

# 第一章

## 创新教育评价的 由来与发展

创新教育评价是运用现代科学的评价思想、理论和方法，以学校的创新办学思想、创新教学体系、创新学生素质及创新教育环境等为对象和要素，通过系统地搜集、整理相关信息，对其作出价值判断并为教育决策提供信息反馈的过程。创新教育评价是现代教育评价最核心的评价形式和内容。要了解创新教育评价的由来和发展，必须首先知道教育评价的产生与发展历程，使我们初步掌握创新教育评价产生的实践基础和历史渊源，为进一步探讨创新教育评价的有关理论与方法奠定知识基础。

评价是对事物的某种价值进行评定和判断的过程，它是与价值概念紧密联系在一起。价值表示主体与客体之间的关系，反映了客体在一定时空内、一定程度上满足主体需要的主、客体相关的对象性联系。凡是有一定价值的事物都存在着评价行为。因此，从广义上说，教育评价是随着教育产生而产生、随着教育发展而发展的。但是狭义上的教育评价，也就是说，具有一定界说的教育评价，它虽然渊源于中国，但正式产生于 19 世

纪 30 年代美国的“八年研究”至今还不到一百年的历史。

## 一 教育评价的渊源

教育评价作为一种衡量教育的质和量的实践活动，其思想渊源于中国。我国封建社会用一套完整的考试方法选拔人才的科举制度，设立专事巡视京外学务的“视学规程”等都属于教育评价初始活动的内容。

早在我国西周时期，统治者就建立了“国学”和“乡学”等用以培养、选拔人才的机构，学校教育中就含有如何“选士”、“取士”的考试制度。在《礼记·学记》篇中就记载了西周“国学”中的考试“选士”的制度：“比年入学，中年考校。一年视离经辨志，三年视敬业乐群；五年视博习亲师；七年视论学取友；谓之小成。九年知类通达，强立而不反，谓之大成。”这里规定了考试周期和考试内容以及“小成”与“大成”的标准，不仅有分析经义、志学专业、研讨学问和学习能力等智育方面的评定标准，而且有树立志向、团结同学、亲师择友等德育方面的评定标准。西汉以后，直至隋，选士制从学校分离出来，形成了察举制。为了选拔官吏，网罗人才，隋炀帝大业二年（公元 606 年）始创“开科取士”、“分科举人”的选拔制度，至清光绪三十一年（公元 1905 年）废“科举”止，前后持续了约 1300 年。经过历代统治者的积累与完善，科举制度形成了一整套完备的体系；从考生条件、考试科目、考试程序到阅卷方法、录取规程等，都有一套标准与措施。中国的“科举制度”通过国际交流与合作，对欧亚其他国家产生了一定影响。据研究，英国学校废除口试，代以笔试，以及英国的“文官制度”等形式，无不受到中国科举制度的影响。

中国的科举制度虽然在检测、评价考生的学力中起过重要

作用，但它无法对人进行全面、公正、科学的考察与评鉴，且自身又存在着许多弊端。上世纪末、20世纪初，我国“废科举”、“兴学堂”，但是在学校教育工作中仍沿袭了传统的考试制度。到了20世纪20年代晚期，虽有人民教育家陶行知先生倡导过教育评价，但实践的范围毕竟有限，研究也难深入。直到20世纪70年代末，尽管我国的社会制度和教育性质早已发生了质的变革，在考试制度上也借鉴了西方的教育测量的理论与方法，尝试过改革，但是在教育评价方面并未发生根本性的改变。当代经过革新的教育测验的方法与技术未能得到普遍推行。在学校教育工作中，传统的考试方法仍占统治地位，更没有发展到现代意义上的教育评价这一阶段。

## 二 西方教育评价的产生与发展

教育评价的思想基础虽然渊源于中国，但是，提出教育评价这一科学概念，进而把教育评价作为一门学科，其理论形成于美国。从西方教育评价的产生与发展上看，它是一个在不断追求评判的客观性的同时更加追求教育性（通过教育评价起到改进与提高教育、促进个性发展的作用）的过程。同时，它也是一个不断由狭义的评价逐渐向广义的评价延伸和扩展的过程。这个过程可以分为下述的三个阶段。

