

综合评价卡：审核评估的新探索

袁益民

(江苏省教育评估院, 江苏 南京 210024)

摘要:根据国家评估改革方案,地方院校的审核评估下放至省级组织实施。某省在审核评估中试用了综合评价卡,作为写实性评估结论意见的一种标识和呈现方式。通过对院校内部质量保障状态的十个维度和各维度的五个达标程度的定性描述,并以坐标和曲线的形式直观表达院校所处的位置。这一创新实践既避免了以往教学水平评估的“优良中差”分等评估的弊端,也规避了新一轮教学工作审核评估结论出现过于大而化之、无法全面客观反映院校的整体实情的风险,受到专家的普遍肯定。未来拟帮助院校更加深入地理解和使用好综合评价卡,为其强化内部质量保障机制打造一个切实可行的专业工具。

关键词:审核评估;综合评价卡;内部质量保障

中图分类号: G64

文献标识码: A

文章编号: 2095-042X (2015) 06-0082-05

doi: 10.3969/j.issn.2095-042X.2015.06.014

根据国家本科教学工作评估新一轮改革方案,地方院校的审核评估(英文中称为 Institutional Quality Audit,意思是“院校质素评审”)将由省级教育行政部门负责,相关工作由具备条件的教育评估机构实施。某省在厚实的中外比较研究的基础上,基于国家整体方案架构下率先发展和创新发展的考虑,在若干方面进行了有益的探索,如:增加了材料复核及反馈的程序,现场考察中强化了师生问卷调查的作用,评后设定了改进工作进展报告和改进回访的制度等。特别是在两所院校的评估中试用了综合评价卡的综合鉴定与定性描述方法,不仅极大地方便了专家的内部分析、相互讨论和整体判断,同时使评估报告的写实性、概括性和实效性得到更好的体现,而且可以较好地规避新一轮评估可能出现的标准把握失之过宽、状态描述失之过偏、评估结论过虚过软的潜在和现实风险。

一、质量原理如何支持综合评价卡

就如质量是个十分模糊的概念一样,对质量管理这个很少得到实证研究的课题也存在各种不同的理解。维斯特福等(Westphal, Gulati, and Shortel, 1997)将顾客聚焦、持续改进、结构化的问题解决程序和雇员赋权增能作为关键质量维度;哈克曼和维琦曼(Hackman and Wageman, 1995)则将顾客聚焦、团队合作、供应者合作伙伴关系、

过程管理和统计及科学工具的使用作为核心维度;加文(Garvin, 1988)和特波尔(Teboul, 1991)识别了质量作为竞争优先、顾客聚焦、质量的执行、质量主动性、组织范围的承诺、最高管理层的领导力和进步主义的劳动力方面的实践等7个维度;波奖(The Malcolm Baldrige National Quality Award)则使用领导力,战略规划,顾客聚焦,测量、分析与知识管理,人力聚焦,过程管理和结果等7个维度;格林(Greene, 1993)更提出了24个全面质量管理维度^[1]。乌波路和欧本(Isaiah O. Ugboro, Kofi Obeng, 2000)将最高管理层的领导力、雇员赋权增能、工作满意度和顾客满意度作为组织全面质量管理最关键的要素。相关文献的作者还包括博文、希尔和希内德等(Bowen, Siehl, & Schneider, 1989; Brower, 1994; Deming, 1982; Roberts, 1994; Camp, 1989; Senge, 1994; Mendelowitz, 1991)^[2]。至于院校质量管理和质量保障的标准则更加多元,每个国家都会有不同的标准,甚至是许多学校都会有自己的标准。

但是,各种研究成果也为我们认识质量及其保障提供了重要的阶梯。无论是全面质量管理的PDCA循环,还是欧洲高校内部质量保障的标准与准则、波奖卓越绩效模式、澳大利亚持续质量改进环路OADRI模型、IPO投入过程产出模型、六西

