

教育评估理论 与实践

北京航空學院出版社

教育评估的理论与实践

北京市高教局	北京航空学院
北京师范大学	北京钢铁学院
北京工业学院	北方工业大学

合 编

北京航空学院出版社

内 容 简 介

本书是从《评估高等学校工作状态理论与方法学术讨论会》的论文中，挑选、整理、编纂而成的。

它的主要内容有：教育评估的理论基础、评估职能、指标体系及指标相关性的分析；研究德育评估、体育评估、课程评估的实践、学生技能评估以及国外教育评估的比较。

它对于各类学校、教育管理部门及教育科学研究机构是有参考价值的。

2592/13

教育评估的理论与实践

北京市高教局及北京航空学院等院校

教育科研机构合编

责任编辑 许传安

* * *

北京航空学院出版社出版

新华书店北京发行所发行 各地新华书店经售

北京航空学院印刷厂印装

* * *

787×1092毫米 1/32 印张：7 字数：154千字

1987年5月第一版 1987年5月第一次印刷 印数：1~6500册

统一书号：17432·056

ISBN 7-81012-000-X/Z·001 定价：1.60元

前 言

《中共中央关于教育体制改革的决定》发表以来，我国教育评估的理论研究和实践都取得了较大的进展。

如果以教育评估最初萌芽的古代考试制度算起，我国的教育评估已有几千年的历史。但是教育评估成为一个独立的研究领域，并逐步向科学化方向发展则是本世纪以来的事。对于我国，则只有几年的时间。教育评估的实践也说明我们的评估理论尚未成熟，实践经验亦尚不充分。为此，为了推动教育评估理论的研究和促进教育评估的实际工作，北京市高等教育学会和北京航空学院在国家教委的支持下，于一九八六年六月十二日～十四日在北京召开了“评估高等学校工作状态的理论与方法学术讨论会”。这次会议是成功的。如果说“高等工程教育评估问题专题讨论会”（“甬泊湖会议”）标志着我国现代的、科学的教育评估工作的发动与起步的话，那么，这次会议表明了我们的教育评估理论与实践，在深度上有延伸，在广度上有扩展。

为了进一步推动教育评估的理论研究和实践，北京航空学院、北京钢铁学院、北京工业学院、北京师范大学、北方工业大学和北京市高等教育局，在这次学术讨论会研究成果的基础上，编辑成《教育评估的理论与实践》一书，以饗读者。该书涉及到教育评估的理论及依据、教育评估的职能与作用、教育评估的实践与研究、国外教育评估的理论与实践等问题。

该书似书，又不是一本完整、系统的书。它是各院校著译的论文，经过一定的整理，按着一定的结构有序地排列起来的。前后的观点、题问的提法及名词术语难免有不一致的地方。内容也难免有重复。尽管如此，它对于各类学校、教育管理部门及教育研究机构还是会有参考价值的。它也许对我国正在起步的教育评估活动会起到一点推动作用。倘能如此，我们也就满足了。

我们相信，经过若干年的理论研究和实践探索，一定能在不断总结自己经验和借鉴国外有益的经验基础上，建立起符合我国国情的教育评估体系与制度。

由于整理、编纂水平有限，难免有失误之处，望本书各章、节著译者予以指正。同时也代表各位著译者欢迎广大读者批评指正。

编 者

一九八六年十二月

目 录

第一章 教育评估理论研究概述	(1)
第二章 教育评估的基础理论	(8)
第一节 教育评估中的若干问题.....	(8)
第二节 系统理论在教育评估中的运用.....	(21)
第三节 教育评估要着眼于总体目标.....	(28)
第三章 评估的职能与作用	(33)
第一节 教育评估职能的多样化.....	(33)
第二节 教育评估的作用.....	(48)
第三节 教育的社会效益评估.....	(55)
第四章 评估指标体系的研究	(63)
第一节 简化评估指标体系的探讨.....	(63)
第二节 评估数据的采集.....	(72)
第三节 指标体系的相关性.....	(81)
第五章 评估的实践与研究	(98)
第一节 校内专业评估的实践与研究.....	(98)
第二节 课程评估的初步实践.....	(102)
第三节 学生思想政治工作评估.....	(106)
第四节 教学计划的质量评估法.....	(112)

