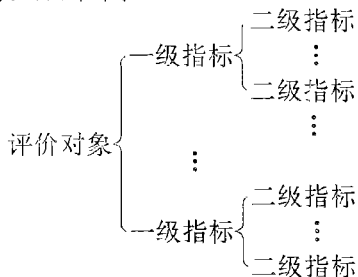


论教育评价指标体系的构建

●李 方

指标体系是教育评价的依据,构建评价指标体系是教育评价方案设计的中心环节,也是难度最大的一道工序。目前人们已经意识到教育评价指标体系的重要性及其构建之困难,但对其结构与构建的研究尚不够。本文主要从理论上讨论教育评价指标体系的结构及其构建问题,旨在为科学而有效地设计教育评价指标体系探明路向。

教育评价指标体系由反映评价对象内涵的指标集及其评价标准和量化符号构成,其主体框架是指标。“评价对象”与“指标”是相对而言的,如图所示。



上图的“一级指标”既可以看作说明“评价对象”的指标,也可以看作“二级指标”的评价对象;若将“二级指标”再行分解,则该“二级指标”又成为下一级(即三级)指标的评价对象。同时,若将上图看作一个更大指标体系中的子体系。那么上图的“评价对象”则转变为更大的指标体系中的一项指标。所以说,用来反映、说明或测量上一级项目的下一级项目

为“指标”;被反映、说明或测量的上一级项目为“评价对象”(上图由左至右表示由上一级项目到下一级项目)。在指标体系中,指标的级数越往上,指标越抽象、概括;指标的级数越往下,指标越具体、独特。

教育评价指标体系可分为一般模型和分类模型(以下简称为“两种模型”)。一般模型是评价某一范围内一般现象的指标体系,它适用于该范围内所有被评价者的评价与比较;分类模型是评价某一范围内特殊现象的指标体系,它适用某一类被评价者的评价与比较。在同一评价范围内,两种模型可以共存。

“一般模型”与“分类模型”是相对而言的。当我们说某评价指标体系为一般模型时,是相对于它所界定的评价范围以内的各种特殊情况而言;当我们说它为分类模型时,是相对于它所界定的评价范围以外的更大范围中的一般情况而言。

两种模型之界说,理清了教育评价指标体系的纵向关系(一般模型与分类模型之间的相对关系为纵向关系),拓展了构建评价指标体系的视野,使我们在构建教育评价指标体系时,既联想到一般情况,又注意到特殊情形,把一般与特殊结合起来,从宏观上、整体上去把握教育评价指标体系构建的路向。

两种模型内部均由不同类别的指标、评价标准、量化符号构成。

从达标水平去考察,有必达指标与期望指标。布卢姆(B·S·Bloom)等人曾把课程目标分为可能达到的目标和希望达到的目标两

类^①。美国管理学家凯普纳(C·H·Kepner)和屈力柯(B·B·Tregoe)将决策目标分为必达目标和期望目标两类^②。教育评价指标体系中的必达指标与期望指标,与上述两类课程目标及两类决策目标的含义相似。必达指标与期望指标的设定分两种情形,一种情形是:必达指标与期望指标是评价同一项指标的指标,必达指标是达标下限而期望指标为达标上限,即达标的最低数值为必达指标,而达标的最高数值为期望指标。如,对教师学历、学校设备等项目,可设置必达指标和期望指标。另一种情形是:某些评价项目本身就是必达指标或期望指标。如,对违纪犯罪率的最低限度的规定为必达指标,达到最低限度规定者,可得若干分,否则扣除若干分。对各种获奖、受表彰、特殊贡献的加分,则属于期望指标。必达指标的设定要充分考虑教育方针政策上的要求和实际上的可能性;期望指标的设定应有超前意识,充分考虑其对评价对象所产生的激励和导向作用。

