

教育部本科教学工作水平评估优秀示范读物



比较与借鉴—— 中外高等教育评估体系研究

● 张德才 陈虹岩 编著



哈尔滨工程大学出版社

教育部本科教学工作水平评估优秀示范读物

比较与借鉴—— 中外高等教育评估体系研究

● 张德才 陈虹岩 编著

哈尔滨工程大学出版社

内 容 简 介

相对于西方高等教育发达的国家,我国的高等教育评估工作起步较晚,因此,借鉴一些西方国家,尤其是高等教育相对发达、高等教育评估体系相对完善国家的一些有益的经验 and 做法,就显得十分重要。本书选择了几个具有代表性的国家,阐述了他们关于高等教育评估的基本做法,分析了不同评估体系对促进本国高等教育质量的作用,着重论证了各个国家与我国在高等教育评估方面的差异,从而能够为完善我国高等教育评估体系、促进高等教育评估规范化、制度化和科学化提供一些有益的借鉴。

图书在版编目(CIP)数据

比较与借鉴:中外高等教育评估体系研究/张德才,
陈虹岩编著. —哈尔滨:哈尔滨工程大学出版社, 2008. 5
ISBN 978-7-81133-331-2

I. 比… II. ①张…②陈… III. 高等教育-教育评估-
对比研究-中国-外国 IV. G649. 2

中国版本图书馆 CIP 数据核字 (2008) 第 049603 号

出版发行 哈尔滨工程大学出版社
社 址 哈尔滨市南岗区东大直街 124 号
邮政编码 150001
发行电话 0451-82519328
传 真 0451-82519699
经 销 新华书店
印 刷 哈尔滨工业大学印刷厂
开 本 787mm×1092mm 1/16
印 张 12.75
字 数 153 千字
版 次 2008 年 5 月第 1 版
印 次 2008 年 5 月第 1 次印刷
定 价 35.00 元

<http://press.hrbeu.edu.cn>

E-mail: heupress@hrbeu.edu.cn

前 言

我国的高等教育评估，从1985年6月原教育部在黑龙江省镜泊湖召开“高等工程教育评估专题讨论会”开始，已经走过了二十多个年头，大致经历了理论研讨、评估试点和稳步发展阶段。以1990年10月教育部颁布的《普通高等学校教育评估暂行规定》为标志，我国的高等教育评估步入了制度化、规范化的轨道。高等教育评估在加强对高等学校的宏观管理、改善办学条件、端正办学指导思想、深化教育教学改革、加强教学管理、提高教育教学质量方面起着十分重要的作用。随着我国高等教育步入大众化教育阶段，如何监督并有效地提升高等教育质量已经成为社会普遍关注的话题。二十年的实践证明：高等教育评估是提高教育教学质量的有效途径之一。

随着我国高等教育评估工作的日益深入，在评估实践中也遇到了一些值得思考和探讨的理论问题和现实问题。诸如，评估指标体系的科学性问题、评估的可操作性问题、评估主客体关系问题、如何建立独立于政府和高等学校以外的评估中介组织问题，以及如何借鉴外国，尤其是发达国家评估成功经验和做法的问题，等等。这些问题如果得不到及时解决，将会影响到高等教育评估的健康发展。

萌生编著本书的初衷缘于作者多年从事高等教育评估的实践与研究工作。1998年，牡丹江师范学院作为20世纪70年代以后建立的本科院校，接受教育部本科教学工作合格评估，那时，我们被抽调到评估工作办公室从事学院的自评自建工作，期间，我们认真研读了许多关于评估方面的书籍，收集了大量国内外有关评估方面的资料，在工作中也积累了较为丰富的实践经验，并撰写了一些相关理论文章，为本书的撰写提供了素材，奠定了实践基础。2005年，我院再次接受教育部本科教学工作水平评估，我们又经历了整个评估过程，对我国高等教育评估的发展历程又多了些理性思考和新的认识。2006年，我

们主持了省教育厅科研课题《高等教育评估比较研究》，开始对中外高等教育评估的异同进行系统研究，通过对国外高等教育体制和评估机制的比较研究，越发觉得国外高等教育体制和评估机制中真是有很多值得我们去研究和借鉴的东西，于是便产生了要写点什么的冲动，一来是对自己多年研究资料、研究成果进行一次系统梳理，二来也可以为我国高等教育评估的完善提供一点可供借鉴的不成熟意见，也正因为这样，才产生了呈现在您面前的这部拙作——《比较与借鉴——中外高等教育评估体系研究》。

本书在写作和出版过程中，得到了牡丹江师范学院领导的大力支持，党委书记、院长修朋月教授在百忙当中高度关注本书的出版，学院研究生管理处在经费上也给予了大力支持，在这里一并表示感谢。

