体现先进的多元评估理念,全面做好主体目标导向、通用标准参照、院校背景分析、同行伙伴比对的工作,如表6所示。其中,参照相同层次和类型学校进行的基准比对工作不应被忽略。

表6 审核评估工作的多元价值追求

| 价值取向 | 解决问题 | 追求目标 |
|------|------|------|
| 常模参照 | 已然 | 水浅水深 |
| 标准参照 | 必然 | 水落石出 |
| 背景参照 | 偶然 | 水到渠成 |
| 目标参照 | 应然 | 水涨船高 |

根据教育部和学界普遍认同的定义,分类评估是指:在院校分类的基础上,“用不同评估标准,对不同类型的评估对象进行评估的方法。如在高等学校的教学评估中,根据不同层次、不同类型的学校,设置不同的体系、标准,运用不同的方法进行评

估。”^[6]根据大家的这一共同认知基础,审核评估不属于分类评估或者分类评估中的一种,而是针对所有“层次和类型”的学校所使用的一种独立的评估模式。

然而,审核评估不搞分类评估,并非评估不需要分层次和分类型进行常模参照。相反,同层次、同类型院校之间的基准比对与院校内部的基准比对一样,是十分重要的,如表7所示^[7]。外部基准比对可以帮助识别自身差距和同行中的最佳实践案例。基准是改进工具,也可以用于展示基于外部参照的适当标准的达成。当这种确立分析基准的做法被用来改进院校的过程和系统以更好地实现其教育使命时,院校还能够更好地抵御那些对其具有扰乱性甚至破坏性的外部目标和标准。基准可以分为部门基准(整体或某项)、通用基准(跨行业)和最佳实践基准等。基准比对包括内部比对、竞争性比对、功能性比对、通用性或同类最佳式比对。有的专家认为,基准比对包括战略性、绩效或竞争性、程序性、功能或通用性、外部性、内部性、国际性、常识性基准比对(Wyld,2010)等。国外院校质量评审中广泛开展同类院校之间的基准比对,广泛使用各类院校及分项评价结果,除了资质框架、各种准则(指南)、绩效表现指标外,还普遍使用其他的外部参照点。

表7 基准比对的定义

| 定义人,定义时间 | 基准比对的定义 |
|-------------------------|--|
| Kempner,1993 | 通过外部关注内部活动、功能和运作,测量和比较一机构与另一机构工作程序的连续的系统过程 |
| Dervitsiotis,2000 | 与竞争者和其他某一特定功能上的最佳者就关键绩效表现指标进行的系统研究与比较 |
| Epper,1999 | 相互学习,改变你所做的和模仿成功行为的一种系统方式 |
| Rizk & Al-alusi, 2011 | 通过审视其他学校的程序和模式并调适使用他们的技术和方法,改进院校的行政程序和教学模式;测量和比较一机构与另一机构的工作程序;对内部活动、功能和运作引入一种外部的关注 |
| Woodhouse,2000 | 内部基准比对出现在一个组织内部与另一个部门作比较的时候 |
| Southard & Parente,2007 | 当一些环境中缺少外部基准比对努力的某种复杂性时,提供一种开始仔细、严谨地识别基准比对执行障碍的受控制的基础 |
| Meade,1998 | 基准比对是搜寻导致卓越表现的最佳实践的正式的、结构化的过程,是关于它们的信息的观察和交流,满足其自身组织需要的调整以及修补性实践的执行 |
| Jackson & Lund,2000 | 一结构化的学习过程,以使那些参与此过程者比较他们的服务/活动/产品,以识别它们的比较优势和弱点,作为自我改进和/或自我管理的基础 |
| Benc,2003 | 基准比对是一寻找、调整和执行出色实践以取得优越表现的经营卓越工具 |

具体的基准比对工具,包括课程与教学评估、出口(毕业)调查、课程档案报告、学业咨询调查、校友调查、雇主调查、培训评估和毕业设计。在学业领域,可以对学生学习、教员教学、课程、研究与培训、社区服务和资源等方面进行基准比对。可以使用学生学习的关键绩效表现指标,譬如:新录取学生数(国别,性别)、巩固率、录取80%以上学业水平的比例、学生的满意率、学生的培训与研究情况、学生的学业与社交表现情况等。也可以使用教员教学的关键绩效表现指标,譬如:教员专业发展活动、教员