从指标内涵性质去考察,有稳定性指标和变动性指标。稳定性指标是在相当长一段时期内涵保持不变的或者在指标体系中仍然保留的指标。它反映教育的客观规律对教育工作提出的共同要求,不会因时间、地点的变化而更换或取消。如,“教学方法”这一指标是以教育规律为评价依据的,在相当长一段时期内,不同国家或地区,教学方法之含义,以及评价教学方法的有效性的标准是不变的、一致的。评价教师的教学工作都必须设置“教学方法”这一指标。在教育评价指标体系中,由于存在着稳定性指标,使指标体系具有相对固定的内部结构。变动性指标是在不同时期,可改变内涵(包括改变评价标准)或可更换的指标。它由一定时期一定社会的政治经济状况、生产力、文化、科技发展水平对教育工作提出的不同要求所导致。例如,“德育”这一指标,不同社会制度下,其内涵和评价标准是不同的。又如,“计算机辅助教学”这一指标,适用于经济发达、教育条件较好的地区的学校,而不一定适用于贫穷落后,教育设备条件极差的地区的学校。再如,“教学语言”这一指标,由于各国或不同地区,不同民族的传统和文化背景不同,“教学语言”这一评价指标的内涵也就不同。由于变动

性指标的存在,使教育评价指标体系具有一定程度的可塑性。设置变动性指标愈多,其可塑性愈大。对变动性指标的研究,使我们注意用发展变化的眼光看待评价指标,面向未来,面向世界,并从实际出发,创造性地构建教育评价指标体系。

从指标的精确度去考察,有硬指标和软指标。硬指标是指达标要求是固定的精确的指标,它强调评价时看“硬件”,包括实物、证件、统计数据、静态资料等。如“教学设备”、“场地”、“获奖情况”、“教师学历”、“学生考试成绩”、“教师队伍年龄结构”等均属硬指标。对硬指标的评价一般比较客观。软指标是指达标要求伸缩性较大、不精确的指标。它强调评价时看“软件”,常采用定性分析评价,包括对事物的性质特点、变化过程、因果关系的分析评价。如“办学特色”、“决策水平”、“教学思想”、“教改状态”、“教育经验的先进性”等均属软指标。软指标具有较多的主观因素。但由于教育现象的复杂性和模糊性,导致教育评价不可避免要使用软指标。硬指标与软指标结合使用是教育评价指标体系结构的普遍现象。

从不同角度去考察指标的性质特征,从而将指标分成不同种类,各类指标之间常出现交叠现象。如某些必达指标,可能同时又是稳定性指标或硬指标。对指标进行分类研究,目的不仅在于启发我们恰当地从各类指标中选择某些类型指标去合理地构建教育评价指标体系,而且还在于指导我们从不同角度去考察同一项指标的内涵,以便加深对某一项指标性质特征的认识,从而有助于评价标准的确定。

评价标准是评价教育实际达到指标程度(或情形)的具体要求。达标程度可分为不同等级,不同等级有不同的评价标准。在教育评价指标体系中,末级指标必须明确规定评价标准。目前,评价标准也有不同分类法。一般从评价内容上分类,如分为素质标准、职责标准和效能标准^③,或分为状态标准与效果标准。我认为,除了内容标准外,还应有参照标准,即从评价的参照点角度去分类,分为社会标准与科学标准。

社会标准是指以社会对教育的规范要求为依据来确定评价标准,它要求达标过程做到社

会规范化。具体说来,评价标准是根据教育方针、政策、法规,以及政府有关部门对教育的统一要求而确定的。它强调价值判断标准的社会统一性和规范性。评价时考察应然现象。评价参照点是外部的、横向的,常与评价对象以外的社会现象作比较。对教育场地、教学设备、教师队伍结构、学生素质、师生比例等指标的评价,可采用社会标准。

