

制定教育评价标准的若干方法分析

沈玉顺¹ 卢建萍²

(1. 华东师大教育科学学院; 2. 华东师大理工学院, 上海 200062)

[摘要] 评价标准的制定是评价方案设计的核心。制定教育评价标准的方法有多种, 不同的方法有不同的依据, 不仅包含着不同的价值选择, 而且包含着评价方法论的选择。本文分析了文献中常见的八种教育评价标准制定方法的原理及其优缺点, 并对实践中如何运用这些方法确定教育评价的标准做了讨论。

[关键词] 教育管理; 教育评价; 评价方法

[中图分类号] G449 [文献标识码] A [文章编号] 1002-5111(2000)02-0021-06

评价从本质上说是一种价值判断活动, 而评价标准则集中体现了评价活动所依据的价值准则。正是评价标准所蕴含的价值准则, 使具体的评价活动获得意义, 决定了评价结论。评价标准的制定, 是所有类型评价方案设计的核心。

制定教育评价标准的方法有多种, 不同的方法代表了对教育的价值及如何增进这些价值的不同理解。本文以教育机构评价为例, 讨论几种常见的教育评价标准制定方法的原理及其实践意义。

一、“教育目标”作为教育评价标准

将“教育目标”作为评价的标准, 是教育评价文献中最常见的决定标准的方法。这一方法源于早期西方教育评价的研究和实践。在西方被称为“教育评价之父”的美国学者 L·泰勒所倡导的教育评价思想, 即所谓的教育

评价的“泰勒模式”, 被一些教育评价研究人员直接描述为“行为目标模式”。这一模式的基本特征, 就是以预先确定的教育目标为标准, 对教育活动的成效做出评价, 即实际教育教学活动的价值按其实现目标的程度加以衡量。以教育目标为标准的教育评价实践, 强调的是教育活动对教育目标达成的贡献, 注重的是实际结果与目标的“一致”或“符合”。

目标是任何有目的、有计划、有组织的人类活动的基本依据。用目标规范组织及其成员的行为, 保证既定目标的达成, 是多数组织管理的基本任务。将教育目标作为教育评价的标准合乎逻辑, 在很大程度上也与人们的经验相吻合。

按照这种方法决定教育评价的标准, 是以教育目标的正确性为基本前提的。因而该方法特别重视教育目标的设计和陈述, 要求目标的表述尽量精确并能够测量。但其存在的问题是: 谁来决定正确的教育目标? 教育

目标的正确性能否预先决定？教育机构是否存在普遍赞成的、具有实质性意义的组织目标？教育目标能否有效地规范教育教学活动？上述问题都是有效的管理实践所必须解决的。

教育是一种复杂的、专门化的社会实践活动，具有周期长，见效慢、不可控制的影响因素多等特点，因而事先确定合理的教育目标并非易事。以“教育目标”为标准评价教育活动，必须特别注意教育活动的上述复杂性质，考虑教育活动预期和非预期的效果。

二、“附加值”作为教育评价标准

所谓“附加值”(Value-added)，即一定时期的学校教育对学生成长发展所带来的积极影响。“附加值”作为一种教育评价标准，源自 70 年代初期首创于美国东北密苏里大学的“附加值评估”方法。该方法试图确定一定时期内学校教育活动对学生增加的价值。其基本假设是，学生入学时的水平与毕业时水平的差异，或学生在校期间的变化情况，可以归因于学校教育；学生变化的幅度，即“附加值”的大小，可以看作是学校、培养计划、课程或教师的教育成就。

实践中，确定“附加值”的基本方法，是对学生的学习成绩进行历时比较。学生在特定学科领域、特定时期两次考试成绩的差异，经过统计上的调整，即被认为是学校教育对学生增加的价值。附加值评估的主要依据，是学生特定学科的标准化考试成绩。

通过“附加值”评估确定学校教育活动的绩效，是一种与传统的教育评价方法迥异的新方法。“附加值”作为教育评价的标准，直接指向学校教育对学生成长发展的实际影响，并据此判断教育活动的价值，对于鼓励学校和教师努力增进教育对学生成长发展的贡献，无疑具有积极的意义。

