

教育评估的理论与技术

陈 玉 琨 编 著

广东高等教育出版社

教育评估的理论与技术

陈玉琨 编著

广东高等教育出版社

449.7
411

教育评估的理论与技术

陈玉琨 编著

*

广东高等教育出版社出版

广东省新华书店发行

广州市红旗印刷厂印刷

787×1092毫米32开本 8印张 175千字

1987年2月第1版 1987年2月第1次印刷

印数1—11,000册

书号 7343·32 定价 1.80元

内 容 提 要

2
-
1
本书较全面地论述了教育评估的理论与方法，它探讨了教育评估的本质、类别、范围及其在教育过程中的地位与作用。本书既介绍了西方教育评估的历史与现代流派，又着重探讨了在我国建立教育评估制度应有的特色。对教育评估的一般过程与组织、指标系统设计的原则与方法、权系数集合的构造问题、测量理论和统计方法在教育评估中的应用，本书都作了较详细的阐述。本书对各级各类教育管理部门、教育科学研究人员、师范院校教育系的师生和教育工作者有一定参考价值，也可用作教育管理干部培训班的教材。

CHJ34117

2
-
1

序 言

《中共中央关于教育体制改革的决定》中提出了要对“高等学校的办学水平进行评估”的要求，这无疑对办好高等教育、管理好高等学校是很重要的。因为教育评估可对学校教育工作起到分等督察作用，以保证办学的基本质量；并通过评估得到对学校教育工作的各种信息，可不断调整学校的教育目标和过程；科学地设计评估的指标体系，对办好学校也将起到指挥定向作用。所以，教育评估从理论与实践的结合上开展研究已提到教育管理科学的重要议事日程。本书的作者在校干部进修班开设《教育评估的理论与技术》课程和参加了一定的教育评估实际工作的基础上，对这一问题作了探讨，提出了一些新的观点，做了一件很有意义的工作。

评估是对客体作出价值判断的过程，如何开展教育评估，它与一个国家的社会制度、社会意识形态有着重要的联系，在不同国家有着不同的做法。本书在分析国外有关资料的基础上，立足于建立具有中国特色的教育评估制度，试图以此为出发点，阐述教育评估的技术和方法。作者的这种态度是值得称赞的，并在编写本书的过程中，力求尝试从马克思主义哲学的角度论述和概括一些问题，并不是就方法谈方法，因而使本书的某些方面也就有了一定的深度。

过去，没有全面地、有组织地、系统地开展对高等学校教育工作的评估，因而它是当前高等教育、高等教育管理改革的重要内容。教育评估的理论研究要走在前面，以指导教

育评估的实践，但又必须力戒空谈理论，评估的理论研究要调查、分析和解决我国今天教育中评估的实际问题。作者在这方面进行了可贵的努力，如在论述“评估指标系统设计的途径和方法”时，就联系了我国高等学校重点学科评估的具体问题。

我国对教育评估的研究起步较晚，本书是作者在这方面初步研究的一些成果。要使教育评估在理论上不断成熟起来，在实践中逐步解决教育评估中的种种问题，并在此基础上形成具有中国特色的教育评估学是完全有可能的，但需要靠大家的努力才能实现这个目标。

薛天祥

1986. 3. 9.