第五节	高等学校体育工作综合评估法·····	(123)
第六节	学生实践技能的评估·····	(132)
第七节	评估偏差心里分析·····	(138)
第六章	苏联的教育评估·····	(144)
第一节	苏联教育评估简介·····	(144)
第二节	苏联高等院校主要课程教学 质量检查指南·····	(152)
第七章	美国的教育评估·····	(166)
第一节	美国教育质量评估活动简介·····	(168)
第二节	美国高等学校的教育评价·····	(197)
第三节	美国教育评估理论与实施方案·····	(205)

第一章 教育评估理论研究概述

《中共中央关于教育体制改革的决定》中提出：“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。”开展教育评估工作是教育体制改革后的必然要求，也是教育体制改革的一个重要方面。国家在扩大了学校办学的自主权以后，必须加强和改进对教育的宏观指导和管理。改变过去管理部门在宏观上没有管好，在微观上又管得过死的弊病。放权之后，管理部门还要掌握学校人才培养的数量、质量及其经济效益，要了解毕业生情况的信息反馈。因此，对各级各类学校的办学水平有必要进行评估。一九八三年在武汉召开的有关会议提出要对重点学校进行评议以后，有的学校在校内组织了教学质量、学术水平、学科发展方向等单项指标的自评，有的学校着手研究适合我国国情的评估制度，研究评估高等学校工作状态的理论和方法。一九八五年三月，原教育部召集了部属高等学校座谈会，讨论如何建立评估方案；四月，又在北京航空学院召开了十五所院校评估工作座谈会；六月二十日至三十日，原教育部在黑龙江省镜泊湖召开了三十八所高等院校（包括文、理、工、农、医、师范等科）和七个部委、四个省（市）高教厅（局）同志参加的“高等工程教育评估问题专题讨论会”。会后还提出了一个高等工程教育办学水平评估试点参考方案。一九八六年六月十二月至十

四日由北京市高等教育学会和北京航空学院联合发起举办了《评估高等学校工作状态理论与方法学术讨论会》，参加会议的有八十人，四十一个单位，其中有综合性大学、师范大学、医科大学、工科高等学校等三十所，六个部委、三个省（市）高教厅（局）。中央教科所、中国科学技术出版社等代表也参加了会议。会议共收到论文六十篇。这次会议是继鏡泊湖会议后又一次大型学术讨论会。与会同志认为，这次会议在鏡泊湖会议的基础上，使教育评估理论与方法的研究又前进了一步。本书就是从这次会议论文中精选、编辑而成的。在此将这次会议研究的几个评估理论问题的情况分述如下。

一、教育评估的内涵问题

什么是教育评估？说法并不完全一致：①评估主要是教师和管理干部在一定条件下，经过主观努力达到目标的程度。②评估的对象不是客体的实体属性，即不是客体本身的本质和规律，而是客体的价值属性。和这种提法相似的，更多的人引用了美国教育评估代表人物贝比Beeby在一九七五年对评估的定义：“系统性地收集和解释证据的过程，在此基础上作出价值判断，目的在于行动。”所谓价值判断，有的同志解释——例如评估某一门课程，在校际之间进行统一测验，以检查学生掌握知识和能力的情况。假定两所学校统测的平均分数都是八十分，如果有一所学校是重点大学，新生录取分数高，另一所是一般大学，录取分数相对的低一些，虽然两所统测成绩是一样的，但后者的价值要比前者高，即说明了办学水平高。因此，有同志提出，评估是根据一定目标，通过系统收集信息，对教育客体作出价值判断的过程。其目

的在于推动人们去完成预定目标。有同志把教育评估的内涵分为四个方面：即系统地收集教育的信息；对收到的信息进行解释分析；对教育的文化、政治、经济价值做出判断；促进教育的改革和进步，不断提高教育水平。