* 收稿日期: 2015-10-20

作者简介: 袁益民(1963—),男,江苏江阴人,研究员,主要从事教育评估研究。

格玛 DMAIC 环路、BSC 平衡计分卡、逻辑框架方法（LFA）和 ISO9004-2 中的服务业通用质量环路等，都为我们探索先进的质量模型提供了借鉴。在

审核评估中根据自己国情确立符合自身的院校素质评审模型既可以借鉴国外质量模型，也可以与国内标准保持较好的匹配性（见表 1）。

表 1 综合评价卡与质量原理及审核项目的匹配情况

综合评价卡	审核项目	PDCA	OADR	IPO	DMAIC	Baldrige	BSC	ISO9004
发展适应度	定位与目标	计划	目标	投入	定义	领导力	使命愿景战略	服务需求
计划符合度	定位与目标	计划	路径	投入	定义	战略规划	使命愿景战略	开发纲领
条件保障度	队伍建设教学资源	实施	执行	投入	定义	人力聚焦	财务层	服务设计
安排有效度	队伍建设教学资源	实施	执行	投入	定义	过程管理	内部流程层面	服务设计
活动整合度	培养过程	实施	执行	过程	定义	过程管理	内部流程层面	服务提供
管理成熟度	学生发展	实施	执行	过程	控制	过程管理	内部流程层面	服务提供
目标达成度	学生发展	检查	结果	产出	测量	结果	内部流程层面	服务结果
组织完善度	质量保障	检查	结果	产出	分析	结果	内部流程层面	服务结果
用户满意度	质量保障	检查	结果	产出	分析	顾客聚焦	客户层	供客评价
学业风险度	质量保障	行动	改进	产出	改进	知识管理	学习层	分析改进

金·卡梅隆和韦斯利·塞恩（Kim Cameron, Wesley Sine, 1999）对组织性质量文化做过深入的研究，提出由低到高的四种质量文化：不重质量的文化、差错检测文化、差错预防文化、创造性质量文化^[1]。在后面两种较为先进的文化中，创造性质量文化重视顾客偏好的内化和顾客需要的预测，更加关注持续性、规律性、恒久性、突破性、超越性和引领性，它比差错预防文化更为可靠。综合评价卡的设计思路符合创造性质量文化的要求。

二、综合评价卡何以评价院校质量

综合评价卡作为审核评估中的一项评估工具，通过对院校内部质量保障状态的十个维度和各维度的五个达标程度的定性描述，试图刻画出院校在质量保障体系建设方面所处的位置，并以坐标和曲线的形式直观表达（见图 1）。综合评价卡采用绿、蓝、黄、橙、红五种颜色分别代表五个达标程度，反映不同的质量管理风险等级，但无数字表达的刻度。

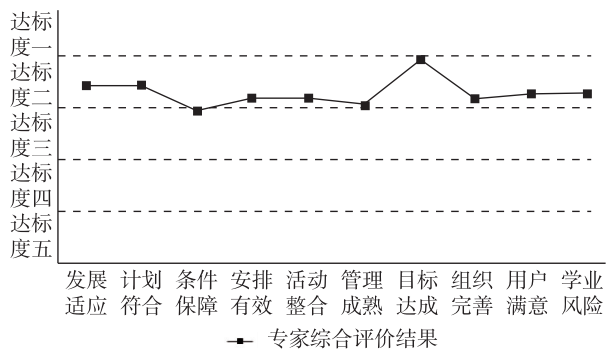


图 1 某高校 A 的综合评价卡

其中，十个质量维度是根据审核评估的审核要素并结合国外院校素质评审标准而转化过来的，而五个达标度是参照“零缺陷之父” Philip Crosby 的管理成熟度方格、能力成熟度 CMMI 模型和组织性

学习的成熟水平等相关理论并参照诸多国家的院校素质评审实践而设置的。综合评价卡与审核要素的对应关系及达成度定性描述见表 2。

综合评价卡的具体评价结论是根据专家的原始评价，通过将专家的定性描述等级转换成百分制而得出的。如图 1 中某高校 A 的综合评价卡曲线就是由 11 位专家根据质量的 10 个维度的达成度独立打分，然后去掉打分最高和最低的各一位，最后再将剩下的 9 位专家打分取平均数得出的。另一所高校 B 的专家组原始评价结果见表 3。该高校 B 的综合评价卡则尝试了另一种等级转换成曲线的方式：使用德菲尔专家法，通过专家组成员的多轮调查交流，在形成最大程度的共识的基础上建立的。从高校 B 的专家组原始评价结果可以看出，专家在“安排有效度”上有较大的分歧，在“用户满意度”上也有一定的分歧，因此采用后一种方法能够作出具有更大共识、也更为合理的判断。