### 一 从教育测量到教育评价

从19世纪末到20世纪初，美国教育心理学家把心理测量原理和方法运用于教育，对教育效果的判定采用了科学量化的成绩考核法，并逐渐形成了教育测量运动。1904年，美国心理学家桑代克（Thorndike, E. L., 1874 - 1949）发表了《精神与社会测量学导论》一书。这是一本在测量学史上划时代的巨著，标志

着教育测量运动的开始。在这本书中，桑代克提出了著名的“凡是存在的东西都有数量，凡是有数量的东西都可以测量”的信条。在以后的教育测量研究与实践中，对桑代克提出的这一信条，不但在技术方面得到了正确应用，而且从理论上得到了进一步的证实。教育测量研究与实践也逐渐发展成为三个分支，即学力测量、智力测量和人格测量；其测量方法到 1928 年为止就有三千种之多。随着教育测量运动的不断发展，人们逐渐认识到，教育测量尽管能使考试客观化、标准化，并把人的知识和能力数量化，可以准确了解个体差异程度，但是，教育测量首先追求的都是客观性，因此，所有难以客观地、数量化地加以描述的事物比如社会态度、情感、意志、创造兴趣、鉴赏力等都属于教育的重要领域，却都被排除在测量之外。这样一来，测量所抓住的只是教育成果中的零散片断和局部的东西，它忽视了教育对象的全面的人格发展。

在 1930 年至 1940 年间，美国进入了测量运动的批判时期，教育测量逐渐为教育评价所代替。1933 年，在美国进步主义教育联盟 PEA 会长艾金 Aikin, W. M. 的领导下组织了由 7 所大学和 30 多所中学进行教育评价实验到 1940 年实验宣告成功。历史学家称这次实验为“八年研究”(1933-1940)。为了评价“八年研究”的成果泰勒 Tyler, R. W. 教授与其他多位教授组织了一个评价委员会，该委员会最后提出一份史密斯-泰勒报告。这份被后人称为“划时代的教育评价宣言”，第一次明确了教育评价的思想。它明确指出：对于教育的成果，不能单单测量受教育者的某些能力和其他特征，而应该评价受教育者由于教育的作用而向着教育目标的成长和发展。由此看来，教育评价的基本思想应该是：

1. 所谓教育就是促使人的行为方式发生变化和改进的过程；

2. 人的形形色色的行为方式的改善与发展的变化，就是教育目标；

3. 评价教育过程，就是衡量这些教育目标达到了什么程度；

4. 因为人的行为是复杂的，所以评价要从各个侧面进行，并且评价不仅要有分析 还要有综合。由此可见 评价的方法仅仅依靠纸和笔的测量是不够的 还须包括行为观察、思维推断等多种多样的手段。

根据这种评价思想，该委员会还以全面发展人的才能为主要目标，悉心研究和设计了一整套教育方法和考核评价方法。

史密斯—泰勒报告的思想 后来在美国发展成为“教育评价运动”。这场运动强调 不仅要评价知识的识记 还要评价应用、分析、综合等高层次的认识能力 以及评价兴趣、情感、价值观、鉴赏力、社会态度等情、意方面的心理特征。至此 教育评价更臻完备，使其成为教育科学的一门独立学科，被全美国所采用，其后传播到国外。

## 二 从教育目标分类体系的建立到目标到达度评价

1956年以来 布鲁姆 Bloom, B. S. 等人建立的教育目标分类学，已成为现代教育评价的重要基石。教育目标分类研究由两方面构成：一方面是教育学的，即对课程和教育目标的研究；另一方面是心理学的，即对能力和价值观、技能的习得和学习，内化过程等课题的研究。这两方面的研究成果，以教育学上的有效性、逻辑性和心理学的依据为准则 进行综合、整理 以具体的行为目标的形式设定出最具体的目标 然后按照发展顺序 层层上达，构成目标体系。建立这种科学的统一的教育目标分类学 使各个教育评价研究者有了共同的语言 使教育评价理论研究得到了更深入的发展，教育目标科学体系成了教育评价活动

的参照基准，使得呈正态分布的相对标准评价逐渐转化为促进个体内部发展的绝对标准评价——即目标达成度评价。这么一来，由于教育目标分类学的建立，使得教育评价完成了下述两个方面的转化。

