

普通教育 评价与测试文选

北京师范大学教育系资料室编

北京师范大学教育学院印

一九八七年五月

编者的话

教育评价，作为现代教育研究的三大领域（教育基础理论研究、教育发展研究、教育评价研究）之一，正得长足进展。教育评价是教育决策的基础，对教育系统具有重要的反馈调控作用，又是教育管理过程的基本环节。是否具有教育评价的能力，是检验广大教育工作者和教育行政管理者教育科研水平和教育管理水平的一个综合尺度。

教育评价的理论和运动，得到了当今世界各国政府的重视，吸引着社会各界热心教育事业的人士的极大关注。美国是最早出现教育理论和方法的国家。近些年来，美国教育评价研究发展很快，其学术水平在北美和西欧国家中处于领先地位。自本世纪六十年代开始，教育评价活动在世界各地兴盛起来，美国、英国、日本、法国、苏联、加拿大、联邦德国等发达国家纷纷建立了系统的教育评价制度。联合国教科文组织已经将是否具有完备的教育评价技术，作为衡量一个国家教育水平高低的尺度。

近二十年来，各种评价模式纷纷涌现。但是，目前对教育评价的研究仍不深入。教育评价的研究可以说是刚刚起步。随着教育事业的发展，也由于国外教育评价研究的影响，国内越来越多的人认识到教育评价的重要性和现实意义。近年来，教育评价的研究和实验工作也在逐步开展。1984年我国正式加入“国际教育成就评价协会”（IEA），

并参加了“第二次 IEA 科学研究”(SISS)活动。中央教育科学研究所也建立了“中国 IEA 国家中心”。全国各地还先后召开了教育评价理论研讨会,从而推动了我国教育评价工作的深入开展。

赵紫阳同志在《关于第七个五年计划的报告》中指出:

“要加强教育事业的管理,逐步建立系统的教育评价和监督制度”。这是教育事业与教育改革深入发展的客观要求与必然趋势。特别是作为国民教育基础的普通教育的评价,是整个教育评价的重要组成部分,更须倍加重视。此外,教育评价的产生和发展是与教育测量的产生与发展分不开的,教育测量是教育评价的基础和重要组成部分,是教育评价的手段。因此,为了便于教育研究人员与广大教育工作者对普通教育评价进一步研究与探讨,我们从收集的众多专题资料中,择选部分文章汇编成《普通教育评价与测试文选》,以供大家参阅。由于时间仓促,水平有限,不妥之处,请读者指正。

在此,我们谨向收入本书文章的作者及提供资料的各兄弟单位,致以衷心的感谢!

北京师范大学教育系资料室

1987年5月

目 录

(一)

评价的基本理论

略论教育评价发展趋势

.....华东师大教育系 赵永年 (1)

略论教育评价的对象与主体

.....东北师大 刘本固 (9)

布卢姆关于认知领域教育目标分类的基本思想和主要内容

.....骆泽民 左和金 (19)

教育评价与教育测量

.....张厚粲 孟鸿伟 (30)

学校评价

我国学校综合评价的现状与趋势

.....湖北省教科所 游心超等 (40)

中小学办学水平评估指标体系

...上海市教科所《办学水平评估课题小组》 (51)

如何评价你们学校的管理状况

——根据诊断卡片评价学校管理的新尝试

.....国立教育研究所室主任 牧昌见 (59)

评价学校主要看校长

.....郑万钟 (69)

参加中学办学水平评价工作的体会	广东省教科所 吴国铭 (72)
教学评价	
教学评价	吴 杰 (77)
关于课堂教学评价若干问题初探	广州市教育局教研室 麦 曦 (100)
教学效率的量化模型及其理论根据	
——谈教育理论数量化研究的关键	张铁明 (113)
课堂教学质量的综合评判	
.....湖北大学教育管理系 席家煊 (126)	
课堂教学效果的评价方法及对传统课	
堂教学的分析	鹿志保 (134)
教师评价	
教师评价	辽宁省教科所 (148)
中小学教师考核评价的方法	黄云龙 (157)
日本中小学教师的培养规格及其评估	东北师大 薛凤德 (162)
评估班主任工作的指标和标准	[苏]B·R·鲍尔特内 (166)
一个校长对教师的评价	[美]拉塞尔·S·比彻 (175)

评价教师教学工作的背景因素	
.....华东师大比较教育研究所 顾建民	(181)
评价教师工作要避免形式主义	
.....北京教育行政学院编译	(187)
中小学教师考核评价的标准及习见性误差	
.....黄云龙	(192)
美国教师评价方法中的弊病	
.....[美]R·S·索等	(199)
学生评价	
学生评价	
.....辽宁省教科所	(210)
对科学地评价学生的几种探索	
.....张伟俊	(219)
班主任应该掌握教育评价的原则	
.....东北师大 刘本固	(231)
学生质量综合评定初探	
.....张鲁阳 李文庠	(238)
学生品德评定指标体系的主成分、权重研究	
.....德育评价研究小组 胡卫等	(246)
中学生品德考查量化问题初探	
.....湖北大学教育管理系 席家焕	(263)
从能力测定谈打好基础,发展能力	
.....江苏省苏州中学 赵振海等	(275)