责任编辑：陈 辉

封面设计：黄培存

书 号：7343·32
定 价：1.80元

目 录

序 言

第一章 概论

- 第一节 教育评估的概念 1
- 第二节 教育评估的主要类型与各类评估的特点 ... 6
- 第三节 教育评估在教育管理中的地位与作用 11

第二章 西方教育评估的实践、理论与中国教育评估

制度的特色问题

- 第一节 西方教育评估的历史发展与现代流派 16
- 第二节 中国教育评估制度的特色问题 34

第三章 评估的一般步骤

- 第一节 评估的准备 42
- 第二节 评估的实施 46
- 第三节 评估结果的处理 54

第四章 评估的组织

- 第一节 评估的组织机构 58
- 第二节 评估中各类人员的作用与组织方式 62

第五章 指标系统

- 第一节 指标的概念与指标和目标的关系 72
- 第二节 指标系统的设计原则及其哲学、逻辑学的
诠释 75
- 第三节 指标系统设计的途径和方法 85

第六章 权集合

- 第一节 权集的本质及其解释 97

第二节	构造权集的一般技术·····	112
第三节	检验权集品质的一般方法·····	120
第七章	分项评分技术(上): 测量技术	
第一节	分项评分问题概述·····	123
第二节	测量的一般原理·····	126
第三节	测量法与改善测量品质的途径·····	137
第八章	分项评分技术(下): 统计技术	
第一节	统计概述·····	147
第二节	统计量与分值的确定·····	155
第三节	假设检验在成果指标定性分析中的应用··	161
第四节	齐次马尔可夫链分析法在教学成果指标定 量分析中的应用·····	179
第九章	评估结果的分析与处理	
第一节	分项评分数据的处理与综合评估结论的 获得·····	192
第二节	评估信息的反馈与数据处理·····	214
第三节	评估的实践与提高评估质量的一般途径··	220
附表	·····	230
主要参考文献	·····	243
后 记	·····	247

第一章 概 论

随着我国教育体制改革的深入，教育评估的问题已经引起了教育工作者的广泛注意。《中共中央关于教育体制改革的决定》有两处提到了教育评估的问题，在谈到加强对高等学校宏观管理时，《决定》提出：“教育管理部门还要组织教育界、知识界和用人部门定期对高等学校的办学水平进行评估，对成绩卓著的学校给予荣誉和物质上的重点支持，办得不好的学校要整顿以至停办。”在谈到培养高质量的人才和出高水平的科研成果时，《决定》又提出：要“根据同行评议、择优扶植的原则，有计划地建设一批重点学科。”这就把教育评估的问题提到了重要的议事日程上来了。

评估是一价值判断的过程，是完整的科学管理过程的一个重要环节。在高等学校办学自主权有了较大幅度的扩大后，积极而又慎重地开展教育评估，是促进各级各类高校不断提高办学水平，使它们更好地为社会主义建设服务的重要措施和保证。它对促进高校的建设和发展具有重要的作用。但是，由于全面地、有组织地、系统地开展教育评估在我国还是一项新的工作，对此，我们还缺乏必要的经验，因此，有必要对它的理论与技术进行深入的研究。

第一节 教育评估的概念

一 评估与评价

评估，一般被认为是一与评价相近的概念。在实践中，

教育评估与教育评价常常被人们当作同一概念，在同一意义上加以使用。因此，深入地研究评估需要首先对评价这一概念略作辨析。

评价 (evaluate) 即评判价值，就其内涵来说，它与价值这一概念有关。在英语中，评价一词由词干“valu”加上词头“e”和动词性词尾“ate”组成。其中“valu”意为价值，词头“e”的意义等同于“out”，意为引出。评价即为引出和阐发价值。从本质上说，评价是一价值判断的过程。

教育评价是对教育的社会价值作出判断的过程。教育作为一种客观的社会活动，它的结果满足一定社会的政治、经济、文化等发展的需要，其现实性和可能性构成了教育的政治价值、经济价值和文化价值，这些价值的总和构成了教育的社会价值。教育评价就是对这一社会价值作出判断，从而推动教育活动的发展。

那么，什么是评估，它同评价究竟有什么区别？谈到评估时我国有的学者认为，“‘评’是评价、评判、评定之意，即弄清对象的价值高低、质地优劣。本来，既然是评定，也就是可以断定了，但是，这种断定又只是属于揣度、推测、估价、估量，只能说‘估计是如此’，还不能说‘绝对必然地就是如此’。这就是‘评估’的含义，即评价与估量。”^①这就是说，评估是一种模糊定量的评价。这一观点是言之成理的。的确，由于教育评估的对象主要是教育机构、教育团体或者是教育计划、方案、政策等等，涉及的因素较多，复杂性程度也较高，因此，对它们的评价按严格、精确的定量分析方法进行还有困难，目前，主要地还是依靠定性与定

① 张尚仁：《认识论与决策科学》，云南人民出版社1985年版，第70页。

量结合、客观统计资料与主观描述资料并重的手段。评估这个词反映了这一事实。在教育界，尤其在西方，人们倾向于把以客观的方法（比如测验的方法）对学生个体的评价称为评定（*assessing*），而把对学生个体以外其它复杂性程度更高的评价称为评估。

当然，无论是在国外还是在国内、评价、评估、评定这些概念的区分本身也是很模糊的、极不严格的。在一般文献中，这些词混用的情况常常可见。甚至在不少文献中，除了“学生评定”这样的概念外，只有所谓“机构评价”、“方案评价”、“需要评价”以及“工作人员评价”这样的词，很少出现“评估”这一概念。