### （一）由心理测定性评价向教育测定性评价转化

历来的心理测定性评价，对于个人的学力，往往不是就其本身进行测定和评价，而是用该人与其他人的相对位置来表示。这种以正态分布为依据的相对目标标准的测定和评价，虽然有其客观性强的优点，却缺乏检验是否达到了教育的要求。教育目标分类学建立起来之后，为人们追求评价的这一教育性原则提供了前提和条件，使之逐步完成了向教育测定性评价的转化。这个转折的标志，就是提倡绝对目标标准的测定和评价。它批判了“正态分布的、一贯而稳定的、个人与个人之间的差异”这种相对目标标准的心理测定性评价程式，采取了达到目标的形式，设定通过相应的教育活动来实现的预期的学力内容，并且以每一个人为基准来测定和评价这些目标是否到达。它关心的不是正态分布的个体之间的差异性，而是每一个人自身内部发展的进展。因此，它不但具有客观性的特点，而且具有强烈的教育性，即通过教育和教育评价促进了个性发展。

### （二）由总结性评价向形成性评价转化

教育目标分类学问世以前的所有教育评价活动，都属于总结性评价。所谓总结性评价，就是一个学期、一个教程或者一个学年结束的时候所进行的评价活动，以便进行分等、鉴定，作出评价结论，或者据此对课程、学程以及教育计划的效率进行研究和改造。这类评价的基本特点是在学习和教学完结之后，就学习或者教学的效率，对学生、教师或者课程编制者作出评判。正是这种性质的作出评判的行动，难免会同时带来在学生、教师和

课程编制者中间引起不同程度的焦虑和抵触情绪。

1967年,美国学者斯克里芬 Scriven, M. 在其所著的《评价方法论》中首先提出了形成性评价的概念。经过布鲁姆等人的丰富和发展,形成性评价已成为当今教育评价的不可或缺的重要内容。一般说来,形成性评价是指在教学过程中而不是在教学过程终结时进行的评价活动。它为修正课程提供线索,为是否需要补充教学提供依据,它向每个学生建议适当的学习任务量。形成性评价的重要性之所以受到重视,是基于人们对于评价的反馈功能的认识,也基于人们对它在教育决策上的重要意义的认识。它的特点是在教学过程之中进行的,目的是为了形成适合于教育对象的教学。因此,教学的各有关方面——学生、教师、课程编制者,均对形成性评价表示欢迎,他们感到,这类评价对帮助他们各自所想做的工作,是非常有用的。

绝对目标评价与形成性评价都追求测定和评价的教育性,其区别在于前者追求的是评价的方法上的教育性,而后者追求的是评价在应用上的教育性。这里要说明的是,凡是形成性评价都是以绝对目标为基准进行的到达度评价。

### 三 从狭义的评价到广义的评价

教育评价有狭义和广义之分,其发展过程是从狭义扩展到广义。上述两个阶段的教育评价,都指的是狭义的教育评价。一般说来,狭义的教育评价的基本特点是以分数为标准,以学生为对象,其任务仅限于学习和生活领域内对反映出学生的成长和发展状况的教育行为给予价值上的判断,以促进教学过程的变革。时至20世纪70年代,为适应现代教育科学管理的需要,人们逐步把教育评价的应用领域拓宽,使教育评价的概念外延适应于已经和正在发生的大量教育现象,因此教育评价概念的内涵由狭义扩展到了广义。广义的教育评价,涉及到教育活

动的一切方面和教育系统的各个环节。其评价的对象已由学科的教育、教学过程和结果扩展到更广阔的领域。诸如对学生、教师、学校管理者的评价等,还有诸如教育方针、教育制度、教育结构、课程设置、师资队伍、教育发展规划、教育科研的学术水平等方面的评价。教育评价的概念内涵的日益丰富和概念外延的不断扩展,说明了教育评价的产生和发展适应了当代教育建设和发展的需要,显示了教育评价的强大生命力。

### 三 我国现代教育评价的历史回顾

随着中共中央、国务院《关于深化教育改革 全面推进素质教育的决定》的全面贯彻落实 我国的教育思想、教育基本理论、教育制度、教育体制 以及教育内容、方法、技术诸方面都要发生深刻的变化,这将促使我国教育在迈向 21 世纪的进程之中逐步进入素质教育时代。在这新时代的大背景之下,我国现代教育评价也将走完第一个阶段的发展历程,步入素质教育评价时期,实现教育评价思想观念的转变和教育评价理论、模式、方法和制度的更新。为此,我们要正确处理继承与创新的关系。只有在总结、整理和深化我国现代教育评价近 20 年来理论研究和实践探索的经验的基础上,才能探索新时期素质教育评价的实践和发展,才能创新素质教育评价的科学理论体系。下面,对近 20 年来我国教育评价的理论研究成果和实践探索经验进行回顾和总结,以使初步了解我国现代教育评价的发展轨迹。

