

对下一轮普通高等学校本科教学评估若干问题的探讨

李志义 朱 泓 刘志军 张晓鹏

摘要: 基于高等教育评估理论, 结合我国高等教育以及高等教育评估制度的发展现状, 分析研究了下一轮评估的几个基本问题——下一轮评估叫什么(名称)、评什么(内容)和为什么(目的)等。正确认识这些问题, 对下一轮评估方案设计非常重要。

关键词: 高等教育, 教学评估; 人才培养; 质量保证

我国于2003年至2008年, 成功地开展了一轮普通高等学校本科教学工作水平评估, 极大地强化了高等学校教学工作, 有力地促进了教学质量的提高, 对建立中国特色的高等教育质量保证金体系积累了宝贵经验。这一轮评估成效显著, 意义重大。目前, 教育部正在组织力量, 研究制定下一轮评估方案。下一轮评估的几个基本问题——下一轮评估叫什么(名称)、评什么(内容)、为什么(目的)和怎样评(方法)等, 是下一轮评估方案设计的核心。本文主要针对上述前三个问题做一探讨, 最后一个问题——下一轮评估怎样评(方法)将在本文的续篇“对下一轮普通高等学校本科教学评估方法的探讨”中进行讨论。

一、下一轮评估叫什么?

上一轮评估叫“本科教学工作水平评估”, 下一轮评估叫什么? 有人建议叫“本科教学质量评估”, 有人建议叫“本科人才培养质量评估”, 还有人认为应叫“本科教学工作评估”、“本科人才培养工作评估”, 等等。这些名称的主要区别有二: 一是主题词是“教学”还是“人才培养”; 二是对这些主题词怎么加定语。无论是哪一个主题词, 前置“本科”定语是共识, 无可非议。但要不要后置定语, 后置“质量”抑或“工作”? 需要探讨。这实际上是评估“名称”与评估“定位”的关系问题。我们认为, 应该将“名称”与“定位”相分离。“名称”应该“宏观”, “定位”应该“具体”; “名称”应该有通用性, 要相对固定; “定位”应该有针对性, 要根据不同时期的重点进行调整。因此, 建议在关键词后面不再加定语。那么, 究竟是本科教学评估, 还是本科人才培养评估? 我们认为, 应该是“本科教学评估”。原因有二: 一

是“教学评估”已是一个国际通用名词, 采用“教学评估”便于国际交流与认可; 二是“教学评估”与上一轮评估有继承, 为人们所习惯, 采用“教学评估”有利于评估“品牌”建设, 形成评估文化。

就“教学评估”的国际通用性和国内习惯性, 我们做了文献调查。检索国外含有人文社科类的大型文献数据库的结果表明, “教学评估”的词条很通用, 但“人才培养评估”的词条寥寥无几。例如对 Journal of storage (JSTOR, 人文、社科、经管、数理化、环境生态过刊全文数据库) 和 Gale-Info Trac Custom (GITC, 人文、社科、科技期刊全文数据库) 分别以 Teaching Evaluation 和 Cultivation Evaluation (包括其他同义词及其不同组合) 为词条的检索结果如表 1。

表 1 国外文献对“教学评估”及“人才培养评估”术语使用情况

检索词组	检索结果(包含该检索词条的文献篇数)	
	JSTOR 数据库	GITC 数据库
Teaching Evaluation	395	673
Cultivation Evaluation	0	0

另一方面, 国外一些著作和官方文件中通常使用“教学评估”的术语, 几乎没有看到“人才培养评估”之类的表述。

对国内数据库(包括中国期刊全文数据库, 中国博士学位论文全文数据库, 中国优秀硕士学位论文全文数据库全文数

李志义, 大连理工大学副校长, 教授; 朱 泓, 大连理工大学教务处处长, 教授。

据库,中国重要会议论文全文数据库,中国重要报纸全文数据库,中国年鉴全文数据库)的联合检索结果(1979—2008年)见表2。

表2 国内文献对“教学评估”及“人才培养评估”及相关术语的使用情况

检索词条	检索结果(包含检索词条的文献篇数)	
	检索词条出现在题名中	检索词条出现在全文中
教学评估	1 840	43 435
人才培养评估	8	215
教学质量评估	859	17 140
培养质量评估	21	507
教学工作评估	143	3 749
培养工作评估	18	326
教学工作水平评估	516	7 805
教育评估	1 758	30 291
教育质量评估	215	6 409

