

中国地方政府绩效评估工具的现状与未来^{*}

——基于德尔菲法的研究

The Status and Future of Tools Applied in Local Government
Performance Evaluation : Research Based on the Delphi Method

王琦玮¹ 马亮² 吴建南¹

(1. 西安交通大学公共政策与管理学院; 2. 西安交通大学管理学院 西安 710049)

摘要 本文运用德尔菲法研究了中国地方政府绩效评估工具使用的现状并预测了未来发展趋势。研究表明,目标责任制、文明机关评选和公众参与的政府绩效评估(包括行风评议和公民评议机关)在现在和未来都将处于非常重要的地位,并且其未来的使用还会略有加强。而在学术界中备受关注的政府绩效评估工具(如平衡计分卡、绩效棱柱、通用评估框架等)在目前的应用程度和未来的应用前景都十分有限。

关键词 政府绩效评估 地方政府 评估工具 德尔菲法

中图分类号 D035

文献标识码 A

文章编号 1002 - 1965(2009)10 - 0015 - 04

在“新公共管理”运动的影响下,各国政府都强调结果、绩效与问责。政府绩效评估成为公共管理领域的热点问题。改革开放以来,绩效评估也逐步在我国各级政府及其组成部门中广泛开展,并受到了社会各界的普遍重视。评估工具是对政府绩效评估的构成要素进行整合的结果,不同的评估工具会使绩效评估产生不同的效果,因此评估工具的选择和搭配对于政府绩效评估而言意义重大。20世纪80年代以来,我国各地先后涌现出目标责任制考核、行风评议、万人评议机关、满意不满意单位评选等形式多种多样、内容各具特色的政府绩效评估活动^[1]。目前地方政府绩效评估工具的使用现状如何,未来的发展趋势怎样,尽管这些问题已经得到了部分关注,但还有待深入研究。

目前政府绩效评估工具的研究主要集中在对政府绩效评估活动的整体设计中,表现为评估的活动、模型、框架、方法等方面的集成系统。虽然评估工具多种多样,概括起来可以包括两大方面:一是已经得到较为广泛使用的评估工具,包括目标责任制考核^[2]、文明机关评选^[3]、行风评议^[4]和公民评议政府^[5]等,学者们对这些评估工具的问题进行了分析并提出了改进建议;二是通过国外和企业的先进经验学习而引入的评估工

具,如平衡计分卡、360度评估、绩效棱柱、通用评估框架^[6]等。还有学者^[7]对国外政府绩效评估方法进行了总结,对标杆管理等方法进行了介绍。然而,目前尚未见有系统梳理政府绩效评估工具使用的研究。本文以中国市级政府作为研究对象,探讨我国地方政府绩效评估工具的现实状况和未来趋势,以期对政府绩效评估研究和实践做出贡献。

1 研究设计

1.1 政府绩效评估工具的类型 我们的研究主要聚焦中国市级政府,包括直辖市和地级市。市级政府在中国政府体系中居于承上启下的关键环节,其所开展的绩效评估活动也更加值得考察。在市级政府中,各种绩效评估工具的使用呈现出百花齐放的状态。通过以往研究的梳理和实践的调研,我们总结了共计10种主要的评估工具,包括:目标责任制、文明机关评选、行风评议、公民评议机关、作业成本法、360度评估、ISO质量认证、平衡计分卡、绩效棱柱、欧盟通用评估框架,以下简称之。

目标责任制^[2]形成于20世纪80年代末,是各级政府普遍采用的一种评估模式。其主要特征在于上级

收稿日期:2009-07-24

修回日期:2009-09-14

基金项目:国家自然科学基金项目“目标责任考核、财政管理与组织创新——面向中国地方政府绩效改进的实证研究”(编号:70873092);教育部“新世纪优秀人才支持计划”(编号:NCET-06-0851)。

作者简介:王琦玮(1986-),女,硕士研究生,研究方向为政府绩效评价与管理;马亮(1984-),男,博士研究生,研究方向为政府绩效管理、政府创新。

政府每年年初通过与下级政府签订目标责任书,详列下级政府需要完成的一系列目标任务,并在年末进行考核奖惩。

文明机关评选最初是为了贯彻落实精神文明建设而开展的评选精神文明建设先进单位活动,主要由党委宣传部门下设的文明办负责,评估内容涉及政府部门的服务态度等“软性因素”,也是政府绩效评估的形式之一。