(二)

测试的基本理论

介绍一种多指的简易统计方法	
——运用Wilcoxon检验*评价学生体质	
.....上海市育才中学 李伟明	(297)
从历史发展趋向看学业成绩考评方法的现代化	
.....广州师范学院 陈一百	(303)
学生的学业成绩测验	
.....陈延泽 译述	(315)
加涅对评定学习结果的理论与技术	
.....李树珍 编译	(340)
测验分数的双重含义	
.....谢小庆	(349)
测试的类型与编制	
中小学教学测验的功能与种类	
.....上海市教育科学研究所 顾志跃	(355)
中学生学习方法测验的编制	
.....赖昌贵	(361)
测试的方法分析	
正确计算多科考试成绩的方法	
.....广东省教育科学研究所 叶佩华 曾桂兴	(376)
中学生成绩评定法初探	
华中工学院《模糊数学》编辑部 李健萍...	(388)
考题要检验——关键在于考分的常态分布	
.....黄贵中	(397)
班级之间考试成绩的评价方法	
.....张敏强	(410)
同一班级不同学科之间考试成绩的评价方法	

.....	张敏强	(415)
同一班级不同学科之间考试成绩的评价方法		
——同张敏强同志商榷		
.....	华中师大教育系82级 吴淑娟	(418)
判别校际教学质量的一种方法		
...上海南市区教育学院 徐涵鑫 孙国良		(420)
教学效果的综合评价		
.....	辽宁教育学院 郭述平	(424)
班级教学质量管理的探索		
.....	广州执信中学教研室 伍百兆	(431)
利用线性规划方法对学校管理进行		
最优化分析.....	邢最智	(441)
教育问题的相关分析.....	姜凤华	(450)
漫谈标准化考试		
.....	北京师范大学心理系	(453)

略论教育评价发展趋势

华东师范大学教育系

赵永年

本文从教育评价六个基本问题出发，探讨一下教育评价的发展趋势，提出我们对这些问题的看法和思考。

一、教育评价的定义

历史上曾有不少人对教育评价下过定义，但至今认识尚未完全统一。美国密苏理大学的 T·D 坦布林克 (Ten Brink) 教授曾做过一项有趣的实验：他选择一些师范生和任课教师作为被试，要求他们在几秒钟内回答：“当你听到‘评价’这个词时，最先想到什么？”结果依人数的多少，师范生对评价的理解是 1. 测验，2. 等级，3. 成绩，4. 不公正，5. 判断；而任课教师的理解是 1. 测验，2. 测量，3. 等级，4. 成绩责任制，5. 对个人秘密的侵犯。由此可见，即使在教育评价开展得比较先进的美国教育界，人们对教育评价的理解也是很不相同的。

在教育评价发展史上，不少教育评价的专家也曾对教育评价下过许多不同的定义。

瑞夫·泰勒 (Ralph Tyler) 1950 年曾首创评价定义，认为评价就是“确定教育目标在实际上被理解到何种程度的过程”。1963 年柯龙贝切 (Cronbach) 等人提出

“评价是为进行决策提供信息的过程。”豪斯（House）史克列文（Scriven）、斯塔夫列宾（Stufflebean）等人一致认为，评价是一种既有描述，又有判断的活动”“是一种对优缺点和价值的评估。”

目前被广泛接受的评价定义是1981年由美国17名代表着12个与教育评价有关组织的成员组成的有关评价标准联合委员会提出的，他们定义教育评价为“对教育目标以及对它们优缺点和价值的评估的系统调查，为教育决策提供依据的过程。”

我们也倾向于接受这个评价定义，因为这个定义概括了当前评价的三个重要组成部分，即，评价依据——教育目标；评价手段——系统调查，当然包括定性和定量地收集资料的一切方法；评价作用——对教育活动作出价值判断，为教育决策提供依据。

评价定义的发展变化反映了人们对评价对象和评价作用理解的深化。评价已不再如其初期被定义的那样，仅是审定评价对象对一些具体指标的到达度，而是对评价对象的价值（包括评价指标体系在内）作全面而系统的调查。