但是，与“附加值”作为评价标准相伴的，有一系列困难的问题。如：将学生考试成绩的变化作为“附加值”，可能导致学校和教师

提高教育质量的努力集中于提高学生特定学科的考试分数；学生成长发展具有多方面的含义，从最好的意义上讲，考试成绩只能评价其中的一部分；而考试分数究竟能否反映学生在特定学科的学业成就，还有很大的争议；学生的发展变化能否全部归因于学校教育，人们看法不一；学生在不同时期的发展水平，是否属于同一层次的内容、能否相互比较等问题，也是一个值得研究的问题。此外，不同学科的教育对学生的“附加值”是否相同？对学生“附加值”大的教育活动是否对社会的“附加值”同样大？

附加值评估方法形成的动因，主要是为了满足政府核定学校教学效能(accountability)的需要，其主要功能是证明学校的教育成就。因此，它很少能够提供关于教育过程的信息，对学校改进教育过程的帮助有限；这种方法主要依赖学生的考试成绩，技术上精确度高，评价本身并不鼓励教师的参与。从证明的角度看，附加值评估的结果要与其他学校相比才有意义，因此，这种方法对于标准化考试的依赖性很大。但由于学生入学水平的不同及入学选拔政策的差异，不同学校之间“附加值”的比较也存在问题。比如，入学选拔性高的学校的“附加值”可能不如入学选拔性低的学校，但由此并不能得出前者对学生成长发展贡献较小的结论。在这种情况下，将“附加值”作为评价标准，难以真正反映学生的成长发展情况，因而也难以合理地评价相关教育活动的绩效。将“附加值”作为评价标准，必须合理处理上述问题。

三、“顾客满意”作为教育评价标准

“顾客”的概念频繁出现于教育管理文献，并流行于许多西方国家乃至我国的校园，主要是 1990 年代中期以来发生的现象。“顾客”这一概念被引进教育领域，“顾客满意”被当作教育评价的标准，是教育领域“消费主义”思想兴起的标志之一。这种方法将教育看作是向其顾客提供“教育”这种特殊服务的

活动,并用顾客的标准评价教育机构服务的质量。“顾客满意”作为评价标准,强调教育的社会价值,强调教育机构的服务功能,强调学校要对社会需要——尤其是学生的需要——保持敏感并及时做出反应,对于促进教育与社会实际生活的结合,推动教育的发展无疑具有积极的意义。

但是,谁是教育的“顾客”?这是一个看似简单,而事实上却十分复杂的问题。通常认为,学生是教育的“顾客”;但毕业生的雇主是否是教育的“顾客”?高层次的教育机构是否属于低层次教育机构的“顾客”?政府、教育拨款机构、外部评估机构、赞助教育活动的社会团体或个人是否属于教育的“顾客”?

将“顾客满意”作为教育评价的标准,还会引起其他一系列复杂的问题。“顾客满意”作为教育评价标准,是以顾客有能力对其自身的利益做出准确判断为假设的。在工商业活动中,“顾客总是对的”是一条“定律”。但在教育领域,顾客是否总是“对的”?很显然,学生作为受教育者,并不一定能够很好地判断自己在教育中的利益及其最佳实现方式。即使学生能够判断,作为公共服务机构的学校以“顾客满意”为标准决定教育教学活动,是否符合公众利益和社会的长远利益?教育有许多“顾客”,使不同的“顾客”满意的标准显然并不完全相同,甚至相互冲突。那么,教育机构应优先满足哪些“顾客”的要求?如何协调不同“顾客”之间的利益冲突?

当代社会对教育的期望之高、要求之多,在历史上是前所未有的。如据英国学者对1980年代以前英语文献中关于高等教育目标的统计,关于高等教育的目标陈述多达1500种。教育“顾客”的要求是否全部合理?学校是否有义务满足“顾客”的所有要求?学校能否满足如此多的要求?上述问题都是以“顾客满意”为教育评价标准这一方法所必须回答的。

四、“良好的教育实践”作为教育评

价标准

根据以往的经验,将人们普遍认为是“良好的教育实践”作为教育评价的标准,是实践中常用的制定标准的方法。我国高校的本科教学评优评估标准,事实上就是按照这一方法确定的。“良好的教育实践”通常是教育工作者所熟知的,易于理解和定义,因而比较容易得到认同和接受。这种方法假定,以前起作用的“良好的教育实践”具有普遍的适用性,而且在将来也会同样有效。