行风评议最早出现于 20 世纪 90 年代中期,一般由中共纪委和监察部门共同设立的纠正行业不正之风办公室负责实施,通过面向社会公众的问卷调查,对政府部门和公共服务行业的服务意识和工作作风进行评估。

公民评议机关是政府“开门评议”,邀请公众对政府部门的办事效率和服务态度进行评议,并根据结果要求被评部门予以整改。其形式与行风评议相似,但主管部门不同,一般是各级党委组织。

作业成本法是一种以作业为中心,通过对组织的所有作业活动进行追踪并予以动态反映,计量作业成本并评估作业绩效的成本计算和管理方法。它首先产生于制造业企业,后扩展到其他行业,并在西方国家一些政府中得到应用。

360 度评估最初源于人力资源管理领域对员工绩效评估的多角度设想,后逐步扩展到其他领域。在政府绩效评估中,多角度的评估意味着来自政府的各个关键利益相关者都可以参与到评估中,从而调和矛盾并平衡利益。

ISO 质量认证是由国际标准组织制定的质量管理体系,通过权威认证对组织质量管理流程进行标准化改造,实现组织质量管理的升级,并对组织绩效评估产生重要影响。

平衡计分卡是将组织战略目标逐层分解转化为各种具体的相互平衡的绩效评估指标,并对这些指标的实现状况进行不同时段考核,从而为战略目标的完成建立可靠的、有执行基础的绩效管理工具。

绩效棱柱则是一个三维框架模型,用棱柱的五个面分别代表公司业绩评估的五个方面:即利益相关者的满意、利益相关者的贡献、战略、流程和能力,其将利益相关者放到最重要的核心位置。

欧盟通用绩效评估框架与平衡计分卡类似,主要采用组织内部的自我评估,为组织需要共享信息并改善沟通,更清晰地了解组织的强势和劣势。

1.2 数据与方法 我们采用改造后的德尔菲法进行研究,了解评估工具的使用情况。具体调查问题为:“在贵城市政府绩效评估活动中,所使用工具的状况如

何?”调查由有关专家对其所在城市的政府绩效评估活动中上述 10 种评估工具的情况进行回答。两轮调查的有效问卷数分别为 34 份和 26 份,可以用于进一步分析。

本文使用第一轮数据中“目前的使用率”和“目前的优先权”两个方面的数据对评估工具的使用现状进行描述,利用第二轮数据中的“未来的希求度”和“未来的可行性”两个方面对未来趋势进行预测,并且增加了专家对“未来的可行性”的确切性判断。数据分析方法包括:a.对现状的分析运用均值和单因素方差分析中的 Sheffe - test 方法,检验多组数据均值之间的差异性并依据均值大小进行分组排序,通过方差分析了解各个工具在得分上的差异性;b.对未来的预测运用离散系数评估专家意见的一致程度,通过配对样本 t 检验分析现实描述和未来预测之间的整体差异。

2 目前政府绩效评估工具的使用率和优先权

我们首先对市级政府绩效评估工具目前的使用率和优先权进行研究,考察哪些评估工具在目前的使用程度较高、使用中受重视的程度较高(表 1)。就目前的使用率而言,使用最多的是文明机关评选、目标责任制和行风评议,接近或高于 4.00;使用最少的是通用评估框架、绩效棱柱、平衡计分卡等,在 2.00 左右。目前的优先权方面,受重视程度最高的仍然是文明机关评选、目标责任制和行风评议,受重视程度最低的依然是通用评估框架、绩效棱柱、平衡计分卡等。上述结果说明使用越多的评估工具在政府绩效评估中的重要性也较高,反之亦然。

表 1 专家对于政府绩效评估工具的现状与未来的打分均值

排序	绩效评估工具	目前的使用率	标准差	目前的优先权	标准差
1	360 度评估	2.68	1.28	2.65	1.23
2	ISO 质量认证	2.56	1.32	2.65	1.38
3	公民评议机关	3.30	1.26	3.34	1.29
4	作业成本法	2.76	1.00	2.97	1.03
5	绩效棱柱	2.19	1.08	2.53	1.14
6	目标责任制	4.09	0.93	4.13	1.06
7	平衡计分卡	2.31	1.31	2.33	1.27
8	通用评估框架	1.97	0.96	2.18	1.06
9	文明机关评选	4.24	1.00	4.06	1.06
10	行风评议	3.85	1.25	4.00	1.11