离职率、预期的学生课程选择率等。在我国本科教学工作审核评估中,应该尤其关注教授为本科生上课和教学投入等情况。

为了整体把握院校内部质量保障的基本状况,审核评估还应针对教学工作建立学业风险评价的框架。澳大利亚的第三级教育质量标准署(TEQSA)就建立了风险评价框架,其主要构件包括对于学业和财务状况的总体风险评价以及四个关键风险领域(监管历史及身份;学生负担、体验和结果;教员概况;财务可行性和可持续性)。风险指标包括完成学业人数、学生负担、未完成学业比率、升学率、完成学业比率、学生满意率、毕业去向、高层次教员带头人、生师比、临时教员、财务可行性、财务可持续性、识别的其他风险等。TEQSA的前身AUQA发现的学业风险包括:课程不是及时更新的;评价不测量相关的学习;不同校区和地点的质量和学术标准水准不一致;给海外师生提供的支持和服务水准不够;新的课程提议由机会而非战略指向驱动;学年制教职员缺乏专业发展;研究型学生缺乏学术交流的空间;对行政和设施的关注大于教学能力等。

综合系统的方法特别需要充分地体现在审核评估的结论标识上。审核评估不做结论的说法是不符实情的,现有方案中也是包含了评估结论和结果的。境外院校质量评审的结论意见还有多种等级标识,有的作为专家内部讨论和评价的尺度,有的直接对外公布,有的是表达各项评估标准的达成度,有的是显示整体的绩效表现水平,有的反映质量保障机制的成熟度,有的代表对质量工作(质量管理可行度和可靠性)的信心水平,如表8所示。

表8 境外院校质量评审结论意见的多种等级标识

| 国别 | 第一等级 | 第二等级 | 第三等级 | 第四等级 |
|------|------------------------------|--------------------------------|--------------------------------|--------------------|
| 芬兰 | Advanced 高级阶段 | Developing 正在发展 | Emerging 开始出现 | Absent 仍然空缺 |
| 瑞典 | Developed 达标度:发达阶段 | Under development 达标度:处发展之中 | Initiated 达标度:业已启动 | |
| | Full confidence 满意度:充分信心 | Confidence 满意度:有信心 | Limited confidence 满意度:有限信心 | |
| 英国 | Broad confidence 满意度:广泛信心 | Limited confidence 满意度:有限信心 | No confidence 满意度:没有信心 | |
| | Desirable 建议:理想层次 | Advisable 建议:劝告层次 | Essential 建议:必要层次 | |
| 澳大利亚 | Commendations 赞扬推荐 | Affirmations 证实肯定 | Recommendations 劝告提议 | |
| 加拿大 | Ensure 保障 | Generally Ensure 总体保障 | Partly Ensure 部分保障 | Not Ensure 不能保障 |
| 香港 | Matured Effort 成熟的努力 | Organized Effort 有组织的努力 | Informal Effort 非正式努力 | No Effort 无努力 |

我国的写实性评估结论使用“值得肯定,需要改进,必须整改”进行标识,与澳大利亚实际使用的院校评审报告的提法(赞扬推荐,证实肯定,劝告提议)比较相近,同

时吸收了英国院校评审报告中建议部分的分等的意涵。为了在借鉴的过程中不至于丢失国外相关设计的精髓,建议在引进时保留国外院校质量评审报告的相关结构要素,如表9所示。

表9 引进院校质量评审结论分等意见标识的适当转换

| 既定标识 | 澳洲标识方法 | 在既定分等结论标识下拟包含的写实性评估意见内容 |
|------|--------|-----------------------------|
| 值得肯定 | 赞扬推荐 | 值得赞扬的方面、堪称业内最佳的实践、可向同行推介的做法 |
| 需要改进 | 证实肯定 | 对自身不足之改进路径的认可,并就需强化方面提出补充意见 |
| 必须整改 | 劝告提议 | 指出学校需进一步改进的方面,提出进一步改进的意见和建议 |

此外,如果评估结论中能够参照国外的普遍做法,指出各项标准的达成情况甚至是对于内部质量保障机制的整体信心水平,并在报告中具体写出已做好的、能做好的、还需做更好的方面,就能更好地体现评估支持不断改进的顶层设计思路。如果在自评和他评报告中过多地强调改进和整改,且在现场反馈时又主要讲问题与不足,缺少对于院校现有做法的必要的“称赞与推荐”,特别是缺失对于院校已经发现的问题及其解决思路和方案的“肯定、证实和断言”,将不利于院校自身对内部质量保障机制的自我定位和自主完善。