科学标准是指以科学上的客观规定和原则为依据来确定评价标准,它要求达标过程符合教育工作自身发展的客观规律,以教育工作过程、方法、措施、结果的科学水平来衡量其价值。评价时,从评价对象自身的实际出发,仔细考察和分析教育工作过程、方法、措施等是否符合科学要求,是否遵循评价对象自身发展的客观规律;判断教育效果是否准确和可信。强调实然现象,常以实践为标准来检验工作过程的科学性。将事物发展的结果与事物发展的初期作比较分析,参照点是内部的、纵向的,是自比而不是他比。譬如对“教改实验”这一项指标的评价,宜采用科学标准,以科学标准来评价实验设计和实验过程。除此之外,“学校领导决策”、“教师工作效能”等指标亦可采用科学标准。

量化符号是教育评价指标体系结构中不可忽略的部分,一般有权数和分数两类^①。这两类数值是用来反映某一个体在整体中的相对地位的。权数常用小数形式,一般把同一级指标群集视为一个整体,整体权数总值为1。此外还有其它形式如百分数等。分数包括指标赋分和等级赋分两种。同一级指标赋分,满分为100分。等级赋分可使用达到度,如,A、B、C、D四等的达到度为0.95、0.85、0.75、0.65。在教育评价指标体系中,既可以单纯使用这两类数值的其中一类,也可以将这两类数值结合起来使用。

对指标、评价标准、量化符号等内部元素的分析,使我们理清了教育评价指标体系的横向关系(横向关系指:指标体系的同一级指标群集中,各类指标之间的关系,各项指标的权数或分数之间的关系,末级指标的各类评价标准之间的关系,各评价等级的分值之间的关系,各类量化符号之间的关系),加深了对教育评价

指标体系内部结构的认识,有利于从微观上、从个别特质上去掌握教育评价指标体系的构建方法和技术。

二

构建教育评价指标体系首先要考虑的是信度与效度问题,因为信度和效度是教育评价指标体系实施的前提条件,是衡量教育评价指标体系构建质量的重要尺度。

关于教育评价指标体系的信度,本文拟从三个方面去定义。(1)同一评价者(指评估组或个人,下同)运用同一教育评价指标体系去重复评价某一对象,考察所得结果的一致程度,可称之为重评信度。如果使用同一个指标体系去评价同一对象,在相隔较短的时间内实施两次(或两次以上),两次(或两次以上)的评价结果越接近,则信度越高,相差越大,则信度越低。(2)不同评价者运用同一种教育评价指标体系在相隔较短的时间去评价同一对象,考察不同评价者所评出的结果的一致程度,可称之为评价者信度。(3)同一评价者,使用两种(或两种以上)教育评价指标体系,几乎同时评价同一对象,考察两个(或两个以上)指标体系所得的结果的真实程度,可称之为事实信度。哪一个评价指标体系所得的结果能贴合实际,能评出真实情况,哪一个评价指标体系的信度就高,即被评的数据和资料的真实可靠性就强,评价结果虚假成份就少,此为信度高;反之,被评的数据和资料与事实不符,评价结果有较多虚伪成份,不能令人信服,此为信度低。

一般说来,指标内涵客观,指标与指标之间相互独立,界线清楚,硬指标比例大,评价标准的规定明确,等级之间界限分明有利于提高信度。如果软指标多,指标与指标之间,等级与等级之间界限模糊,评价者较难区分指标之间的差异,难以掌握评判准则,则会降低信度。此外,信度还受评价过程、方法,以及评价者的价值观、理解能力、专业水平、评价工作的业务训练程度等因素的影响。

教育评价指标体系的效度是指评价指标、评价标准、评价的量化符号等元素设计的准确性和代表性。它要考察的是我们想要评什么(即预期目标)与评价指标体系可能评价到什么

程度之间关系,两者愈接近、愈吻合,说明评价指标体系的各种元素的设置愈准确,愈有代表性,使教育评价工作更有效,此为效度高;相反,两者偏差大,指标体系缺乏应有的代表性,从而使教育评价工作成为无效的劳动,此为效度低,甚至无效度可言。

在实际中,教育评价指标设置不合理,指标权重分配或赋分不严谨,评价标准确定不科学,等级赋分不恰当等因素,都会降低教育评价指标体系的效度,而且有可能会误导教育工作。

教育评价指标体系效度的高低,与构建者对指标体系构建依据理解的准确程度,对指标内涵理解的透彻程度,对评价工作目的明确程度,对实际情况的了解程度等因素的关系尤为密切。