2.1 目前的使用率 为了进一步对目前的使用率进行分析,我们进行了单因素方差分析(见表 2)。可以发现,调查专家对政府绩效评估工具使用率的打分存在显著差异,据此我们可以根据目前的使用率将评估工具分为四组,各组之间显著差异,各组内部差异不显著。目标责任制和文明机关评选的得分最高,其次是行风评议、公民评议机关,再次是作业成本法、360 度评估、ISO 质量认证,最后是通用评估框架、绩效棱

柱和平衡计分卡。得分较低的评估工具多由国外研究者提出而在企业中应用广泛,在我国政府部门的使用还需要进一步的修正或完善。

表 2 政府绩效评估工具目前的使用率方差分析输出表

分组	维度	N	Subset for alpha = .05			
			1	2	3	4
A	通用评估框架	31	1.9355			
B	绩效棱柱	32	2.2188	2.2188		
B	平衡计分卡	33	2.3030	2.3030		
B	ISO 质量认证	33	2.5455	2.5455		
B	360 度评估	33	2.6364	2.6364		
C	作业成本法	34	2.7647	2.7647	2.7647	
D	公民评议机关	34		3.2647	3.2647	3.2647
E	行风评议	34			3.8235	3.8235
F	目标责任制	33				4.0303
F	文明机关评选	34				4.2059
Sig.			0.477	0.139	0.127	0.27

在受调查的城市中,特别是沿海的城市,文明机关评选声势浩大。文明机关创评对各单位改进机关作风、提高干部队伍素质产生了积极的效果^[3]。其评估指标的定性程度较高,评估流程一般为部门对照评选标准进行申报,并不是所有部门都参与,因此文明机关评选更像一种荣誉申报而非对于绩效的评估。

目标责任制是由上级政府制定目标,由下级政府执行完成,并由上级政府进行考核。与其他评估工具相比,它具有很强的控制力^[2]。首先,目标的制定比较清晰明确,类似这样的指标是易于考核的。通过目标考核,可以使政府部门不断下放和分解自身的目标,从而分担了风险。其次,在结果使用方面目标责任制的力度强于其他工具,特别是针对于某些特定考核项目会采用一票否决制,使考核对象在目标履行过程中尽量达成目标以避免惩罚。

2.2 目前的优先权 在目前的优先权方面,德尔菲专家给出的打分结果与使用率的排序上具有高度相似的特征,说明目前使用比较多的评估工具即为优先权比较高的评估工具。

表 3 政府绩效评估工具目前的优先权方差分析输出表

分组	维度	N	Subset for alpha = .05	
			1	2
A	通用评估框架	29	2.1379	
B	平衡计分卡	31	2.2903	
B	绩效棱柱	31	2.5161	
B	ISO 质量认证	32	2.5938	
B	360 度评估	32	2.6250	
C	作业成本法	33	2.9697	2.9697
C	公民评议机关	33	3.3030	3.3030
D	行风评议	33		3.9697
D	文明机关评选	33		4.0303
D	目标责任制	32		4.0625
Sig.			.076	0.131

评估工具的使用与评估目的密切相关。目标责任制作为确保上级了解下级政府和相关领导工作执行情

况的重要绩效评估工具,其排序靠前也是在情理之中。在目标制定和考核的过程中,上级政府和领导都起到了关键的作用。相对其他的绩效评估工具,正是目标责任制强的控制力和“自上而下”推行使其在绩效评估的工具中使用具有最高的优先权。

在政府部门中通过绩效评估了解外部公众的需求和满意度则更加重要。推动管制型行政向服务型行政转变的时期,绩效评估中的公民导向尤为重要^[8],行风评议的评估主体并非上级而是接受政府服务的公众,因此在公民导向的背景下,行风评议和公民评议机关受到了政府部门和学者们的关注。这两种工具中,行风评议目前的优先权高于公民评议机关。