二 评估与价值

评估作为价值判断的过程，它的特点与价值的特点有关，因此，这里我们有必要再对价值这一概念略作探讨。价值在哲学界至今仍是一歧义异见，众说纷纭的概念。在西方哲学中，有截然对立的两种观点：一方为客观主义。它认为，价值是客观对象固有的本性，是独立于人类的认识、情感和行为之外的；另一方是相对主义。它不同意客观主义的这种价值论，不同意关于价值是客体固有所本性的观点，它认为价值乃是用来表达个人对事物与客体的感情的。现代存在主义哲学则进一步主张，价值乃是个人所创造的。这样，在他们看来，价值是完全主观的、相对的东西。事实上，这两种观点都是错误的。按照马克思主义的观点来看，价值是一用以表明客体对主体特殊效用关系的概念。“‘价值’这个普遍的概念是从人们对满足他们需要的外界物的关系中产生的。”^①价值是主体人

^① 《马克思恩格斯全集》第19卷，第406页。

的需要同外部世界的一种关系，价值表示在实践的认识活动中，客体的存在、属性和合乎规律的运动变化结果向主体接近的现实性和可能性。价值是一以主体的一定需要、意图、愿望，一定的目的、目标、指向为准绳来衡量客体效用的。因此，价值是主体性与客观性的统一。价值的这一特点决定了作为价值判断过程的评估区别于其它活动的一个重要特征：它是主体性与客观性高度统一的活动。这是评估的基本特点。

诚然，评估作为一种价值判断的活动，它又不同于其它形式的价值判断的活动。1981年，由代表着美国十二个与教育评估有关组织的十七名成员在对教育评估下定义时，指出评估乃是“对教育目标及它们价值的判断的系统调查，是为教育决策提供依据的过程。”^①这就指出了评估是通过系统地收集信息，按照严格的科学程序，有计划，有组织进行的活动。因此，它是关于对象的一种较为深刻的，对于它的发展变化具有重要影响的价值判断过程。在现代的学校管理中，教育评估起着一种特殊的信息反馈机制的作用，它通过对现状与目标之间距离的判断，有效地促使被评对象不断逼近预定的目标，不断提高工作质量。由于这个原因，在现代管理中，评估已越来越多地受到了人们的重视，它已成了现代科学管理的一个重要环节。因而，它也成了管理科学研究中一个日益受到人们关注的领域。

三 评估与测量

价值判断一般地说是以事实判断为基础的。事实判断是

① Nevo, D. "The Conceptualization of Educational Evaluation: An Analytical Review of the Literature", Review of Educational Research (Spring 1983)。

一与价值判断既有联系又有区别的概念，它是对事物的存在、属性、运动变化的规律，人的活动及其结果的检验。在英语中，一般用系词“to be”来表达。它的结果是作出肯定或否定的句子，关于事实的陈述是它的表现形式。而价值判断则是根据事物能否实现人的目的来决定，它是在对事实检验的基础上，对客体——事物和人的行为以及结果可能的或者事实上已经满足主体——社会和个人一定需要的程度作出判断。一般地它包含着判断者对“ought to”（应该如何）的认识。因此，它们是有区别的，但它们又是有紧密联系的。

教育评估同教育测量有着紧密的联系。它们之间的关系就是价值判断与事实判断的关系。测量是依据一定的标准对事物的某一属性作出事实判断，进行赋值的过程。测量是评估的基础，但是测量并不就是评估。

首先，与评估相比，我们可以说测量是一纯客观的过程。客观性是测量质量的首要指标。客观的测量期望测量者能尽可能地排除各种主观因素的影响。不同的测量者，对同一事物的测量，除了允许的误差外，它的结果应是相同的。评估是在测量的基础上进行的，但它却超出了描述，而试图去确定行为的价值。对同一事物，由于评估者的价值观不同，评估的结果有时可能大相径庭。前面已经指出，在本性上，教育评估并非是纯客观的，而同时具有客观性和主体性这两个基本属性。它的客观性就是它是在对学校教育工作状况和结果的客观测量、客观的检验、客观的描述的基础上进行的，它的主体性就是它又是以主体的需要和目标为准绳来评判的。

其次，与评估相比，测量是一相对地说较为单一，即只

重视量的获得的活动，而评估则是一较为复杂，包含着多重活动的过程。评估，包括定性与定量的评估在内，都是对客体质的差异作出判断的过程。所谓定量的评估只不过是在认识客体间数量差异的基础上，对它们质的差异作出它更精细的区分而已。