为了制订具体的评判尺度,这里试图用综合评价卡的形式来写实性地报告各项评估标准的达标情况,如表10所示。综合评价卡分10个质量维度,对应于审核评估范围的24个相关联的审核要素,每个质量维度又分成5个达标程度。10个质量维度与24个审核要素的对应性是比较好的,除了按照质量管理的自身逻辑将原来放在教学资源条件保障度中的“专业设置与培养方案”放到了计划符合度中、将原来放在质量保障的组织完善度中的“质量监控”放到了管理成熟度中,绝大部分是完全对应的。综合评价卡的形式只是代表了一种系统诊断方式,它不属于开处方(Prescription,规定用药)。一方面,“开处方”这个词从境外专业评估实践的角度来看基本上是个贬义词,另一方面,鼓励学校在较小压力下自主选择才是专业评估服务之道,过于强势的整改要求只会造成学校对失败的恐惧,而并不能激发学校对学习的渴望。

表10 审核范围与质量评审细则的对应关系^[7]

| 质量的维度 | 相对应的审核要素 | 达标度一 | 达标度二 | 达标度三 | 达标度四 | 达标度五 |
|-------|---------------------------------------|------|------|------|------|------|
| 发展适应度 | 办学定位、培养目标 | 完全适应 | 比较适应 | 初步适应 | 不很适应 | 很不适应 |
| 计划符合度 | 人才培养中心地位、专业设置与培养方案 | 非常符合 | 较好符合 | 基本符合 | 不够符合 | 很不符合 |
| 条件保障度 | 数量与结构、教育教学水平、教师教学投入、教师发展与服务、教学经费、教学设施 | 充分保障 | 较有保障 | 基本保障 | 保障不力 | 没有保障 |
| 安排有效度 | 课程资源、社会资源 | 切实可行 | 操作性强 | 具可行性 | 可行性弱 | 无可行性 |

续表 10

| 质量的维度 | 相对应的审核要素 | 达标度一 | 达标度二 | 达标度三 | 达标度四 | 达标度五 |
|-------|----------------------|------|------|------|------|------|
| 活动整合度 | 教学改革、课堂教学、实践教学、第二课堂 | 有机融合 | 有效整合 | 初步磨合 | 勉强凑合 | 机缘巧合 |
| 管理成熟度 | 招生及生源情况、学生指导与服务、质量监控 | 成熟期 | 智慧期 | 启蒙期 | 觉醒期 | 不成熟期 |
| 目标达成度 | 学风与学习效果、就业与发展 | 达成目标 | 基本达标 | 部分达标 | 尚未达标 | 远离目标 |
| 组织完善度 | 教学质量保障体系 | 信心充分 | 较有信心 | 尚有信心 | 信心有限 | 缺乏信心 |
| 用户满意度 | 质量信息及利用 | 十分满意 | 比较满意 | 还算满意 | 不太满意 | 很不满意 |
| 学业风险度 | 质量改进 | 未有风险 | 风险较小 | 有些风险 | 风险较大 | 风险很大 |

其实,教育部的相关文件已经规定了评估报告的内容,明确要求评估报告“应在全面深入考察和准确把握所有审核内容基础上,对各评审项目及其要素的审核情况进行描述,并围绕审核重点对学校本科人才培养总体情况作出判断和评价,同时明确学校教学工作值得肯定、需要改进和必须整改的方面。”切实贯彻好这一规定,不仅能够使我们的审核评估与国外的院校质量评审真正做到实质等效,而且对于新一轮评估的成功实施至关重要。

参考文献

- [1] 袁益民. 教育评估的体制创新[M]. 南京:凤凰出版集团江苏科技出版社,2007:116-117.
- [2] Hilary, W. & Rob, C., Regulation and Quality Assurance in Higher Education[R]. 2011-10-24.
- [3] William F. Massy. Education Quality Audit as Applied in Hong Kong[EB/OL]. <http://www.unc.edu/ppaq/docs/Audit/Audit.html>, 2014-07-11.
- [4] John M. & Archie M. Self-evaluation[Z]. Routledge Falmer, 2002.
- [5] 袁益民. 教育评估改革:以学校为主体,以质量为核心,以服务为导向[J]. 常州大学学报:社会科学版, 2014(1):1-4.
- [6] 教育部高等教育教学评估中心. 中国高等教育评估词汇[M]. 北京:高等教育出版社, 2010: 66.
- [7] 袁益民. 审核评估:如何有效使用专业的评审方法及技术[J]. 高教发展与评估, 2014(1):16-24, 48.

(收稿日期:2014-11-11;编辑:伊夫)