二、教育评估的基础理论问题

什么是教育评估的基础理论，在什么理论指导下建立适合我国特点的评估制度，是评估的理论研究中带根本性的问题。有同志认为，高等教育评估对象是大小不同的系统。对这些对象的评估，就是对系统的评估。由于对象是一个系统，因此理所当然的要以系统理论为指导，应用系统理论中整体性原理、反馈性原理和有序性原理。与此不同的看法是：教育评估的理论基础是教育学、管理学，而“三论”是研究自然、社会现象的方法论，不是教育评估的基础理论，只不过是借用来研究评估理论的方法。要评估的是教育，就要了解教育的内部律规和外部规律；评估又是管理中的一环，要研究评估的目的和方法，需要有管理学作指导，同时还要运用教育心理学、管理心理学、教育测量学、教育统计学等诸方面的研究成果。有同志提出教育评估的基本理论应该包括三个层次的内容：①定义、内涵、本质、目的、作用以及在教育中的地位；教育与政策的关系、评估的价值观念等等；②制定评估指标体系的基本理论；③教育测量与教育评估方法的理论。

三、教育评估的职能问题

有人提出，教育评估可以区分为不同的目标。为了达到不同的目标而制定的指标体系就显示了教育评估的不同职

能。有人认为，《决定》规定了教育评估的“上”“下”限，因此制定专业和学校评估指标时，应把评估分为“鉴定”和“评价”两大类。所谓鉴定就是资格审查。它只涉及到办学基本条件和标准。指标条目只有“够格”与“不够格”，“通过”和“不通过”之分。“不够格”，“不通过”者不得进入“评估”。评估就是对办学水平的评估。对各指标条目应分设卓越、良好、及格、不及格四档。所有学校和专业都应定期地被“鉴定”，而被评估的单位，则只是一部分。评估的目的是鉴别办学成绩卓著的学校（专业）。

美国的教育评估，突出的是鉴定和评选的职能；苏联的教育评估侧重了评比和估价。有人在研究了美国和苏联的教育评估制度后，提出了建立中国评估制度的设想。他设想我国评估制度应具有四种职能，其具体内容是：

(1) **鉴定合格的职能**：区分合格与不合格。所有学校都要通过合格线。同时，还应制订开办新的高等学校的条例，防止滥设新校和盲目升格。

(2) **评比职能**：评比是一种社会主义竞赛，它使已经通过鉴定的学校不断地获得前进的动力。

(3) **评选先进的职能**：其目的就是要形成或扶植一批具有先进水平的高等学校。评选先进职能又有条件评选，水平评选，效益评选等几个方面。

(4) **估价成就的职能**：这是高教评估职能中难度最大的一种职能。它可分为验收性估价，发展性估价和推广性估价。

四、简化指标体系的探讨

一九八五年十二月国家教委发出开展高等教育评估研究

和试点的通知，教育评估也进入了一个新的阶段。这个阶段的一个重要特点就是许多理论问题的研究都是根据实际工作中出现的问题提出来的。如在评估的初期阶段，提出的评估指标体系方案中，有的末级指标达到200~300项，少的也有几十项。指标数量太多，权重就要分散。虽然可以增加主要指标的权重，但实际上增加也是有限的。我国有上千所高校，每所都要用成百个的指标去评估，工作量很大。因此如何简化指标体系已成为当务之急。有的人认为要从明确评估的职能入手。只有明确评估的职能，才能建立起简明扼要的指标体系。还有的提出把全国高校分为不合格的、合格的和高水平的三种类型，先用简化指标筛选出两头，中间的多数就好评定了。第三种办法是对评估指标体系作相关性分析。这种方法主要是运用《概率与数理统计》中主成分分析法。一个指标体系中，各个指标常常有一定程度的相关，应把这些指标转换为互相独立的指标。新的指标是原来指标的线性组合。经过这种转换，指标数量会有较大的减少。在讨论中，许多人认为运用主成分分析法简化指标体系，可能是今后简化指标体系的重要方法。不过由于评估指标分类归并简化之后，指标的内涵就模糊了，实际评估中运用就会发生困难。精确分析指标间的相关函数，工作量也是很大的，但并不是说这种方法就不可取。