调查专家认为,在现行的绩效评估工具中,平衡计分卡、绩效棱柱和通用评估框架等评估工具的优先权较低。虽然它们在学术研究中较广泛,在实践中也有了一些探索,但并没有大规模展开。原因主要在于这几种评估工具在中国地方政府的应用才刚刚开始,尚不够成熟。

3 未来政府绩效评估工具使用的希求度和可行性

对于市级政府绩效评估工具在未来的希求度和可行性,我们也进行了分析,以考察调查专家认为哪些评估工具在未来需要予以重视,哪些评估工具在未来使用更加可行,以及他们对其回答的肯定程度(见表 4)。

表 4 专家对政府绩效评估工具的未来预测结果

排序	绩效评估工具	未来的希求度		未来的可行性		未来可行性的确定性	
		众数	变异系数	众数	变异系数	众数	变异系数
1	360 度评估	3	0.30	4	0.37	3	0.30
2	ISO 质量认证	3	0.35	3	0.35	3	0.28
3	公民评议机关	5	0.18	5	0.22	3	0.27
4	作业成本法	3	0.26	3	0.26	3	0.26
5	绩效棱柱	3	0.27	3	0.27	2	0.29
6	目标责任制	5	0.17	5	0.25	4	0.24
7	平衡计分卡	3	0.39	4	0.32	2	0.30
8	通用评估框架	3	0.44	3	0.41	2	0.40
9	文明机关评选	5	0.35	3	0.26	3	0.27
10	行风评议	5	0.30	4	0.25	4	0.29

3.1 未来的希求度 可以看到大部分绩效评估工具的变异系数仍然大于 0.25,即专家对于绩效评估工具在未来三年的希求没有达成一致。不同于目标责任制是自上而下推行的绩效评估工具,具有较高的可复制性,而且在各个地区的希求度比较一致,其他的绩效评估活动在不同地区的需求是不同的。不同地区的绩效评估工具使用存在着很高的差异,由于不同的专家来自不同的地区,所以对于评估工具在未来的希求度的使用很难达成一致。

另外还有一些因素会影响专家对绩效评估工具的认识。由于地区的差异,不同地区的绩效评估工具在运用中也可能存在一些差异,即使名称一致,在实际使

用的过程中可能也存在不同的内涵,所以在不同地方的需求也就不一样。特别是针对于由国外引入的绩效评估工具,由于其在应用的时候必须根据国情和地方的具体情况进行调整。即使是在我国根植的绩效评估工具,在不同市级政府的使用情况也存在差异。这是专家打分过程中离散程度较高的原因。

尽管专家对目前绩效评估工具的使用率和优先权存在着不同的看法,但是在对于未来希求度的分析中,可以看到公民评议机关和目标责任制不但平均分最高,同时也具有较低的变异系数,这表明专家一致认为,这两种绩效评估工具在未来中国市级政府中仍会得到重视。专家认为,在未来市级政府仍然会将目标责任制作为评估的重要工具,通过这种工具督促政府完成既定的各项目标,实现上级政府的控制能力;公民评议机关在未来评估工具使用中,将会与目标责任制一起,成为最为重要的绩效评估工具,公众对于政府的绩效评估将会纳入到整个绩效评估体系中,从而充分地了解公众对于政府的需求和意见,进一步提高组织的服务意识和水平。

通过第二轮调查,数据仍然没有得到有效收敛。然而这一结论基本符合不同地区市级政府的希求。同时使得我们有了进一步的反思,即绩效评估工具在未来的使用会呈现出什么样的景象。较高的变异系数表明未来绩效评估工具在不同地方的需求是不同的,这样的现象可能存在着多样的原因,评估工具未来的希求度会受到政府绩效评估体系自身的影响,在目标责任制等传统评估工具使用比较成熟的地区,市级政府会倾向于创新并对绩效评估提出更高的要求,包括提高评估体系整体的信度和效度、完善指标等。

3.2 未来的可行性 在未来的可行性方面,除了欧盟通用框架以外,所有评估工具的可行性得分均大于 3 分,其中目标责任制、公民评议机关和行风评议的可行性最高,并且变异系数均小于 0.25,说明专家较为一致地认为这三项绩效评估活动具有最高可行性。