从表2可以清楚地看出,“教学评估”的认可度极高,而“人才培养评估”的认可度极低,后者还不到前者的千分之五。“教学评估”的认可度甚至远高于大家熟知的“教学工作水平评估”(在文章题名中后者仅为前者的28%,在全文中后者仅为前者的20%)。值得注意的是,“教学评估”的使用频率甚至高于它的上位概念“教育评估”。

事实上,“高等教育”、“人才培养”和“教学”这三个概念在高等教育领域是有明确界定的。“高等教育”是“人才培养”的上位概念,“人才培养”是“教学”的上位概念。按现代教育理论与实践,高等教育具有三大功能,即人才培养、科学研究与社会服务。“人才培养”包括了“育人”(狭义上指德育)和“育才”(狭义上指智育)两方面,其中“育才”主要通过“教学”来实现。然而,由于不同的应用场合及人们的习惯,往往将“教学”这一概念提升到“人才培养”这个层面上使用,这可理解为是对“教学”概念的延伸,是对它的广义理解。实际上,上述的“教学评估”,其含义与“人才培养评估”基本相近。

二、下一轮评估评什么?

前已述及,下一轮评估叫“教学评估”更为妥当,但不一定叫什么就评什么,这就是名称与定位的关系问题。一般来说,定位应在名称的范畴之内。上一轮评估名叫“本科教学工作水平评估”,题目不大,但内容不少,几乎涉及了学校工作的方方面面,差不多成了“学校工作水平评估”,定位过于宽泛。上一轮评估之所以“扰民”成为一个焦点问题,除了个别被评估学校办学条件差、管理不到位、欠账太多,需要“临时抱佛脚”外,评估本身量大面广也不能不是一个重要原因。因此,下一轮评估一定要“定好位”,“缩小面”。

做好下一轮评估的定位,首先要考虑当前高等教育的核心问题是什么。现阶段高等教育的核心问题就是“质量”。胡锦涛总书记在党的十七大报告中明确指出,要“提高高等教育质量”。温家宝总理在国家科技教育领导小组研究制定《国家中长期教育改革和发展规划纲要》会议上,做了题为“百年大计、教育为本”的重要讲话^[1]。针对“高等教育问题”,温总理指出:“从长远看,我们不仅要不断扩大高等教育的规模,满足群众对高等教育的需求,更重要的是要提高高等教育质量,把提高高等教育质量摆在更加突出的位置。”特别强调要“建立和完善高等教育质量保障体系”。教育部周济部长在实施高等学校本科教学质量与教学改革工程视频会议上的讲话中指出:“按照党中央、国务院的决策部署,‘十一五’期间,高等教育需要切实把重点放在提高质量上。”“贯彻落实党中央、国务院切实把高等教育的重点放在提高质量上的决策部署,必须落在具体行动上。”因此,经国务院批准,教育部、财政部联合下发了2007年“1号文件”,决定实施“质量工程”。“质量工程”自2007年实施以来,已经取得了巨大成效。教育部高教司张尧学司长在下一轮本科教学评估课题研讨会(2008年10月14日,北京师范大学)上的讲话中指出:“高等教育就是要抓教育质量,而且这应该是各个学校办学的目标,评估要为提高高等教育质量服务”;“评估无论叫本科教学工作评估还是叫人才培养工作评估,其重点就是评质量”。显然,“质量”是目前高等教育的核心问题,评“质量”就抓住了这一核心问题。

做好下一轮评估的定位,还要考虑我国高等教育评估现状。上一轮评估为下一轮评估奠定了坚实基础。上一轮评估至少在以下三方面取得了显著成效:一是提高了认识,使教学工作的中心地位得到了进一步强化;二是加大了投入,使办学条件得到了极大改善;三是规范了管理,为建立中国特色的高等教育质量保证体系积累了宝贵经验。上一轮评估本着“以评促建、以评促改、

以评促管”，以及评估内容和指标体系的导向作用，将评估的重点引向了“投入”和“硬条件”。下一轮评估应该将重点投向“过程”、“产出”和“软环境”，要更注重发展内涵。有学者认为，下一轮评估应该着重“评软不评硬、评动不评静”。那么，“软”、“动”指的是什么？是“质量”。也就是说，经过上一轮评估后，下一轮评估应该定位在“质量”上。

那么，应该评教育质量？教学质量？还是人才培养质量？准确地讲，教育质量是培养质量的上位概念，它不仅包括了培养质量，还包括科研质量和服务质量等；培养质量是教学质量的上位概念，它包括教育和教学两方面的质量；教学质量是指教与学两方面的质量。教育质量太宽泛，教学质量太狭窄，下一轮评估应该评“人才培养质量”。