由于我们水平有限，加之占有资料不足，错误和疏漏之处在所难免，恳请专家不吝赐教。

张德才 陈虹岩

2008年3月

目 录

第一章 高等教育评估概述	1
第一节 教育评估的历史与发展	1
第二节 高等教育评估的概念	9
第三节 高等教育评估的目的和作用	16
第四节 高等教育评估的理论基础	18
第二章 法国的中央政府主导评估模式	22
第一节 法国高等教育管理的基本体制	22
第二节 法国高等教育的质量评估机制	27
第三章 德国的地方政府主导评估模式	42
第一节 德国高等教育管理的基本体制	42
第二节 德国高等教育的质量评估机制	48
第三节 德国高校评估对拨款的影响	50
第四章 英国的委托独立机构评估模式	54
第一节 英国高等教育管理的基本体制	54
第二节 英国高等教育的质量评估机制	58
第三节 原英联邦体系高等教育质量评估的基本特点	68
第五章 日本的双重结构评估模式	79
第一节 日本高等教育的结构特征	79
第二节 日本高等教育的质量评估机制	86
第三节 日本与英国高等教育评估制度的差异	97
第六章 美国的多元化评估模式	99
第一节 美国高等教育管理的基本体制	99
第二节 美国高等教育的质量评估机制	110
第三节 中美高等教育评估制度的差异	117
第七章 我国高等教育评估的回顾与展望	131

第一节	我国高等教育评估的历史回顾	131
第二节	我国已经实施的评估的主要类别和项目	136
第三节	我国高等教育评估的基本特点、发展趋势与展望 ...	139
第八章	我国高等学校办学水平综合评估	146
第一节	我国高等学校办学水平综合评估的目的和依据	146
第二节	我国高等学校办学水平综合评估的基本内容	154
第三节	我国高等学校办学水平综合评估的组织	156
第四节	我国高等学校教育评估的一般程序	159
第五节	我国高等学校教育评估的队伍建设	170
参考文献		196

第一章 高等教育评估概述

高等教育评估,是按照一定的指标体系,采用科学的、客观的方法,系统地收集、整理评估客体有关教育活动及其效果的数据资料等信息,并对之加以分析,形成对其社会价值的判断的过程。从专业学科角度来说,高等教育评估是教育评估学下设的一个分支学科,与其他学科一样,也经历了一个产生、发展到完善的过程。作为教育评估大系统下的子系统,高等教育评估毫无疑问要遵循教育评估的规律,因此,研究高等教育评估,首先应该从教育评估入手。

第一节 教育评估的历史与发展

研究教育评估的历史,必须遵从史实,进行必要的考证,根据文献资料来证实、说明、作出科学的判断。同时,还要兼顾国内外有关教育评估的文献资料,既要考证国外有关教育评估的文献资料,又不能忘记中国教育评估发展的历史。

一、中国教育评估的历史发展脉络

教育评估起源于中国,这是许多国外教育家公认的事实。

远在公元前 551—479 年东周时期的孔子就有教学评价思想。孔子是第一个创设有一定规模的私立学校的教育家。孔子弟子先后有三千人。孔子将其分为三类,“中人以上可以语上也,中人以下不可以语上也”(《雍也》)。这就是说,对中等以上水平的学生才可以传授高深的学问。在三千弟子中高才生只有七十二人,孔子对他们也有自己的评价:德行最高的有颜渊、闵子骞、冉伯牛、仲弓;善于言语的有宰我、子贡;长于政事的有冉有、子路;擅长文学的有子游、子夏。孔子按各人特点进行教学,给予解答。这就是在对人评价基础上的因材施教。

施教。

我国对学校进行评价的萌芽是从隋唐时期的科举制度开始的。公元 607 年隋炀帝提出十科举人，这改变了隋文帝凭门第高低和文章华美取士的保荐用人制度。在十科举中有“文才秀美”一科，即进士科，以考诗赋为主；南北士族人才凭文才竞争高低，取得士位。这是科举制度的开始。到了唐太宗时期，士主要来源于学校生徒和州县乡贡。那时学校分六种：国子学、太学、四门学、律学、书学、算学，统称国子监。由国子监推荐的学生到授有士位，必须经吏部考试，考的科目有秀才、明经、俊士、进士、明法、明字、明算等，其中明经（主要考帖经）、进士（主要考诗赋）两科尤为重要。那时的文人多愿应进士科，以表现自己的才能。考试及格称为及第，进士及第人数仅为考生的百分之一二。从此，科举作为一种取士的制度定下来，这实际上是以考试测验学力向朝廷推举人才的一种评价学校毕业生的制度。当时的名人如房玄龄（唐开国名相，隋末进士）、韩愈（唐代文学家，曾任监察御史，贞元进士）、白居易（唐代诗人，曾任秘书省校书郎，贞元进士）都是进士出身。可见，科举制在我国古代人才选拔上是起了积极作用的。它一直被沿用到 1905 年袁世凯奏请废除为止，共有一千三百年的历史。它是一套以各地预选和全国统考相结合，以政治条件、文化水平和身体状况综合评价为基础的人才选拔制度（如明嘉靖年间规定考生如有“离经叛道诡辨邪说”的，即使考中也要“默为民”；元代规定“患废疾”者不许应试）。我国的科举制度对许多国家公职人员的选拔有重大影响。如日本在公元 7 世纪就已开始采用这种制度；明治维新和第二次世界大战战败前，在所实行的高级文官考试中以及现在国家公职人员高级职务考试制度中仍保持着这种影响。又如法国在 1701 年参照我国科举制度建立了他们的文官考试制度。