在对未来的可行性分析上,专家对于未来评估工具使用可行性的确定程度最高的为目标责任制。目标责任制是现行条件下与我国政治体制结合最为紧密的评估工具,因此无论是现在还是未来,专家均认为目标责任制始终是最为重要的工具。尽管所有评估工具的众数都已经在 3 分以上,但是对于欧盟评估框架、平衡计分卡和绩效棱柱等在未来的可行性的确定程度上仍然偏低。这三种评估工具均是从国外引进的,如何使其具有“中国特色”还需要进一步探索。

在对未来的希求度和可行性的分析上,可以看到通用评估框架都排在最后一位,即通用评估框架在未

来中国的应用前景并不被专家看好。尽管通用框架可能对我国有一定的借鉴意义,但是它对于绩效评估框架和内容的统一、评估体系的配套建设^[6]都具有较高的要求。因此专家们对于其可行性仍持谨慎态度。

4 政府绩效评估工具使用:现状和未来的比较

上文分别对政府绩效评估工具的现状与未来进行了考察,以下将对现状与未来进行比较,以期发现更有价值的结果。我们首先对评估工具目前的使用率与未来的希求度之间进行比较分析,然后对未来的希求度与可行性之间进行比较。

通过对于市级政府绩效评估工具在目前的使用率和未来的希求度进行 t 检验后发现,文明机关评选、目标责任制在现在的使用率和未来的希求度之间并不存在显著差异。

文明机关评选和行风评议的得分在未来的希求度上显著低于目前的使用率,表明专家对这两种评估工具的未来前景并不看好。虽然这两种评估工具目前的使用程度较高,但由于体制约束,政府自行组织的公民评议仍然存在许多缺陷难以克服。因此,专家们对这两种评估工具在未来的发展并不鼓励,尽管它们的重要性因为体制惯性并不会降低。

除上述以外的其他评估工具未来的希求度都显著高于现在的使用率,意味着它们都存在充足的提升空间,并且会受到更多的重视。这表明专家们认为这些评估工具可能会在未来更加受到重视,同时也表明现有评估工具的使用不足。显然,对于这些在国外和企业中应用普遍的评估工具,专家们仍然寄予厚望,希望它们能够在未来得到更多的利用。

5 结 论

通过两轮调查数据的分析,我们对中国市级政府绩效评估工具使用的现状和未来的研究发现,评估工具的使用在现在和未来表现出如下特征:

目标责任制和公众参与的政府绩效评估是目前最重要的评估工具。目标责任制是目前使用频率最高、优先权非常高的评估工具。公众参与的政府绩效评估也受到广泛关注,公民评议机关可能会在未来取代或整合行风评议,成为新的公众参与的评估工具。目标责任制主要使用客观的结果指标,一般由统计部门提供;公众参与的政府绩效评估则需要外部公众的参与并通过主观感知评估获得信息,因此二者一内一外、一主观一客观,未来或可构成“内外兼具”的政府绩效评

(下转第 23 页)

用率和优先权均不高。长此以往,就偏离了政府绩效评估活动开展的本意。因此,想要进一步提高绩效评估活动的使用效果,就应该多关注于绩效评估结果使用在政府管理方面的贡献。

c. 通过绩效评估对组织成员进行培训的效果不大。在各种途径中,地方政府最善于利用绩效评估的机会,向上级领导和同级人大说明其成就,表明自身在可持续发展和争取公众满意方面取得的成绩,以进一步增加获得晋升的机会。作为评估激励方式的一种,为公务员提供培训机会却只是组织学习的要求,公务员虽然可以通过技能或知识的提高,了解到自身优势、劣势和潜力的空间。但是,该方式对组织成员职位的提升与权力的增加没有直接的作用。使得该种方式对组织成员的激励效果受到限制。

d. 以绩效评估为工具,将评估结果运用于预算上还比较困难。专家们普遍认为在现阶段使评估结果服务于预算的使用率和优先权很低,这说明因为体制和观念上的原因,专家认为将评估结果与预算挂钩在现阶段是不可行的。这也在一定层面上说明为什么我们国家试图引入绩效预算的努力到现在都没有太大的成效。