三、审核评估为何需要综合评价卡

首先，审核评估中采用综合评价卡是为了更好落实该项评估的功能定位。作为我国首次系统引进的一项院校评估模式，审核评估的根本性的定位是对院校的内部质量保障机制及整个质量工作进行评审。审核重点是学校人才培养目标与培养效果的实现状况（实际上就是内部质量保障有效性这一核心）以及适应度、保障度、满意度等。综合评价卡便是根据质量保障本身的不同维度加上每个维度的若干达标程度编制而成的，它既反映了质量建设内在的规律性，又与审核评估的审核范围具有很高的匹配度，是帮助评估者全面认识和深度理解审核范围的一种解码器，也是内部质量保障机制及院校质量工作评价的一个很好的参照系。

表 2 综合评价卡与审核要素的对应关系及达成度定性描述 [4]

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
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	成熟时期	智慧时期	启蒙时期	觉醒时期	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

表 3 某高校 B 的专家组原始评价结果

质量的维度	专家 1	专家 2	专家 3	专家 4	专家 5	专家 6	专家 7	专家 8	专家 9	专家 10	专家 11
发展适应度	比较适应	完全适应	比较适应	比较适应	比较适应	比较适应	比较适应	比较适应	比较适应	比较适应	初步适应
计划符合度	较好符合	较好符合	较好符合	较好符合	非常符合	较好符合	较好符合	较好符合	较好符合	较好符合	基本符合
条件保障度	较有保障	基本保障	较有保障	较有保障	基本保障	基本保障	较有保障	较有保障	较有保障	基本保障	基本保障
安排有效度	操作性强	切实可行	切实可行	具可行性	操作性强	具可行性	切实可行	具可行性	操作性强	具可行性	具可行性
活动整合度	有效整合	有效整合	有效整合	有效整合	有机融合	初步磨合	有效整合	有效整合	初步磨合	有效整合	初步磨合
管理成熟度	智慧时期	智慧时期	智慧时期	智慧时期	智慧时期	启蒙时期	智慧时期	智慧时期	智慧时期	启蒙时期	启蒙时期
目标达成度	达成目标	基本达标	达成目标	基本达标	基本达标	基本达标	达成目标	达成目标	达成目标	基本达标	基本达标
组织完善度	尚有信心	较有信心	较有信心	较有信心	较有信心	较有信心	较有信心	较有信心	较有信心	较有信心	尚有信心
用户满意度	比较满意	比较满意	十分满意	比较满意	还算满意	比较满意	比较满意	比较满意	十分满意	比较满意	还算满意
学业风险度	风险较小	风险较小	风险较小	风险较小	风险较小	风险较小	风险较小	风险较小	风险较小	风险较小	有些风险

根据教育部的实施方案，审核评估的结论不搞通过不通过或者评出优良中差，而是将在评估结束后提供写实性报告。那么，这个写实性报告怎么去写实呢？

我认为写实性评估报告要真正写实，不只是一要通过实证手法呈现评估发现的事实，而且要结合实际的成绩和问题作出实质性的评估结论并就评后改进作出建设性安排。具体来说，一是要依据实际证据。写实性报告首先要严格按照所获取的证据来写，“有七分证据不说八分话”。二是认可改进实绩。对于学校按照上一轮评估中专家针对问题提出的改进建议所取得进展的实际情形，要作出是否已改进或已改进到何种程度的判断和描述。三是要具具体判断实情。根据评估实情作出具体的结论，不仅要指出已经做得很好因而值得赞扬的方面和做得还不够因而需要改进的方面，同时还要特别强调针对学校在自评中发现的问题以及已经在自评报告中提出的解决路径，提出同行认可性的意见以及建议继续强化的意见。四是突出实际问题。围绕参评院校的突出问题或者学校自己选定的特定课题，作出专

题性的深度剖析，提出相应的建设性意见。五是给出实质结论。使用内外部基准比对、三角定鉴及痕迹追踪等方法，发现并指出实际存在的学业及学术风险，作为作出总体性判断的依据，以明确指出学校在内部质量保障以及绩效表现方面实际所处的位置或阶段。六是关注后续实效。针对评估后如何处理评估发现和评估结论，提出适当处置意见，使评估工作真正服务于评后的改进。

在这些写实性的报告环节中，作出总体性的判断并给出实质性的结论是至关重要的。教高[2013] 10 号文明确要求，审核评估首先要“在全面深入考察和准确把握所有审核内容基础上，对各审核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状况作出判断和评价”。这实际上是对评估内容的全面性、评估发现的深入性、评估聚焦的选择性、事实描述的总体性和价值判断的终结性提出了实质性的要求。那么在评估报告中又怎么才能体现这些要求呢？我认为，综合评价卡便是可以较平衡地体现这些要求的评估结论的一种直观呈现方式。它既能涵盖证据收集的