至于学校教育评价的另一种形式——教育视导，则早在汉代就已在我国出现。那时，太学教师除了在学校教书外，有时还要视导地方的政教。教育视导活动发展到清末光绪年间，形成了较为完备的教育视导制度。当时的学部（即朝廷的教育管理机构）设视学官 12 人，专为巡视京外学务；各省设提学使司，下设省视学。

1909年颁布《视学官章程》，将全国分为12个视学区，每区按年派视学官2人。每年视察三、四个区，每三年视察一遍（实际上并未完全实行）。民国元年（1912年）中华民国政府建立教育部，次年即公布《视学规程》，全国分八个视学区，每区派视学2人，视察普通学校教育和社会教育，视察事项主要有：（1）教育行政；（2）学校教育；（3）学校经济；（4）学校卫生；（5）教职员执行任务情况；（6）社会教育及其设施；（7）总长特命视察事项。这种教育视导制度一直延续到全国解放以后。

二、国外教育评估的历史演变

学校教育评价虽然在我国有悠久历史，但始终未得到应有的发展。主要原因首先归结于封建社会的桎梏，地主阶级通过政权把教育严格控制在自己手里，用科举制度把他们自己的子弟培养成政权中的官吏，现代学校教育得不到发展。其次，从1840年鸦片战争开始的帝国主义列强入侵，使中华民族长期处于危难之中，半封建半殖民地社会不允许学校教育得到健康发展。在这种情况下，谈不上开展学校教育评价工作，无法研究教育评价理论。

学校教育评价形成独立的研究领域，并逐步向科学化方向发展，作为教育科学的一个组成部分，是19世纪末和20世纪内的事。国内外许多教育家也公认在学校教育评价的理论和实践方面做得比较全面、比较完整的代表性国家是西方世界的美国。西方学校教育评价的发展大体分成三个时期。

1. 萌芽时期（从19世纪中叶到20世纪30年代，前后约80年）

在这个时期内，学校教育评价主要是围绕对学生学力测量怎样客观化、标准化的问题进行的。以往，学校对学生学力测验的基本方法是口试，缺乏严格的客观标准。1845年，美国教育家梅恩（Mann）首先在马萨诸塞州波士顿文法学校引入书面考试、统一试卷。1846年，英国格林威治医学校教师费舍尔（Fisher）对书面考试的成绩标准进行了研究。1879年，德国心理学家冯特（Wundt）在莱比锡大学建立了世界上第一个心理学实验室，摸索了一套心理测量的方法。1882年英国

科学家高尔顿 (Galton) 在伦敦设立了人类学实验室, 对人类心理遗传及个体差异进行研究并设计了许多统计方法。冯特和高尔顿的研究成果对学生学力测量产生了较大影响, 在他们的影响下, 1894 年美国心理学家卡特尔 (Gatell) 在哥伦比亚大学试用各种测验考核本校大学生, 这是心理测量对教育测量的一种推动。1897 年美国教育家莱斯 (Rice) 发表了他八年来对一万六千名学生所做的拼字教学测验成果, 这种长时期对大量学生进行教学测验的研究引起人们对教育测量方法研究的普遍关注。1904 年美国心理学家桑 (Thorndike) 发表《心理与社会测量导论》, 提出“凡是存在的东西都有数量, 凡有数量的东西都可测量”的论断, 促进学力测验 (如斯顿 1908 年的数学标准学力测验) 和智力测验 (如比奈和昂利的实用智能测验)。1905 年的进一步发展, 对教育测量的客观化、标准化起了很大的推动作用。以后, 在西方, 各种标准化试卷被大量编制出来。