参 考 文 献

- [1] Lee R D. The Use of Program Analysis in State Budgeting: Changes Between 1990 and 1995[J]. Public Budgeting & Finance, 1997, (17): 18 - 27
- [2] Xiaohu Wang. Performance Measurement in Budgeting: A Study of County Governments[J]. Public Budgeting & Finance, 2000, 20(3): 102 - 118
- [3] Kaplan RS, Norton D P. The Balanced Scorecard Measurement

- that Drive Performance[J]. Harvard Business Review, 1992, 2(1): 71 - 79
- [4] Busby J S, Williamson A. The Appropriate Use of Performance Measurement in Non - production Activity[J]. International Journal Operation and Production Management, 2002, 20(3): 336 - 358
- [5] 邱百鸣. 对于企业绩效评估结果的分析[J]. 企业研究, 2004(11): 69 - 70
- [6] 李辉. 关于我国政府绩效评估结果使用问题的研究[J]. 山东省经济管理干部学院学报, 2008(8): 22 - 24
- [7] 段延锦. 结果使用、组织学习与组织绩效: 基于西部某市行风评议的探索性研究[D]. 西安: 西安交通大学, 2008
- [8] 爱德华·拉齐尔. 人事管理经济学[M]. 北京: 北京大学出版社, 2000, 1: 216
- [9] Loch CH, Tapper. Implementing a Strategy - driven Performance Measurement System for an Applied Research Group[J]. Journal of Product Innovation Mangement, 2002, 19(3): 185 - 198
- [10] Bruns, William J. Responsibility Centers and Performance Measurement[M]. Boston, MA: Harvard Business School, 1993
- [11] Behn, Robert D. Why Measure Performance? Different Purposes Require Different Measures[J]. Public Administration Review, 2003, 63(5): 586 - 606
- [12] Meyer M H, Roberts E. Focusing Product Technology for Corporate Growth[J]. Sloan Management Review, 1998, 29(4): 7 - 16
- [13] Mascarenhas R C, Searching for Efficiency in the Public Sector: Interim Evaluation of Performance Budgeting in New Zealand[J]. Public Budgeting and Finance, 1996, (16): 13 - 26
- [14] Melkers, Julia, Willoughby, Katherine. The State of States: Performance - based Budgeting Requirements in 47 of 50[J]. Public Administration Review, 1998, 55(1): 66 - 73
- [15] 周雪光. “逆向软预算约束”: 一个政府行为的组织分析[J]. 中国社会科学, 2005(2): 132 - 143

(责编:白燕琼)

(上接第 18 页)

估体系。

理论工具和模型的应用程度有限。调查发现,在学界颇受关注的平衡计分卡、绩效棱柱和通用评估框架在中国市级政府中的应用前景并不看好。如何将这些评估工具与中国国情结合起来,尚需进一步研究。引入国外和企业中先进的评估工具是未来政府绩效评估中的一个难点,这需要充分考虑在中国国情和行政体制下如何对这些评估工具进行改良。地方政府需要在绩效评估较为完善的基础上进行创新,并且选择评估基础良好的部门进行试点。

参 考 文 献

- [1] 吴建南, 阎波. 地方政府绩效评估体系的路径选择——福建

的分析[J]. 中国行政管理, 2008(2): 25 - 29

- [2] 叶裕民. 论改进和完善地方政府目标管理责任制[J]. 统计研究, 2000(6): 35 - 39
- [3] 吴建南, 高小平. 行风评议: 公众参与的政府绩效评估研究进展与未来框架[J]. 中国行政管理, 2006(4): 22 - 25
- [4] 吴建南, 庄秋爽. “自下而上”评估政府绩效探索: “公民评议政府”的得失分析[J]. 理论与改革, 2004(5): 69 - 71
- [5] 吴建南, 郭雯菁. 绩效目标实现的因果分析: 平衡计分卡在地方政府绩效管理中的应用[J]. 管理评论, 2004(6): 22 - 27
- [6] 孙迎春, 周志忍. 欧盟通用绩效评估框架及其对我国的启示[J]. 兰州大学学报(社会科学版), 2008(1): 34 - 43
- [7] 张小玲. 国外政府绩效评估方法比较研究[J]. 软科学, 2004(5): 1 - 4
- [8] 周志忍. 政府绩效管理研究: 问题、责任与方向[J]. 中国行政管理, 2006(12): 13 - 15

(责编:刘武英)