五、关于评估的方法问题

在讨论中，与会者认为以下几点应引起注意：①评估的方法要注意简易、实用和方便。特别在微观评估中，参加的人数很多，评估方法要为一般人所理解和掌握。②评估涉及到各种因素，不宜在评估中不恰当地使用数学方法，把社会

系统当成简单的工程技术系统来处理。有的人提出用层次分析法。它虽然是一种较好的方法，但计算工作量大，不易被基层工作人员所接受；对模糊矩阵法也有一些不同意见，这种方法虽然比较简单，有人认为计算结果常常与实际情况有较大出入。一般地说，加权求和法比较简单实用，又容易被理解，可在更多的场合下使用。③评估中涉及到一系列主观因素。指标体系和权重的确定、对主观因素的定量描述，不可避免的存在误差。计算方法的精度高，不但不能消除这些误差，有时反而加大这些误差。因此，有人提出，计算方法只要有一定的科学性就可以采用，而不必追求精确性高。④收集数据涉及到人的因素时要防止挫伤群众的积极性。例如，在组织学生参加课程评估时，要防止把评估的矛头对准教师，增加对教师的压力。⑤评估应以自我评估为主，国家和社会评估应建立在自我评估的基础上。但有人持相反的意见，认为国家评估的着眼点是社会评估，目的是要建立对高等教育评价和监督制。也有人认为：来自社会用人部门的反馈信息无疑在评估中具有重要作用，但他们不是评估的主导力量；教育评估要注意教育过程和工作状态，它们属于教育科学研究的范畴，社会部门不甚熟悉。因此对教育过程的评估，要更多地依靠教育人士，教育的业务部门。他们是教育评估的主力军。

六、关于教育评估的比较研究

这次学术讨论会共收到研究或介绍苏联、美国及联邦德国教育评估论文十五篇。与会者认为，研究外国的教育评估制度，对建立适合我国国情的教育评估制度是有益的。研究过程中，要充分识认中国与西方教育评估制度的区别，否

则就难免走西方走过的弯路。有人认为，建立中国教育评估制度必须注意以下几个问题：①坚持以社会需要为依据，以教育目标为准绳，促进学校主动适应社会需要的能力。西方出现的一些评估流派，反对教育评估中价值观念的收敛性，要求创造出各种新方法适应价值观念发散性，即价值观念高度不统一性的特点。对此必须仔细加以鉴别。②在组织上要官民结合，坚持学术上的权威性和政策上的合理性；③在具体实施中，要注意科学性与可行性相结合，坚持辩证唯物主义的方法；④广泛吸收各国长处，分析西方现代教育评价的特征，引以为鉴。

人们的认识来源于社会实践。任何理论都是社会实践经验的概括和总结。我国教育评估还处于试评阶段。评估理论研究还只是开始。随着评估试点工作的开展，定会有许多新的成功经验需要上升到理论。也会有许多问题需要借鉴其它学科的理论进行研究和解决。本书只能反映这个阶段的研究成果，随着评估工作开展，定会出现成熟的评估理论和办法。这也就是下一阶段的任务了。

（北京航空学院 傅雄烈）

第二章 教育评估的基础理论

第一节 教育评估中的若干问题

一、教育评估发展的历史概况

教育评估就其历史发展来说，是从对学生的学力检测开始的。在我国，自隋炀帝置进士科始(公元606年)，延续了一千三百年之久的封建科举制度，就是一种关于学生学力检测、评估的制度。然而，教育评估成为一个独立的研究领域，逐步向科学化方向发展则是本世纪以来的事。“评估”成为一个科学的概念则更晚。