评估与测量是有紧密联系的，任何评估都把测量作为基础，是测量的深化与发展。从教育评估的历史演变看，教育评估也是在教育测量的基础上发展起来的。历史的分析与逻辑的分析结论是一致的。但是，如上所述，它们之间是有着重要的区别的。

测量应是一客观的事实判断过程。但在实际工作中，测量者的价值观念往往也会对测量产生重要的影响。哥德曾经说过：“我们见到的只是我们知道的。”^① 这就是说对同一对象，不同的观察者有时也会得出不同的观察结果，除了难免的误差外，观察者观念的不同是一个重要的因素，它对观察结果的获得产生着重要的影响。人们对实验工作者有这样一个要求：一走进实验室，他就应是客观事实的忠实观察者，不要把自己所想见到加到实际所见到的事实中去。我们也应对教育测量者提出类似的要求。

第二节 教育评估的主要类型与 各类评估的特点

教育评估除了根据其对象可分为机构评估、方案评估与工作人员评估等类型外，还有终结性评估与形成性评估，伪

^① [英]贝弗里奇著，陈捷译：《科学发现的艺术》，科学出版社1979年版第103页。

评估、准评估与真评估以及硬评估与软评估等多种分类。下面我们对各类评估的特点作简要的分析。

一 总结性评估与形成性评估

评估就其主要目的与用途的不同，可分为总结性评估与形成性评估两大类型。总结性评估以对一个阶段的工作作出总结为目标，它的主要用途是对被评对象作出鉴定、区分等级、预言被评对象未来发展的可能性。总结性评估又称最终产物的评估，一般地它特别关心教育活动的结果，而基本上不涉及教育活动的过程，是事后的评估。学生的毕业考试和高校鉴定是总结性评估的很好的例子。由总结性评估的信息主要是提供给各类决策人。比如，通过高校鉴定，决策人将决定是否对被评院校给予进一步资助。

形成性评估与总结性评估不同，它特别重视评估的问题诊断，计划调节功能，而不太重视评估的分等鉴定功能。形成性评估以对决策计划、方案、活动提出改进意见为目标，它的主要用途是对获得改进工作的依据，以期提高被评估的决策、活动本身的质量。形成性评估是过程中的评估。“事前测验”是形成性评估的典型例子。如在教学计划或教材编好后。为了给教学计划的修订或教材的修改提供必要的信息，评估者从将要使用它们的学生中抽取小样本，让他们事先试用，然后通过测验等方法从中得到有关信息。这种评估就是形成性评估。

总结性评估、形成性评估这两个概念是由美国学者斯克裡芬 (M. Scriven) 在其1967年所著《评价方法论》中首先提出的，最初斯克裡芬只把他的这两个概念运用于课程评估中。后来，人们逐步地把它们的应用领域拓宽，使它们适用

于对一切评估活动的分类。

这里需要指出的是，总结性评估与形成性评估是一对相对的概念，而不是相对立的概念。美国《教育评价大百科》的编者安德生等人曾经正确地指出：“一个好的总结性评价可能还具有形成性评价的含义。”^①事实上，就象“过程中”与“过程末”是相对的一样，总结性与形成性也是相对的。一个小过程的结束可以看作是大过程中一个阶段。因此，只要我们方法得当，一次特定的评估是可以同时发挥其总结性的与形成性的综合功能的。

二 伪评估、准评估与真评估

教育评估活动根据其与评估定义的一致性程度又可分为伪评估、准评估与真评估。著名的教育评估专家斯塔弗尔比姆(D. L. Stufflebeam)等人曾经对评估下过定义，他们认为，评估研究是为了帮助一些人对一些教育对象的价值作出判断和加以提高。根据这一定义，他们考察了各种以评估名义进行的活动，认定，实际进行的许多活动是同这一定义相矛盾的，因此，根据活动与定义相吻合的程度他们把评估分成三类：

1. 伪评估。伪评估是一种政治倾向型的活动。进行这种评估的目的主要是为了实现政治控制，或者是为某个方案、某个过程或某个机构制造出一个能引起人们好感的假象，而不在于作出真实的价值判断。一般地，在这种评估中，控制这一评估的主持人总要故意地制造一些假象，隐瞒一点对

^① Anderson, S.B. et al., "Encyclopedia of Educational Evaluation" (1975).