#### 一 我国现代教育评价的产生

真正意义上的教育评价的研究与实践活动,是 20 世纪 80 年代初才在我国兴起的。

党的十一届三中全会以后,我国各条战线都开展了拨乱反

正，国家建设把战略重点从以阶级斗争为纲转移到以经济建设为中心的轨道上来。在教育界，自从恢复高考制度以后，改革和完善高校统考招生制度成为一个重要课题。北京师范大学心理学系运用现代教育测验的原理和方法，对 1978 至 1981 年各年度高考各科试题进行了评价 结果发现 试题的信度、效度、难度和区分度都存在着很大的问题。在国家教委领导下，从 1985 年起，在广东省开展数学、英语两学科标准化考试改革试验；在上海进行了单独命题的探索；并在部分省区试行中学毕业会考的基础上，减少高考科目。通过改革试验，引进吸收了国外先进的教育测验和教育评价的原理与技术，在命题的科学化、客观化、标准化以及数据处理现代化诸方面有了长足的进步。同时，对于中学的教学思想、课程设置、教学内容、教学方法与手段的改革，也都产生了较大的推进作用。

1983 年，上海市教育局组织有关力量，依据布鲁姆教育目标分类学理论，编制并实施了全国第一个初中平面几何标准化测评方案，开了国内教学目标评价之先河。接着，对学生的学力评价逐渐扩展到对品德、体质的评价，在全面性、科学性诸方面都有较大进展。与此同时，普遍注意把终结性评价与形成性评价有机地结合起来，重视自我评价，充分发挥教育评价的改进、激励、调控、反馈诸功能。这一切已成为学生评价的良性趋向。总之，由于学生质量评价方面取得了明显的进展，为学校教育工作的综合评价奠定了基石。归根到底，任何学校教育工作所追求的终极目的，就是最大限度地提高教育质量，全面提高办学效益 多出人才 出好人才。

自从 1985 年中共中央关于教育体制改革的决定提出基础教育实行“地方负责、分级管理”的原则之后 教育行政管理部门如何在宏观上管理好学校，学校如何实现自身管理的科学化、民

主化，成为教育界面临的迫切需要解决的实践课题。有鉴于此，国家在‘七五’教育科研规划中明确提出了‘逐步建立系统的教育评价和监督制度’的任务，从而大大推进了学校教育评价的进展。随着国家和地方政府各级教育督导机构的建立，教育领导部门把工作的注意力和着重点同时放到了对学校教育工作的督导评价上，它促使学校教育评价朝着系统化、多样化、全方位的方向发展。

## 二 我国现代教育评价的实践与研究成果

回顾 80 年代初以来我国教育评价理论的研究与实践的探索活动，可以看到我国开展的学校教育评价，其主攻方向是围绕评价方案的设计与制定、实验和实施，在实践中修改完善以及从理论上抽象概括这一中心课题展开的。近 20 年来，全国各地研制的教育评价方案不下数千种之多。我国的研究工作不只是停留在方案的简单设计与实施上，而是通过对评价方案的研究和实践，对学校教育评价的目的、对象、定义、原则、结构、功能、过程，以至目标、指标、方法和制度等问题进行系统的探讨，逐步构建了学校教育评价的基本理论研究、指标体系研究、方法体系研究、评价制度研究等四大研究领域，力求构建具有中国特色的学校教育评价理论体系。

### （一）关于教育评价基本理论研究

近 20 年来，我国教育评价理论工作者借鉴国外教育评价优秀成果，联系本国具体实践，以马克思主义的认识论、价值论和方法论为基础，运用系统科学、逻辑学和现代数学，探讨了学校教育评价的基本属性，提示了它的科学内涵及其外延。学校教育评价是以建立在一定的教育价值观基础之上的目标体系为依据，运用现代科学方法与技术，解析学校教育系统的状态变量，对其社会价值进行判断，并为教育决策提供信息反馈的过程。