二、评价的功能

教育评价有什么作用？这个问题的认识有一个发展过程。

早期的教育评价集中于对学生成绩进行考核，常常是用来验收教育质量是否达到了予期的要求，评价主要被用于鉴定、审查、总结学生的学习质量。

1967年Scriven第一个在仔细分析了评价的作用后，提出了“形成性评价”和“终结性评价”的区别。在他之前已有

人认知到这两种评价所起的不同作用。在评价的形成性功能中，评价是被用来改进和发展一个正在进行着的活动，教育的形成性评价是为了帮助改善或不断完善一个教学方案和教学活动而进行的；而终结性评价是用于教育效果的鉴定，核实或人员的选拔。区别评价的这两种功能是重要的，这不仅是由于评价的目的不相同，而且采用的评价方法和收集数据的范围，实施的时机也不相同。这种划分为评价积极干预教育活动增添了活力。

Stufflebean把评价划分为“预先评价”和“追溯评价”。他认为评价的两个功能在于，一是旨在为决策服务的预先评价功能。另一个是为教学效果核实服务的追溯评价功能。我们看来，这种划分和Scriven的划分并没有实质的不同，一个是从形式上加以区别，一个是从时间上加以划分。

有些评价专家认为评价还有第三种功能，即激励和增进认识的心理和社会政治功能。这种评价的作用在于鼓励评价者所希望出现的行为，促进公众关系等。

谈到第四种评价功能的人比较少，他们看到在正规组织中总是由上级评价他的下级，因此认为评价还有执行权威的行政功能。

我们现在不能断言评价的哪一种功能最重要，但不同的功能可以用不同的方法，在不同的时间实施，这就要求我们在评价设计的初期就确定一个具体评价方案的专门功能，从而确定与之相适应的评价途径和方法。有的同志在开展评价时，急于求成，企图通过一次评价就要解决所有的问题，这除了对评价寄予过多的期望外，恐怕还不了解评价有不同的

功能。一个评价方案最好以发挥其一个功能为主，这样评价目的单一，也较容易选择合适的评价方法。

现在的倾向是，越来越重视评价形成性功能，这也许和教育的特点有关。实际教育工作者总是注意教育方案的发展和教育过程的改善，而并不热衷于排名次，争高低。评价的功能从早期对价值的判断到重视改进和发展的形成性功能，看来也是一个历史的趋势。CIPP模式的创始人Stufflebean强调：“评价最重要的目的不是证明，而是改进。”这个趋势反映了教育评价指导思想的转变，这个转变可以概括为一句话，过去的教育评价是要选拔适合教育的儿童，而如今的教育评价是为了创造一个适合于儿童的教育。

三、教育评价的方法

为了反对纯主观的，经验式的定性的评定方法，桑代克(Thorndike)曾提出一个很有名的论断：“凡是存在的，总存在于一定的数量之中。”为了更明白地表示出他们的主张，麦考尔(Macal)又补充道：“存在于数量中的任何东西都是能够测量的。”在这样的思想指导下，一个时期内，在美国，追求对教育现象精确的量化，探讨用数学的定量方法来分析教育现象形成了一个运动。这对于打破传统的偏见，认为定量的数学方法不适宜用于描述教育现象，使教育研究方法向精确化科学化的方向上发展了一大步。但是，随着时间的推移，人们渐渐地认识到纯粹的数学公式解决不了所有的教育问题，在今天的科学发展水平上，对教育现象的全部量化既不可能也无必要，因此一些定性的方法如问卷法、谈话法，专家判别法等等也被常常用于评价的设计

中，然而，这并不是历史的倒退，而是在新的高度上定性和定量的结合。一切被证明的是有效的定量方法非但没有被放弃，相反由于电子计算机的出现而被更广泛地应用着，即使使用的是定性的描述方法，在具体处理时往往也通过一定的量化技术运用数学的方法进行分析。

定性的分析和定量的计算的有机结合，这也是教育评价方法发展的一个趋势，尽管还有人对方方法持极端的态度，但是柯龙贝切说得好：“一个评价者不表明他是忠实于一种定量的方法，还是忠实于一种定性的描述性的方法，这将是明智的。”在教育评价中要兼用定性和定量的方法，这使得我们在选择和宣传教育评价方法时有一个清醒的认识，但是由于我国教育理论界长期来缺乏定量的科学研究，不重视数学方法，把结论建立在一些个体经验的描述上，因此强调一下定量的研究方法，普及一点教育测量学和教育统计学的基本原理，在当前是必要的。

四、自我评价

由于任何评价直接或间接地影响到评价对象的利益或潜在利益，因此现在越来越多的评价者提出要慎重解释评价结论，甚至有人根本反对对被评者下结论。他们提出，一次评价是否需要作出正式报告，是否需要下结论都要根据被评者的愿望而定，而评价者只需把他看到的情况，收集到的数据，各种人员的想法如实地告诉被评者就可以了。