在十九世纪上半叶或更早的一段时期，在学校中广泛流行的考试缺乏标准，因而，也就缺乏考试应有的客观性。虽然1702年在英国剑桥，1845年在美国波士顿先后以书面考试取代了以往的口试，但是，考试的客观性问题仍然是一令人头痛的问题。针对这种情况，1864年英国格林威治医学校(Greenwith Hospital School)教师菲奢(G. Fisher)收集了许多学生的成绩，汇编成了成绩量表，试图为当时的考试提供一个可供参考的标准。但是，由于种种原因，菲奢的这一工作没有引起教育界的重视，因而未能产生多大的影响。

在教育界引起人们对测验问题极大关注的是著名的莱斯(J. M. Rice)拼字测验。1897年，莱斯发表了他对20所学校，16000名学生所作的拼字测验的结果。莱斯测验表明：

八年中每天花45分钟时间进行拼字练习的同每天花15分钟进行练习的成绩并没有多大的差别。莱斯的这一结论震惊了教育界，也遭到了不少人的反对，但却引起了人们对测验问题的普遍关心，推动了教育测验、测量问题的研究。

与此同时，心理学、心理测量的发展，对教育测量起了很大的推动作用。在十九世纪与二十世纪交替的这些年代中，人们的研究精力主要集中在教育评估中的一个基本问题——教育测量的客观化问题上。为此，教育评估学家把这一阶段称为“测量运动”阶段。所谓客观式测验（Objective test）和常模参照测验（Normal referenced test）就是这一时期的产物。它们在解决学力测验的标准化和客观化方面取得了很大的进展。

自三十年代初始，教育评估的发展进入了一个新的时期。国外有人把这一阶段称为“评估运动”阶段。1929年西方经历了一场空前的经济危机。危机的结果是经济萧条、工厂倒闭，一大批青年因无工作可做，只能涌向中学。他们原无上大学的要求，也无上大学的兴趣。但是，当时美国的中学是处于大学入学考试指挥棒的作用下的。中学的课程为大学入学服务，不适应失业青年的需要。这就使得学生与学校课程之间发生了尖锐的矛盾。在这种情况下，泰勒（R.W. Tyler）受命进行了一场为时八年的课程与评价的研究。研究的结果使他们得出这样的结论：课程与评估是教科书中心主义。测验的内容只是记诵教材的知识内容，是极为片面的，并不能真正反映学生全面的发展要求。为此他们提出了一套以教育目标为核心和依据的课程编制与测验编制的原则，试图以此把社会的要求、学生的需要反映在课程与评价之中。当时泰勒的工作取得了很大的成功。所以这一阶段也常被人

们称作教育评估史的“泰勒时期”。根据泰勒的理解，教育评估就是衡量实际活动达到教育目标的程度。由于根据这一理论进行的评估活动能够获得关于目标完成情况的信息，有助于发现存在的问题，改进教学工作，因此受到了较为广泛的欢迎。

1957年苏联人造卫星的上天，在美国朝野引起了极大的反响。他们经过深刻的反省，一致认为，美国在科技方面的落后，归根到底反映了教育上的落后。为此，他们花了极大的精力着手进行教育改革。在教学领域中，布鲁纳的“结构课程论”就是这一时期的产物。在教育评估领域，也有人重新审视了当时所流行的评估理论。他们对评估理论再审查的结果认为，当时的评估理论已经“重病缠身”，如评估以目标为中心和依据，那么，目标的合理性又根据什么来判断；教育活动除了达到预期的目标外，还会产生各种非预期的效应与效果，这些非预期效果要不要评估。如此等等。经过大约十年时间的徘徊，1966年后，美国评估流派大量地涌现出来了。教育评估到了一个迅速发展的时期。为了对这一时期的特点有一概要的了解，我们需要对这些流派择其要者作些简要的探讨。

二、西方现代教育评价主要流派

（一）行为目标模式 行为目标模式的代表是泰勒的理论。行为目标模式与心理学中行为学派有重要联系。它把教育方案、计划的目标用学生的特殊成就来表示，并把这一行为目标当作教育过程和教育评估的主要依据。根据这一模式，教育评估就是判断实际活动达到目标的程度。预定的目标决定了教育活动，同时也规定了评估就是找出实际活动偏