按评价的主体不同，可将学校教育评价分为自我评价和他人评价；按评价的时空变化，可将学校教育评价分为诊断性评价、形成性评价、总结性评价以及二次评价；按评价的指标标准，可将学校教育评价分为相对评价和绝对评价。同时，从教育评价内在的本质联系出发，确立基本原则；由此对其对象范畴、结构功能、过程结果等实质性问题展开了一定深度的探讨，从而初步形成了我国教育评价科学理论体系的雏形。比如：北京提出的“教育型目标调控模式”、上海提出的“协同自我评价模式”、河北提出的“指标量化、文字描述模式”以及湖北黄石提出的“两体一制模式”等都初步形成了有中国特色的教育评价理论。

## （二）关于指标体系研究

建立一套科学的、切合实际的、具有指导性的指标体系是整个学校评价工作的关键所在。指标体系研究，始终是 20 年来我国现代教育评价研究的重中之重。建立指标体系，要力求反映学校教育活动内部的本质联系。指标体系具有直观、简明的系统结构和完整地揭示学校教育工作内涵的内容。从结构内涵上说，指标体系是一个定性与定量相结合，具有层级对应关系的量标系统，它由要素集和权数集有机组成。要素集是由能反映评价客体某种特质的因素所组成的、定性与定量相结合的标准系列。权数集是指渗透于要素集之中、与每一项指标相对应的加权值，是按指标的层级对应性有序排列而成的权重数表。从结构形式上说，指标体系是由指标框架和评价要素内涵两个层级组成。指标框架包括指标序号、指标名称以及归一制权数表等部分组成。指标序号是用排序法记录指标的符号，用它来表示某一项指标在指标体系中所处的位置。指标名称是根据指标所包含的意义对指标命名，用它来概括某一项指标的内涵。归一制权数表是表示各级各类指标相互关系以及每一项指标在整

个指标体系中的地位 and 作用的指标量化数值表。评价要素内涵是关于指标名称的含义的阐释，它是用描述性语言或某种数学法则揭示指标所包含的本质特征以及评价要素的基本规格和要求。它从质的和量的方面确定了每一项指标标准的内在含义或指标标准的数值界限。指标的评价要素内涵的提示，使笼统的抽象的指标要素化、具体化和数量化，使指标体系具有可测性和可操作性。

指标体系的设计，一方面要坚持导向性、科学性、民主性、客观性、规范性和可测性等各项基本原则；另一方面要严格按设计程序进行。首先，把学校教育工作目标作为一个系统的状态变量，应用系统分析技术揭示各状态变量的本质特征和内在联系，建立目标函数；其次，对目标函数所涉及的自变量进行定性、定量分析和逐级分解，确立三个级别的指标和各指标下的评价要素序列；再次，根据每一项指标及要素各自在指标体系中的权重，分别赋予相应的加权值，然后，把它们有机地联系在一起，就构成了一个完整的指标体系，即：(1) 建立目标函数；(2) 确立三级指标；(3) 完善指标系统（揭示评价要素）；(4) 构造权集。

指标体系设计出来后，只有通过理论研讨、试点检测、综合评议等理论与实践论证，证明它具有科学性和可行性之后，方能应用于实际。

### （三）关于方法体系研究

评价方法指的是为达到某种评价目的、根据评价原则选定的有利于各种评价力量协同工作的一整套活动方式和操作程序。具体地说，它是计划、组织、调控评价实施活动的技能和手段，以及搜集、整理、优化评价信息资料的知识 and 经验。教育现象的复杂性和评价过程的多阶段性，决定了评价方法的多样性和整体性。近 20 年来，全国各地开展教育评价活动研究和选用

的方法技术有数百余种，构成了我国教育评价的方法体系。这个体系由评价方法论、评价实用方法和评价操作技术等三个层次有机构成。方法论层次以马克思主义哲学为指导，以系统分析方法为核心，并与逻辑方法、数学方法互相渗透、互为补充有机构成的方法论框架。评价实用方法系统处在中间层次，它是在方法论指导下，围绕评价活动和评价信息这两个重点内容来进行并为之服务的。为评价活动的过程管理和控制提供有效手段，为价值判断提供全面、客观、准确的信息资料。评价操作技术系统则处于方法体系中最低层次，它具体为评价管理和搜集信息资料服务，由评价方法所决定。运用各种评价方法搜集信息资料是通过许多具体的评价技术和操作工具来完成的。