“良好的教育实践”这一概念外延宽泛,可以涵盖教育活动的整个过程。而在实践中,确实常用于教育过程的评价。这一特点使评价活动能够指向教育过程的改进。比之哪些仅仅注重输入或结果评价的方法,对学校改进工作更有帮助。但从操作的角度看,“良好的教育实践”通常存在于绩效良好的学校。这些学校通常具有较好的师资、设施设备、生源、财政等办学条件和良好的社会声誉等优势,这些有利条件可能是大部分学校所不具备的。因此,评价实践中所谓的“良好的教育实践”,通常只能是“比较好”的、多数学校经过努力可以达到的标准。这一做法会使不同学校之间的差异合理化,使那些达到标准的学校安于现状,对于学校追求更好的实践可能产生不利的影响。

“良好的教育实践”作为评价标准的另一个缺陷是,一旦它们成为组织的价值规范,会妨碍学校对科技发展、社会需要变化、教师队伍变化、学生群体变化等因素的适应能力,从而演变成为学校改革和创新的阻力。此外,由于学校具体条件的特殊性,“良好的教育实践”有时可能确实并不一定适合本校的实际。因此,运用这种方法制定教育评价标准,首先必须研究何谓“良好的教育实践”,确保依此制定的评价标准确实是“良好的”。

五、“最优”作为评价标准

“最优”作为评价标准,源于工业质量管理中的“基准法”,即将本地区、本国甚至全世

界同类学校中“最好的”实践作为评价标准,并用其衡量本校的教育活动。这种方法假定,存在适用于所有情形的最好方法。这种方法不允许将自身与其他组织的差别合理化。相反,它认为,自身与“最优组织”的根本区别,就在于没有采用与其相同的方法。

“基准法”的含义是“持续不断地寻求能够创造最高绩效的最佳实践”。“基准法”属于一种质量改进策略,一个寻求确认绩效一流的特殊实践或组织,以及这类组织如何获得其示范地位的发现过程。该方法的目标是,为管理者提供一种用于评价其组织内部活动的外部参照标准,其根本目的,是获得持续的竞争优势。我国学者关于建设世界一流大学的研究策略,所采用的事实上就是这种方法。英美等国流行的“大学排行”所提供的有关评价数据,就是所谓的“最优大学”或“最优系科”的标准。

“基准法”的优点是,首先,它可以为特定学校的教育活动建立较高的标准,提高组织及其成员的期望水平;其次,通过确认和检查同类组织高质量的产品和服务,可以促使自己持续不断地努力改进。“基准法”属于一种过程驱动的方法,它允许任何类型的组织将其自身的绩效和实践活动与可得到的最佳绩效标准进行比较,以帮助管理人员理解用于达到这些标准的活动。

在教育领域,以“最优”作为评价标准可能产生的主要问题是,学校的活动及其运行环境与企业有很大的区别,学校之间存在的许多差别,主要是由体制性因素而非学校管理本身造成的;教育活动的输入、过程与输出要素众多,找到关键性的“最优要素”并不容易;影响教育活动的因素很多,而其中有许多因素是学校或教师无法有效控制的。此外,与“良好的教育实践”作为评价标准相比,“最优”作为评价标准的适应范围更为狭窄。对多数学校来说,追求普遍意义上的“最优”,可能并不是合适的管理目标。因此,在采用这种方法时,确定“最优”的范围也是相当重要的。

六、“自身进步”作为教育评价标准

将自身的进步情况作为评价标准,是指将一段时期内组织或其成员取得的成就,与自己以前的成就相比,事实上是以评价期开始时的现状为依据衡量进步情况。新的评价以上一时期的成就结果为基准。因此,在以“自身进步”作为评价标准的方法中,标准本身处于变化之中,而且呈现不断提高的倾向。以“自身进步”为标准的评价,以发展为导向,强调评价对象及其环境的特殊性,注重评价对象及其环境中可以控制的因素,目的在于鼓励组织及其成员发现和发挥潜力。

将自身进步作为评价标准,对于评价的范围没有限制,也不要求评价技术上的精确性。这种方法通过对教育机构各方面进步信息的收集和分析,确认组织或其成员的发展成就,鼓励组织及其成员不断进步、不断提高。它不排除任何组织成员的参与;相反,它需要所有组织成员的积极参与。由于该方法将评价的焦点集中于对成功因素的鉴别,任何进步都将得到肯定和鼓励,因此不会对组织及其成员构成威胁,有利于评价目的的实现。