即使主张评价需要有正式报告的人，现在也认为评价的结论在一般情况下应该告诉评价对象，要仔细听取本人的意见，给评价对象发表自己意见的机会。

这种越来越重视评价对象在评价过程中的地位，重视评价对象的自我评价的趋势，反映了教育评价过程中，再也不是把评价对象仅仅看成是一个待测试、待检查的客体，而是把评价对象看成是一个教育评价的积极参与者，把自我评价看成是评价活动一个重要的环节。应该强调评价者和评价对象之间在评价的整个过程中不断地对话，然后互相修改自己的观点，求大同，存小异，使评价结论尽可能地取得一致。

重视评价对象的自我评价，这也许和重视改进的形成性评价有关。教育过程既然是个人的发展过程，那么对教育的评价过程，就当然不仅是评人的过程，也应该包括自评，包含评价对象的自我认识，自我分析，自我提高的过程。

要重视自我评价，就要尊重评价对象。从评价的标准开始，到评价过程中所用的评价工具，一直到评价小结……都应该听取评价对象的看法和意见。这样做一来利于提高收集的评价信息质量；二来有利于评价对象自己发现问题，并改进工作；三来也有利于评价结论为评价对象所接受，防止可能出现的对立情绪。当然，强调自我评价并不是意味着削弱评价者的作用，也不是期望在任何一次评价中评价者总是能和评价对象取得一致意见，而是希望当对评价报告存在不同意见时，这种不同意见有机会发表；当评价者和评价对象之间的不同意见不能通过对话取得一致时，应该分析不一致的原因，并在评价报告中反映出这种分歧。一个评价者有义务也有责任听取评价对象对自我形象的估计，而不要如Cronbach等人所嘲讽的那样：“是一个专门雇来决定‘对’、‘错’的裁判员”。

五、民间学术团体

目前，在西方越来越重视民间学术团体的作用。许多评价活动实际上都是由民间的评价机构来进行的。

比如美国最庞大的鉴定教育情况的组织——全国高等教育鉴定委员会和它下属的近六十个鉴定组织都是民间组织。它们定期对各级各类高等院校进行鉴定，鉴定的结果代表了教育界、学术界的同行对各校的工作和办学水平的评价。美国联邦政府和州政府关心鉴定的结果，但不直接插手鉴定工作。这样就避免了对评价结果负有责任。

当然评价机构是民间的，并不等于说这个机构所作出的评价结论是没有意义和影响的。一个权威性的民间评价机构所作出的结论往往得到全国教育界乃至整个社会的公认。让民间学术团体来承担越来越多的评价活动，有利于把一些行政活动学术化，使评价过程在一个较为宽松的环境中进行，也有利于大家接受评价的结论。

六、对评价的再评价

近年来，一些评价研究人员已经越来越一致地认识到，评价报告写成之后，并不意味着评价活动的结束，而且还需要对评价进行再评价。这也反映了对评价结论的慎重态度。关于再评价的标准有许多不同的提法，主要可以归结为四条标准：1. 实用标准，即保证评价是为实际需要服务，并能结合实际被应用。2. 可行性标准，即保证评价是现实的和慎重的，是能够进行实施的。3. 适用标准，即保证评价是合法地，合乎道德和社会舆论地实施，4. 准确标准，即保证

评价人员具有充分的专业知识，能够正确地传达所需要表达的信息。

对评价的再评价，可以用来衡量一个评价是否有再实施的价值和推广价值，对于改进评价活动，提高评价质量是十分必要的。再评价的活动总是在一个评价实施以后，这一点上给我们的启示就是：我们对评价的探讨和研究不能仅仅停留在理论上，而应该付之实施，只有通过实施后的再评价才能提高我们的评价的水平。

[《上海教育科研》1986年第5期]

略论教育评价的对象与主体

东北师范大学

刘本固

在教育评价中评什么，由谁来评等问题，这是急待解决的问题。如果这个问题解决不好，只从评价方法上做文章，说得通俗一点就是还不知道做什么，就想研究怎么做，这就有些本末倒置。因此，本文拟在探讨教育评价的基本特点的基础上，探讨教育评价的对象与主体，力求促进教育评价理论研究和教育评价实践活动的开展。

一、现代教育评价的特点

（一）教育评价的包容性

众所周知，现代教育评价的范围扩大了。传统教育评价一般只把学生作为评价对象，而现代教育评价则把教育的全部领域作为评价对象。这是因为青少年的成长与发展，是受多种因素制约的，是学校、家庭及社会等多种因素共同作用的结果。从教育评价的具体内容来看，传统教育评价注重考评学生记忆教材的程度，而现代教育评价却着眼于人的全面发展，从德智体等几个方面评价教育对象。

（二）评价目标和教育目标的一致性

教育评价的最终目的是全面实施教育目标。从这个意义上讲，教育目标就是评价目标，评价目标又是教育目标，两者