学校教育评价方法体系的核心内容是评价信息管理方法系统，它由三个层次构成。第一层次，评价信息搜集整理法。它是评价者根据指标内涵，搜集、整理和优化评价信息资料过程中所采用的技术和手段。包括信息源确定法、信息获取法和信息整理法等。第二层次，评价信息测定法。它是使评价活动由事实判断向价值判断发展的桥梁或中介，运用这类方法技术，可将评价客体的质量水平确定为某一（优、良、中、差）等级。包括定性测定法和定量测定法。第三层次，评价信息计量法。它是以一种的方式将学校教育现象和过程的变化与一事先确定的标准进行比较，并依据这一客观标准，用数值或文字语言给学校教育状态变量赋“权”的过程。一般包括量化计量法和文字描述法。

对于每一次特定的教育评价活动都要从方法体系中选出适用的某几种或十几种方法、技术以构建特定的方法技术子系统。

#### （四）关于评价制度研究

在学校教育评价实践中，评价制度以原则、计划、规则、规程、政策、条例、规定等法规性、行政性、经验性或科学性的管理

手段制约评价者和评价对象的行为规范,对整个评价工作进程起到组织、控制、协调和促进作用,使评价活动遵循一定程序,有规律的发展,以保证最终实现预定的评价目标。近 20 年来,全国各地在教育评价实践中,初步建立了组织管理、程序管理和质量管理等评价制度,并在实施过程中不断得到修改和完善,逐步形成了我国的教育评价制度体系。

组织管理制度。当前,我国学校教育评价组织的建立是与一定的教育评价理论和实践研究相联系的,是为适应评价理论研究和实践活动的需要而产生的。我国现阶段的评价组织大概可分六种类型:(1)督导式评价组织;(2)行政式评价组织;(3)选拔式评价组织;(4)科研式评价组织;(5)职能式评价组织(学校自评组织);(6)矩阵式评价组织(他评、复评组织),这六种评价组织具有不尽相同的组成和结构形式以及不同的组织功能和作用,各自适应不同类型的教育评价活动组织管理的需要。

程序管理制度。学校教育评价活动的完整过程可分准备、实施和结果处理等三个阶段,每一阶段的评价工作都要严格按科学程序进行。准备阶段:(1)设计评价方案;(2)制订评价规程;(3)开展评价学习宣传。实施阶段:(4)搜集评价信息;(5)整理评价信息;(6)计量评价信息;(7)编写评价报告;(8)建立评价档案,结果处理阶段:(9)审计评价质量;(10)反馈评价信息;(11)总结经验,表彰先进;(12)诊断问题,改进工作。每一程序都规定了相应的评价对象、内容和任务以及应采取的步骤和方法,促使评价活动科学化和规范化。

质量管理制度,主要包括学校自评调控制度、评价质量审计制度和评价档案制度。自评调控是地方评价领导机构派出督导小组,采取参与、指导和督促、检查等形式,对各学校自评工作进行实地指导和管理。(1)参与、指导评价调查;(2)参与、指导资

料整理与要素量化;(3)参与、指导自评报告的编写;(4)检查、督促、解决自评过程中的疑难问题。评价审计是对评价工作本身进行监督和检查,也称二次评价或后续评价,即对学校评价工作过程、步骤、方法进行审核并作出相应的评价。评价档案是关于学校评价的信息资料库。每次评价工作完毕后,对评价的实施方案、总结、报告、卡片、指标数据等文件材料及时进行归档、选择分类、编号、集中保存,建立评价档案,以资利用。

### 三 我国现代教育评价尚待解决的问题

回顾过去,在总结经验教训中,我们发现,前 20 年的教育评价中存在着下列一些尚待解决的问题。

首先,学校教育评价不应该是外在力量强加给学校一次性管理活动,而应成为学校完整的科学管理过程中的一个重要环节,是学校赖以生存和发展的内在基础。换句话说,评价本应是学校的自觉行为,是学校教育活动求得自我调控、自我发展、自我完善的内在活力。学校教育评价应以形成性自我评价为主,让学校教育的主体构建自己的教育价值形象。但是,过去的学校教育评价却往往都是在教育行政部门或督导部门控制之下的总结性综合评价。这种评价活动的主动权掌握在地方行政官员手里,以某种特定的评价方案为准则,实施对学校教育的检测和评定,学校则完全处于被动应付的地位。由于评价指标不尽科学,缺乏科学的方法和切实有效的评价制度,一方面,评价者有意识或无意识地将教育评价概念泛化,以自己有限的主观经验去评判学校教育的价值,作出关于学校教育成就优劣的结论,并与奖惩挂起钩来;另一方面,学校为了求得最佳评价结论,获得奖励,不惜文过饰非,“报喜不报忧”,从心理上,把教育评价当成外在力量强加的一种无益于学校教育的行政行为,并深为反感。因此,这种教育评价活动不仅不能达到促进科学管理和提