为了提高教育评价指标体系的信度和效度,构建者必须掌握有关方法和技术。世界上20世纪初对量表和调查问卷的研制^⑤,为我们构建教育评价指标体系积累了可参考的经验。以后,泰勒(R·M·Jyler)关于教育目标以及“目标达到”教育评价模式的研究^⑥,波帕姆(J·Popham)、马杰(R·Mager)和格兰兰德(N·Gronlund)等人关于各种“行为目标”的研究^⑦,加涅、梅里尔(Merrill·M·D)、达尼洛夫、叶希波夫、巴班斯基等人关于教育目标分类理论和方法的研究^⑧,特别是布卢姆等人的《教育目标分类学》、《教育评价》^⑨,以及其他有关研究成果,值得我们借鉴。近十多年来,我国学者也曾进行过有关探索^⑩,给我们以颇多启发。这里,仅在别人研究的基础上,就构建教育评价指标体系的方法和技术问题,谈谈自己的一些认识。

(一) 关于构建教育评价指标体系的依据

由于教育评价指标受社会因素及教育规律所制约,所以,构建教育评价指标体系的主要依据应包括:(1)教育方针、政策、法规;(2)教育理论和知识;(3)教育规律;(4)教育工作实际。前两条为主观依据,主要解决构建者的理念问题;后两条为客观依据,强调的是实证问题。构建指标体系时,构建者应认真查阅有关文献资料,全面掌握有关信息,明确指标体系的大方向;要深入研究教育科学理论

知识,打好扎实的理论基础;与此同时,深入实际考察,总结经验与教训,并开展调查研究,掌握教育客观规律。

(二) 关于评价对象的逐级分解

科学地逐级分解评价对象是提高评价指标体系效度之关键,但要科学而有效地逐级分解,并非轻而易举。一般情况下,分解评价对象与理解评价对象的内涵有关。只有透彻理解评价对象的内涵,才不致于分解时茫无头绪。对评价对象内涵理解愈正确、深透,愈容易分解。诚然,对评价对象内涵理解的程度,不仅与构建者智力水平(如理解能力)密切关联,还与构建者专业基础知识、专业理论功底息息相关,主要包括逻辑学、心理学和教育学方面的知识和理论^⑪。例如,欲将“师德”这项评价对象分解为若干细项指标,以便构建一个师德方面的评价指标体系,为此,构建者务必对道德结构、教师修养、德育理论等有一定的了解和研究,才能对“师德”的内涵有正确而深刻的理解,这样,才能根据评价目的之需要及实际情况,将“师德”逐级分解。

然而,有时人们对评价对象内涵的理解是多向的,可以这样理解,也可以那样理解。在这种情形下,应根据评价目的需要去选择分解路向。若评价对象内涵相当丰富,一项评价对象可分解成许多细项指标,在这种情况下,则应根据评价目的需要来确定细项指标之取舍,从而选取最重要的指标。

但是,有时根据评价对象的内涵和评价目的需要去选定指标,不一定符合实际情况,因而不一定可行。故此,务必了解实施评价指标的可能性,包括主观可能条件和客观可能条件。根据实际可能去设置指标。如果实践已经证明行不通的指标,则应进行修改或更换。

逐级分解评价对象反映了思维的层次性和复杂性。在教育评价指标体系中,由于评价对象与指标的相对关系,使我们对其的思维是多层次的。思维的层次性,使我们在教育评价指标体系的纵向上思考得更深入、更具体;由于指标类别的多样性,使我们对其的思维又是复杂的,思维的复杂性,使我们在教育评价指标体系的横向上思考得更全面、更周到。