所有范围，给数据的深入挖掘提供指引，又为确立评估重点理清了思路，为作出整体性结论搭建了框架。可以说，离开了这种综合性的评价工具，审核范围难免显得面面俱到，数据整合难免出现支离破碎，讨论分析难免偏离问题焦点，而评估结论难免陷入观点失衡。综合评价卡贵在综合，它顾及从投入到教学过程和学习产出的质量生成的所有基本环节，关照从规划、实施到检测、行动的质量管理的整个过程，贯穿从目标、路径到执行、结果和改进的质量评价主线，为专家作出更加全面深入的思考、评估机构作出更加平衡准确的决定、参评院校进行更有针对性的自我剖析和后续改进，提供了一项有效工具。

其次，审核评估中采用综合评价卡是为了能更好地帮助专家开展评估。审核评估采用专家进校前分头准备、在校时“全面考察、各有侧重、深入交流、独立判断”的方式介入评估工作。尽管有统一的审核范围及基本的评估程序，但是专家的评估活动大多是单独分散进行的。因此，集中的讨论和同一的评审尺度就显得尤为重要，而综合评价卡就为专家这样的内部分析、相互讨论和整体判断提供了一个基本的尺度。它既可以方便专家们在同一个平台上分析证据、交流各自的观点，还为他们寻求共识、作出判断提供了一个便捷的工具。如果使用得当，它可以帮助专家去伪存真、克服片面性，高效地形成专家个人及专家组集体的总体评价意见。由于专家组是在现场考察后半段即将形成评估结论前使用这一工具，所以使用过程并无任何障碍，网上填写综合评价卡更是只需要短短的半分钟之内的简单点击就可以轻松完成。综合评价卡的使用不仅不会增加专家的工作负担，而且是使他们的工作更加条理化、简便化的一个辅助性手段。

第三，审核评估中采用综合评价卡是为了能够规避该项评估的主要风险。审核评估的顺利实施仍然面临诸多困难和挑战。中国本身的院校评估的专业化之路还十分漫长。一方面在教育的管办评分离未有落实的条件下，评估机构远非是独立专业的权威部门，其独立性、专业性和权威性明显不足，评估运作过程受到种种制约和干扰，另一方面评估者的专业素质和业务水平无法满足发展需要，对评估工作原则、目标功能定位、评估程序以及方法与技术的认识明显不足，甚至对于作为后启外发的评估这一专业行当的许多基本概念还没有很好地消化吸收。同时，由于涉及审核评估的本体性研究严重不

足，学界及组织实施部门对于该评估模式的基本概念、关注焦点、相关程序、核心价值和技术路径等的认识容易出现偏差，既有该评估模式的核心要件在设计上出现缺胳膊少腿的现象，又有具体实施过程中遭遇穿新鞋走老路问题的风险。有的院校甚至已经出现了完全背离审核评估初衷、督促院系购买档案柜集中整理“完善”档案的情况。

审核评估作为我国首次引进的一种评估模式和已经实施的一个独立评估项目，在本土化改造后能否整合一致地服务于院校的内部质量保障与持续改进，目前来看仍然存在适切性、可行性和有效性方面的管理风险，包括评估标准与院校质量目标匹配性不够的风险、评估方法与评估目的匹配性不够的风险以及评估证据与评估结论匹配性不够的风险等。审核评估实施过程中最重要的风险和挑战是，由于矫枉过正，容易从上一轮本科教学评估的极其划一性和高厉害性走上本轮评估随意性和无利害的另一个极端，造成参评院校简单应付和轻松过关的思想泛滥，使作为质量主体的学校师生的参与大打折扣，使完善内部质量保障机制的评估目标最终落空。从目前的审核评估实际运作机制、专家培训内容、具体评估环节、院校自评过程和审核评估报告来看，这些困难和挑战是现实的风险。