高教育质量的目的 相反会产生负效应 挫伤了一些学校的办学积极性。这就向我们提出了建立学校教育评价应以自我评价为主的运行机制的问题。

其次 教育评价是一种关于教育事物的价值判断过程 它是以人们的教育的价值观为基础的。持不同教育价值观的人,对同一教育事物的评价结构必然有差异。在教育管理活动中,教育管理的主体和客体因各自从事的实践不相同,就必然产生不相同的教育价值观。教育评价应该是主体与客体协同一致的行为 应竭力使两种不相同的教育价值观统一起来。这就是说 在设计指标体系时,应尊重学校的意志,让学校参与整个设计过程,使学校感到设计出的指标体系对学校教育活动具有有效的指导作用。以此 再来看我国过去的指标体系 有相当多的是依据教育行政部门的“长官意志”凭着评价者有限的主观经验 按照泰勒的行为目标模式仿制出来的。这种指标体系曾对我国现代教育评价的起步起到一定的促进作用,但是随着学校教育评价研究和实践的不断深入,愈来愈显示出了其弊端。怎样建立起由教育行政部门(教育督导部门)教育科研部门和学校共同组成的“三位一体”矩阵式的评价指标体系研制组织 成为我们面临的迫切需要解决的问题。

第三 教育现象的复杂性、随机性和模糊性等特点 对教育评价的科学性程度提出了很高要求,教育评价是建立在运用现代科学技术与方法(如数量统计与测量、定性分析、系统分析、模糊数学与电子计算机技术)对原始材料进行科学处理的基础之上的。据我们调查了解,过去的学校教育评价虽然采用了数百种方法与技术 但是 只要稍加分析 就可以看出 许多方法基本上都是传统的经验型方法。在定量分析方面,只涉及到一些粗浅的测量统计知识,对数据处理以及用以描述评价对象价值特

性的分值，也只是采用简易的数学模型推导出来的。同样，在定性分析方面，也只是采用文字描述的方法去刻画评价对象的性状。然而，教育评价要求以客观统计资料与主观描述资料并重，定性分析与定量分析结合的手段，科学地判断评价对象的价值。而过去的教育评价所采用的方法技术，往往远不能达到教育评价活动本身对科学性程度的要求。因此，其评价结果的信度和效度不会很高，评价结论也难以符合评价对象的客观实际。那么，怎样在研究和实践中突破现行的方法技术程序，建立科学的切实有效的评价方法体系呢？这又是一个尚待解决的问题。

第四 在过去的教育评价中 自我评价的机制始终没有建立起来。我们要问，为什么评价非得依靠行政长官的意志，为教育行政部门的中心工作服务，而不能成为学校科学管理的内在驱动力？为什么评价过程中总是掺杂了人为的情感因素和不必要的心理障碍？为什么科学的教育评价活动难以得到绝大多数学校及教育行政部门的认同？我们认为，最重要的原因是我国尚未建立起成龙配套的教育评价制度。我国尚未建立健全的常规教育评价组织机构，更谈不上制定教育评价条例、规程、规则等工作制度。教育评价工作的开展，对于评价的目的与意义、对象与范围、原则与规则、方法与步骤、组织与领导、时间与空间、规程与程序、周期与频率、结果处理办法等一系列问题，都没有制度的制约。因此，从根本上说，这种没有评价制度管理下的教育评价活动与从前的“评比”和“评先进”没有本质上的区别。怎样以教育法学、行政学和政策学为指导，建立适应社会主义中国国情的教育评价制度，这也是我们面临的一个需要解决的迫切问题。

第五 实践表明 我国过去开展的各类教育评价活动基本上都是按照某一事先制定的评价方案进行的。评价方案成了指导和制约评价实践活动的惟一准则。在这里，评价方案的科学