这种方法的主要缺点是:第一,它要求将评价的注意力集中于组织的成功之处,可能会使努力的重点指向易于获得可见成就的活动,从而忽视教育过程中一些真正重要的东西;第二,由于评价标准来自组织自身,缺乏必要的外部参照标准,可能会使组织及其成员对其自身的进步情况难以做出客观、准确的判断,从而影响组织取得更大的成就。

七、“生产力”作为教育评价标准

所谓“生产力”,即投入与产出之比,也就是人们通常所说的效率。将“生产力”作为评价标准的动因,在于提高教育资源的利用率,其背景则是在社会对教育的需求持续增长、期望不断提高的同时,公共教育财政支出难以满足教育发展的需要。因此,实践中以“生产力”作为教育评价标准的方法,通常以学校或校内的教学研究部门为评价对象,试图评

价其资源利用情况。八十年代以来,在政府支持下,许多西方国家的有关机构开发了一些旨在评价学校生产力的所谓“绩效指标”(Performance Indicators),并将学校的绩效评价结果与公共教育资源的分配联系起来。实践中,典型的绩效指标通常是定量的。

以“生产力”为评价标准的方法,源于经济学中的成本—效益分析。这类方法将教育活动看作是将一定的输入转换成输出的过程。因此,它所评价的,事实上是教育系统将特定的输入转换成特定输出的能力。该方法关注的核心问题,是提高资源的使用效益。对于成本观念一向比较淡薄的教育机构来说,这种方法有助于提高其成本意识,促使其更好地使用有限的教育资源。

但是,这种方法也存在许多有待深入研究的问题:其一,教育的输入、输出要素很多,其配置机制十分复杂,确定特定教育系统的输入、输出要素并不容易。一些经济学家在关于教育成本—收益的分析中所确定的输入输出要素,有时多达数十个,甚至上百个,使得这类分析成为一件十分复杂的工作。由于教育的产出比较难以定义和测量,因而客观准确地评价学校的生产力,是相当困难的事情;其二,在教育领域,成本概念含义比较复杂,确定合理的生产力指标的依据目前并不清楚。许多人认为,与经济活动相比,教育活动的成本并非越低越好;其三,由于上述原因,评价结果用于决策的可能性受到许多限制;其四,评价标准的量化,会使学校只关心可以量化的指标,容易忽略教育活动中一些重要但却难以量化的因素;其五,将“生产力”作为评价标准,从其方法论上说不需要关注教育过程,因而评估活动对于教育过程的改进很少有直接帮助。西方一些国家运用绩效指标评价学校的经验表明,绩效评价具有正负两方面的效应。在西方国家,以“生产力”为标准评价学校,还是一种有争议的方法。

八、“组织质量”作为教育评价标准

在复杂多变的当代社会,组织适应环境的能力是其生存和发展的关键因素。所谓“组织质量”(Institutional Quality),事实上是指组织取得预期成果的能力。传统上对于教育的评价,多以教育活动的输出或输入为对象,关注的是特定的方案、课程或学生学习的条件或结果。将组织取得期望成就的能力作为评价的对象,只是新近才受到研究人员和实际工作者的重视。所谓“高质量的组织”,就是能够把握机遇,在其选定的领域或方向迅速取得突破性成就,获得竞争优势的组织。发展到第三代的全面质量管理(TQM)所追求的,就是这样的组织。美国的马尔科姆·巴尔里奇国家质量奖(Malcolm Baldrige National Quality Award)、日本的戴明奖(Deming Award)以及欧洲质量奖(European Quality Award)的评选方案,分别提供了“高质量组织”的标准。1987年由国际标准化组织发布的ISO9000质量保证标准,也属于这样的标准。

“组织质量”作为教育评价标准,将评价的焦点从传统上对具体活动及其结果的关心,导向对组织取得预期成果的能力的关心,使评价的指导思想发生了重大的转变。这一方法的基本假设是,各种组织活动并非一个个孤立的事件;影响组织各种活动及其效果的因素,也不是独立起作用的,它们是组织适应环境变化,有效满足预期需要的能力的体现。高质量的组织具备迅速有效地满足顾客需要的能力。比如,1996年的美国马尔科姆·巴尔里奇国家质量奖评选方案,从“质量领导”、“质量信息及其分析”、“战略质量规划”、“人力资源开发与管理”、“过程质量管理”、“质量与操作结果”以及“以顾客为核心和顾客满意”等七个相互联系的方面,确认了高质量组织应该具备的标准。