就拿审核评估报告来说，目前过于淡化整体性结论、排斥等级性描述，所谓写实性报告不作通过或不通过的判断、不给等第。似乎只有这样才能克服上一轮评估的弊端，而事实是除了利害度的降低和随意性的提高，传统评估的许多做法仍然是可以一切照旧的。我们一直习惯于在肯定主流的同时指出问题，而指出的问题往往在很长的历史阶段仍然存在，有些还变得更为严重。如果我们本轮评估最终采用没有多少约束性的轻描淡写和不痛不痒的文字进行所谓写实，那么最后的结果就会与上轮评估高压低效下的虎头蛇尾和最终皆大欢喜的高优秀率没有什么实质的两样，只不过是一些人的弄虚作假变成了更多人的只虚不假。看似有利于全面总结和诚实反思，但却实际缺少了问题聚焦和评估实效。不少参评院校将审核评估看成可以不讲标准、不分高低、不看结论的“简略版”甚至“休闲版”评估，在自评中师生的参与少之又少，在评估后实质性的改进更加难以推动。恐怕有的学校连审核评估究竟是评什么都没有搞清就已经皆大欢喜地完成评估任务了；而有的学校连PDCA都没有翻译对就在一片赞扬声中接受评估的圆满成功了；有的学校

在用纯粹的传统评估手段应对完这么轻松的评估后就已经迫不及待地开始到处去介绍这种“新式”评估的参评经验了；有的学校只是匆匆地推销完自己的从未实际运作过的质量保障体系就已经率先通过这类评估了；而有的学校从别的学校那里随意克隆一个质量环路也就顺理成章地过关了；更不用说有的学校依靠传统的购置档案柜和整理材料的方式将之对付过去便无疑显得十分安全和轻松了。

其实，境外的审核评估报告一般均要求给出整体性结论，对于是否通过评估或在何种程度上通过评估普遍要求作出明确判断，不仅评估过程中要根据所获证据作出相应的判断，而且也要在审核评估报告中以差别化的等第描述来反映评估结论。芬兰甚至把评估结论设为通过或不通过，2005 开始的首轮评审中就有较高比例的院校没有通过而得到了需要“重新评审”的结论；有个国家不仅给出不同等级的结论，而且确定不同的评估周期。所以，评估报告不能一味地强调低利害，只要能够少一些人为的压力、多一些专业的建议，学校就会对评估结果多一份理解和支持。其实，学校对那些没有实质内容的评估报告才是最反感的。

审核评估中采用综合评价卡可以帮助我们较好地规避掉该评估项目的上述这些主要风险，无论是对于弥补我们教育评估专业化水平的不足和对审核评估认识的欠缺，还是推动该项评估顺利启动和平

稳开展，都是有益的。特别是综合评价卡能够在通过定性评价手段保持较低利害度的前提下，通过写实性呈现方式坚守了评估的有用性。每个参评学校拿到这样的一张评价卡，不用担心与别人比高低的压力，而是可以看到与众不同的、个性化的、独一无二的、真实的自己。审核评估中采用综合评价卡可以帮助院校更好地认识自身的优势与不足，特别是发现内部质量保障的实际现状，看出目前的薄弱环节和未来的努力方向，为院校评后的持续改进提供支持。作为一项新近提出的还需要不断完善的评估工具，综合评价卡还需要进一步细化评审准则，为专家及参评学校更好地使用编写出一个简要的指导与培训材料。同时，还可以进一步开发和完善相关的配套工具，以便更好地对院校质量建设的相关趋势和因果关系等进行分析、描述和呈现。

参考文献：

- [1] Cameron Kim, Wesley Sine. A framework for organizational quality culture [J]. Quality Management Journal, 1999, 6 (4): 7-25.
- [2] Isaiah O Ugboro, Kofi Obeng. Top management leadership, employee empowerment, job satisfaction, and customer satisfaction in TQM organizations: an empirical study [J]. Journal of Quality Management, 2000 (5): 247-272.
- [3] 袁益民. 审核评估：如何真正做到正确认识，确保平稳开展 [J]. 高教发展与评估, 2015 (1): 1-16.

Comprehensive Progress Report Card as a New Tool for Institutional Quality Audit in Jiangsu

Yuan Yimin

(Jiangsu Agency for Educational Evaluation, Nanjing 210024, China)

Abstract: Jiangsu Agency for Educational Evaluation adopts a comprehensive progress report card as an assessment tool in its recent institutional quality audit of two institutions of the Province. The report card consists of 10 dimensions of internal quality assurance plus 5 levels of attainment for each dimension. The status of the internal quality assurance of an audited institution is interpreted by a certain attainment level of each dimension, and it can also be expressed by a coordinates curve. In the two cases of institutional quality audit, the comprehensive progress report card has served as a useful tool for helping the auditors to make their judgments about the status of the internal quality assurance system of an audited institution. It can also be used as a framework for guiding institutions in data mining and conducting follow-up activities of continuous improvement.

Key words: institutional quality audit; comprehensive progress report card; internal quality assurance

(责任编辑：沈秀)