什么是人才培养质量？至今尚缺乏一个确切的定义，学术界也有不同看法。我们不妨对其做如下的引申。ISO9000 将质量定义为：“一组固有特性满足要求的程度”。对人才培养来说，这种“要求”的主体主要是指国家、社会群体（即用人单位）和个人（即培养对象）。满足国家和社会群体的“要求”，体现的是“社会本位”的教育价值观；满足个人的“要求”，体现的是“个体本位”的教育价值观。人才培养的“固有特性”，体现于人才培养目标，通过培养过程而形成，主要表现在知识、能力、素质等方面。据此，“培养质量”可理解为：培养目标满足国家、社会群体和培养对象的要求（需求和期望）的程度。

人才培养质量的可测性和可评性，尚是一个有争议的问题。人才培养质量是一个状态特性，它犹如产品质量，反映的是产品本身的性态。它来自于产品加工过程，但与产品加工过程并不是完全对应关系。也就是说，高质量的加工过程总会加工出高质量的产品，但质量不高的加工过程有时也会加工出高质量的产品。换句话说，保证了某一时期的培养质量，不一定能保证任何时期的人才培养质量；只有保证了人才培养过程的质量，才能保证人才培养质量的一贯性和持续性。另一方面，许多工业产品的质量是直接可测的，但人才培养质量的可测性较差。代表人才培养质量的许多特性具有潜在性（例如素质）和依时性（例如能力），是无法用外显的和状态指标来测量。因此，评估人才培养质量的科学性和操作性值得商榷。

人才培养质量保证体系具有过程特性。ISO9000 将“体系”定义为：“相互关联或相互作用的一组要素”。人才培养体系的指向是人才培养目标，作用于人才培养过程。据此，可将“人才培养体系”理解为^[2]：为实现培养目标所实施的培养过程中，相互关联或相互作用的要

素。也就是说，抓住了这些要素及其联系，保证了这些过程的质量，就能保证人才培养质量。因此，高校应该建立人才培养质量保证体系。人才培养质量保证体系是否完善和有效，决定了人才培养质量是否能达到要求。一个完善和有效的人才培养质量保证体系，不仅能保证人才培养质量，而且能保证人才培养质量的持续性，同时还能保证人才培养质量的不断改进。如果将下一轮评估定位于评估本科人才培养质量保证体系，不仅能间接地评估人才培养质量，而且能评估质量保证和质量改进的能力（即提高质量的能力），同时对建立中国特色的高等教育质量保证体系具有重要意义。

三、下一轮评估为什么？

如果我们将下一轮评估的目的确定为“保证和提高人才培养质量”，就会使评估目的与前述的评估内容完全契合。

按照教育评估理论，教育评估有形成性评估（formative evaluation）和总结性评估（summative evaluation）两种类型^[3]。形成性评估注重的是过程，目的指向是“改进”；总结性评估注重的是结果，目的指向“分等”。发展性评估是形成性评估的一种，它是以现代教育的科学发展观为指导思想，以“立足现在，回顾过去，着眼未来，促进发展”为基本理念，以满足人民群众对高质量的高等教育需求为出发点，以发展过程、进步幅度、发展潜力和发展水平为核心内容，注重个性和特色，以促进高校及师生自主的持续、快速、健康发展为根本目的的一种教学评估^[4]。

如果将下一轮评估的目的确定为“保证和提高教学质量”，就意味着下一轮评估选择了“发展性评估”。在下一轮评估方案设计中，一定要保持评估内容、评估方法以及评估结果与评估目的的契合性。也就是说，一定要使评估的目的性和方法论相契合。如果它们的契合性很差甚至不契合，在评估实践中就会出问题。上轮评估的目的是“以评促建、以评促改、以评促管”，但对评估结果设计了四个等级（优秀、良好、合格、不合格）。前者体现了发展性评估的目的，后者体现了总结性评估的结果，二者之间不契合。正是这种偏差，在评估实践中，被评学校在争取得到理想的评估结果上所花的精力远远大于评估目的“促建、促改、促管”的初衷。于是就出现了“优秀情结”。被评学校在地方教育行政主管部门、社会、学校教职员、学生及其他利益相关者的重压之下，争取得到“优秀”的评估结果几乎成为别无选择的目的。“以评促建、以评促改、以评促管”，在许多场合下“为评而建、为评而改、为评而管”。

上轮评估中出现一些大家反映比较强烈的问题,大多与评估方案设计时在评估的目的性与方法论的契合性出现的偏差有关。例如:

——弄虚作假。为什么会“弄虚作假”,因为个别学校达不到某些评估指标的要求,但还要得到一个更好(甚至“优秀”)的结论。为了追求评估结果,在评估准备阶段,按照评估指标体系进行“建”、“改”、“管”,“建”不好的、“改”不了的以及“管”不到位的就“造假”。

——形式主义。为什么在评估专家考察期间会出现“形式主义”,因为评估专家可根据自己看到的、听到的和感受到的对学校的本科教学工作水平做出判断,进而进行“分等”。为了使评估专家感受到学校对评估的“重视”,精心策划专家进校的“评估汇报会”,不惜成本准备“文艺演出”,周密安排各级政府“领导会见”;有的学校甚至对教室、食堂、宿舍、图书馆、校园等主要场所进行了周密布置,对相关人员的言行进行了精心安排;有的高校还对规定的几个座谈会进行了多次“彩排”,什么人讲什么话都有精心设计。专家进校犹如“兵临城下”,专家看到的、听到的和感受到的,大都是精心准备过的,失去了本来状态,以形式化取代了常态化。事实上,对于发展性评估来说,其主要功能就是“诊断”和“开处方”,如果失去了正常状态,“诊断”的结果和开出的“处方”就失去了意义。

——腐败现象。当然,这只是很个别的现象,而且很大程度上取决于个别“专家”的素质,但在评估设计上也有下一轮评估值得思考的问题,那就是评估结论由谁说了算。能不能获得“优秀”,在很大程度上取决于评估专家的一票。这就使得个别学校将重视评估结果,取向于重视评估专家;将重视评估专家,取向于一些不正当手段。

——分类指导分类评估。上轮评估标准统一、方法单一,固然需要改进,但分类指导和分类评估就一定能满足社会和学校对评估的期待?恐怕并不尽然。这里仍是评估的目的性和方法论的契合性问题^[5]。如果是总结性评估,通过评估要对学校的本科教学水平进行分等,则分类指导和分类评估就是前提。一般来说,分等越细,分类就得越细,否则就失去了可比性,评估结果就会失去合理性。然而,如果是发展性评估,评估的目的重在通过评估来发现问题、找出改进的方向、提高教学工作水平和培养质量,那么评估结果对学校类型和层次的依存度就会显著降低。一个最好的例证就是 ISO9000 质量

管理体系认证。由于该认证是判断申请认证的组织所建立的质量管理体系是否符合 ISO9000 标准要求,并不对组织的质量管理水平进行分等评价,尽管不同行业、不同性质、不同类型的各种组织(如工厂、银行、商场、医院、学校、甚至幼儿园、敬老院等)千差万别,但只有一个通用性标准 ISO90001,并没有考虑“分类指导”,其结果已被人们广泛认可。

事实上,用一把“尺子”去量不同的学校固然不好,但多几把“尺子”就未必能解决学校的个性化、多样化发展问题。最好的做法是把这把“尺子”交给学校,让学校用自己的“尺子”来量自己的“个头”。这正是审核评估模式(详见本文的续篇“对下一轮普通高等学校本科教学评估方法的探讨”)的特点之所在,这也是 ISO9000 认证的基本思想之所在。

因此,分类指导和分类评估的本质问题,是评估的目的性与方法论的契合问题。如果采用审核评估模式进行评估,就可以将外部的“分类”变成了学校自己的“定位”,将“分类”权交给了学校自己,让学校用自己的“尺子”来量自己的“个头”。

以上主要讨论了下一轮评估叫什么,评什么,和为什么等问题。概括起来,下一轮评估应叫“普通高等学校本科教学评估”,应该评“人才培养质量保证体系”,目的是“保证和提高人才培养质量”。正确认识这些问题,对下一轮评估方案设计非常重要。

[项目来源:教育部高等学校本科教学分类评估项目。参加本研究的还有丁堃、仲秋雁、李旦、李亚东、陈峥等]

参考文献:

- [1]温家宝.百年大计,教育为本[OL].中央政府门户网站 http://www.gov.cn/jdhd/2009-01/04/content_1194983.htm, 2009-01-04.
- [2]李志义.研究型大学如何构建本科人才培养新体系[J].中国高等教育, 2008(13/14): 34~37.
- [3]陈玉琨.教育评价学[M].北京:人民出版社,1999.
- [4]周光明.高校发展性教学评估研究[MA].中南大学硕士学位论文论文,2007.
- [5]孙建荣.对“分类指导分类评估”提法的思考:评估目的性与评估方法论[J].辽宁教育研究,2008(11): 33~36.

[责任编辑:杨裕南]