1980年代末以来,上述质量标准对教育评价研究和实践产生了重要影响。英美等国及我国的一些高校在其管理实践中,分别采用了上述标准。这类标准产生于关于“最优管理实践”的理论和经验研究,明显属于较高

的标准之列。

这一方法的优点是,将评价的重点集中于组织的能力,超越了对具体方案、计划、活动或组织成员个体的关心,旨在提高整个组织适应环境变化的能力,使组织能够迎接任何挑战。在以快速变化为特征的当代社会,这种方法无疑是一种立意较高的评价策略。从其内容来看,为评价组织的质量,要对组织各方面的情况进行全面的、系统的、长远的考虑。这种方法事实上可以看作是对前述各种方法的综合运用。

从实践的角度看,这一方法存在的问题主要在于,所谓的“组织质量”标准,通常源自对工业组织的研究结果,由此获得的评价标准是否适用于教育机构,还是一个需要通过实践检验的问题。从一些高校的经验来看,采用工业质量标准评价教育机构的做法,不易为学术人员所接受。由于教育活动的特殊性,寻找一种能为所有学校接受的“组织质量”评价标准可能是困难的。

前面主要从方法论角度对几种常见的制定教育评价标准的方法做了分析讨论,旨在辨明确定教育评价标准的各种方法的基本原理及其优缺点。教育评价是一种复杂的社会活动。教育评价理论和实践中涉及到的价值问题,也许比任何其他类型的社会活动都要多、都要复杂。教育评价标准的制定,不仅反映了人们对于教育价值的理解和选择,而且反映了人们对于增进其价值的策略与方法的理解和选择。从实践的层面看,教育评价的标准事实上还决定了教育评价活动的实际效果。了解评价标准的性质和功能,是有效实施评价的关键因素。任何希望运用评价手段改进教育的努力,都必须准确地把握教育评价的原理和方法,做到手段与目的一致。否则,不仅难以取得预期的结果,而且可能事与愿违,甚至得不偿失。

[参考文献]

[1] Allen, M. (1988), *The Goals of Universities*,

SRHE & Open University Press.

- [2] Camp, R. C. (1989), *Benchmarking-The Search for Industry Best Practice that Lead to Superior Performance*, Quality Press.
- [3] Cave, M., Hanney, S., Kogan, M., and Trevett, G. (1988), *The Use of Performance Indicators in Higher Education: A Critical Analysis of Developing Practice*, Jessica Kingsley Publishers.
- [4] Craft, A. (ed.) (1994), *International Developments in Assuring Quality in Higher Education*, The Falmer Press.
- [5] Kells, H. R. (1993), *The Development of Performance Indicators for Higher Education: A Compendium of Twelve Countries*, OECD.
- [6] Madaus, G. F., and Stufflebeam, D. (1989) (Eds), *Educational Evaluation: Classic Works of Ralph W. Tyler*, Kluwer Academic Publishers.
- [7] Kogan, M. (ed.) (1989), *Evaluating Higher Education*, Jessica Kingsley Publishers.
- [8] Merli, G. (1996), *The Third Generation Total Quality Approach*, The TQM Magazine, Vol. 8, No. 6, pp30—36.
- [9] Moodie, G. C. (ed.) (1986), *Standards and Criteria in Higher Education*, SRHE & NFER-NELSON.
- [10] Meyerson, J. W., and Massy, W. F. (eds.) (1994), *Measuring Institutional Performance in Higher Education*, Peterson's Guides, Inc.
- [11] Tomlinson, H. (ed.) (1992), *In Search for Standards*, Longman.
- [12] Tuijnman, A. C., and Postlethwaite T. N. (1994), *Monitoring the Standards of Education*, Pergamon.
- [13] Wilbrink, B. (1997), *Assessment in Historical Perspective*, Studies in Educational Evaluation, Vol. 23, No. 1, pp31—48.
- [14] Winn, B. A., and Cameron, K. S. (1998), *Organizational Quality: An Examination of Malcolm Baldrige National Quality Framework*, Research in Higher Education, Vol. 39, NO.5, pp491—512.

(本文责任编辑:王 俭)