(三) 关于评价等级和标准的确定

教育评价一般用2至5个等级,但究竟使用多少个等级为好?作者1994年从事广东省实验中学教育评价指标体系设计课题研究^⑫,曾遇到过这一问题。作者认为,等级数应根据指标的类型和性质来确定。如,评价学校领导和教师是否“坚持四项基本原则”,用二个等级就比用四个等级好,因为在实际要求上无需太精确,在具体评价时也无法太精确,这是由指标内涵的性质决定的。

目前,有的教育评价指标体系采用四级分等,两级定标的办法,将评价划分为A、B、C、D四个等级,对A、C两个等级规定具体的评价标准,A为优等,C为中等,A级和C级之间为B等,C级以下为D等。这种办法设计简单,操作方便,有可取之处。

评价标准的确定,主要解决以下两个问题。

1. 标准类别的选择。要根据不同类型指标的不同特点,综合运用各类评价标准。例如,对教改实验这项指标的评价,在参照标准上,可先确定用科学标准,然后,在内容标准上,可确定使用状态标准和效果标准。这样,就把不同类型的标准综合运用了。又如,对学生学习水平的评价,在参照标准上可采用社会标准,而在内容标准上则可采取效果标准。对教师队伍结构的评价,在参照标准上可用社会标准,而在内容标准上则可采用素质标准。

2. 各等级评价标准的规定。各个等级所规定的评价标准,界限要清楚,使评判者容易区别开来;应按各条标准的内在联系,由高至低顺次编排,显示出等级之间的递阶性。

(四) 关于指标和等级的数值分配

如何科学地确定指标与等级的数值,是构建教育评价指标体系的难题之一。国内外曾采用的做法颇多,从人员上说,已从早期的凭个人经验素确定,发展到集体讨论或咨询法,越来越重视同行和专家的意见;从方式上说,已从主观随意确定,发展到注重调查研究,按一

定科学程序进行规范化确定。目前,问题的关键是参与人员,要求他们:(1)必须具备相应的知识经验、较强的判断能力和比较鉴别能力;(2)必须明确赋值的依据,掌握好原则;(3)要有认真的态度、严谨的作风和高度的责任感,严格按照科学程序去进行指标和等级的数值分配。这三点是使指标与等级的数值分配趋向合理性、科学性的重要因素。

注:

①〔美〕布卢姆等著,邱渊等译:《教育评价》,华东师范大学出版社,1990年,第10—12页。

②马超山、温善策著:《教育决策论》,辽宁教育出版社,1988年,第213—218页。

③张玉田等编:《学校教育评价》,中央民族学院出版社,1987年,第38—42页。

④周海华、周煦编著:《教育评价及评价体系》,广西教育出版社,1989年,第三至第五章。

⑤戴忠恒编著:《心理与教育测量》,华东师范大学出版社,1987年,第一章,第三至第七章。李方编著:《现代教育科学研究方法》,广东高等教育出版社,1989年,第4—6页。

⑥龚伟民:“教育评价的发展及评价模型的演变”,载华东师范大学教科院编印《教育评价资料集》,1985年,第5页。

⑦〔英〕菲利浦·泰勒等著,王传廉等译:《课程研究导论》,春秋出版社,1989年,第19—50页。

⑧侯光文著:《教育测量与教学评价》,明天出版社,1991年,第79—83页。李秉德主编:《教学论》,人民教育出版社,1994年,第53—56页。

⑨〔美〕布卢姆等编,罗黎辉等译:《教育目标分类学》,华东师范大学出版社,1986年。布卢姆等著,王钢等译:《布卢姆掌握学习论文集》,福建教育出版社,1986年,“前言”,第1—2页。

⑩陈玉琨编著:《教育评估的理论与技术》,广东高等教育出版社,1987年,第五、六章。

⑪布卢姆提出教育目标分类的三个原则:教育学上的原则,逻辑学上的原则和心理学上的原则(参见注⑨,《教育目标分类学》第7—9页)。

⑫李方著:《现代教育管理技术》,华南理工大学出版社,1995年,第363—406页。

作者单位 华南师范大学

(本文责任编辑:宗秋荣)