随着教育测量的发展, 用学生学力测验结果来评价学校教育计划有了萌芽, 与学力测验相关的升级率、退学率、教学效率等指标及其评价方法开始出现。例如, 19 世纪中叶美国有些地区开始用学力测验结果进行校际比较, 20 世纪初美国有些州和市将所属几十所学校用学力测验的数据对其教育成就和效率进行调查等。与此类似, 另一类学校教育评价——鉴定, 在美国开始萌芽。由于 19 世纪末美国院校发展过快, 其中有些显然不合格, 于是国会要求对院校下一个定义, 一些教育协会着力于研究怎样对学校以一个最低标准进行鉴定。1896 年美国北中部院校协会率先提出鉴定学校的五方面要求, 1901 年开始对所属中学进行鉴定, 1916 年鉴定大学院校的工作已成为该协会的重要任务。

可见, 在本时期内, 以学力测验客观化、标准化为重要对象的教育测量研究得到蓬勃发展, 对学校的教育评价 (包括对学校的鉴定) 进入萌芽阶段。在学校教育评价中, 评价的主要对象是学校的教育成就和效率, 评价的基本依据是学生学力测验的数据, 评价的主要手段是教育测量, 测量的主要内容是学生对教材的诵记。

2. 形成时期（从 20 世纪 30 年代到 50 年代，前后约 20 年）

在这个时期内，出现了以教育目标为依据的“泰勒评价模式”，形成了一套比较完整的学校评价体系，学校鉴定制度日趋完善。

教育测量虽不失为学校教育评价的重要手段，但它不能充分反映涉及因素较多、复杂程度较高的学校教育（特别是高等学校教育）过程。尽管美国在 1928 年，标准心理和学力测验的方法多达三千多种，仍不能把诸如社会对学校教育需求、教师教学和学术水平、学生的思想品德等影响教育质量的重要因素反映进来，因此它不断受到人们的质疑，对它的批判也就随之产生。与此同时，美国发生了一次严重经济危机：工厂倒闭，大批青年没机会就业，纷纷涌向中学，而中学的课程设置主要为考大学服务，这就使学生的就业需要和学校课程之间发生尖锐的矛盾。在这种情况下，美国一些教育家成立了“进步主义教育联盟”。他们边批判旧教育边寻找新出路，边进行学校教育改革探索边开展学校教育评价，为此成立了由俄亥俄大学教授泰勒（Tayler）领导的评价委员会，在卡内基基金会资助下进行长达八年的“课程与评价的研究”，这就是美国教育史上著名的“八年研究”。泰勒对评价研究的基本思路如下。

（1）教育是改变人类行动方式的过程；教育目标是学校希望实践的那种改变人类行动方式的愿望。

（2）教育计划能够根据对该目标实际完成情况的分析进行评价，所以教育评价就是一种衡量达到教育目标的过程。

（3）人类的行动是复杂的，所以要从各方面进行评价；教学活动的类型和方法，是评价的重要样本和对象，它不仅是可分析的，而且也是可综合的，评价要注重综合性，注重学习的各个领域。

（4）作为评价方法仅靠用纸和铅笔的测验是不充分的，应采用包括观察行动、调查、评定在内的广泛方法；评价的性质能够直接影响教与学。这种思路强调：“学校教育评价的目的是衡量学校教育活动达到教育目标的程度，包括教育目标所规定的掌握知识的程度、预期学生应达到的价值观、心理状态、知识应用、能力水平等方面的程度，而教育测量只是它的手段，这种手段应因评价客体的不同而不同”。根

据泰勒思路所设计的学校教育评价方案和所开展的评价活动能够获得关于教育目标完成情况的信息，有助于发现存在的问题、改进教育工作，因而受到广泛欢迎。

与此同时，鉴定在美国也得到正常发展。从20世纪20年代开始，美国大学协会一直进行高等院校的鉴定工作，后来移交给美国教育理事会，1949年成立国家鉴定委员会，协调全国高等院校的鉴定工作。他们认为鉴定的作用是：认可一所高等学校或专业教育计划是否达到了预定的资格和标准。从这里可以看出泰勒强调“教育评价的目的是衡量学校教育活动达到教育目标程度”的思想。

可见，在本时期内，学校教育评价以教育活动的各个领域（如课程、教育计划、学生学习效果等）为主要对象，评价的基本依据是教育目标，评价的主要手段除了教育测量以外还包括观察、调查等多种方法，评价注意综合性，这是本时期与萌芽期的重要区别。

3. 发展时期（从20世纪50年代至今，约60年）

在这个时期内，学校教育评价出现百家争鸣的形势，各种新的教育评价模式纷纷涌现。从20世纪30年代到50年代，泰勒的教育评价理论一直占据着西方世界的主导地位，可是到了1957年，情况发生了变化。这时，前苏联发射了第一颗人造卫星，这件事震动了美国朝野，人们开始怀疑美国学校的教育质量，认为这是学校教育中忽视基本理论的结果，于是举国上下呼吁教学改革，并为此投入了大量资金，由此使得对学校教育效果的评价显得至关重要。正是在这种形势下，学校教育评价的思想和方法得到长足发展。其中具有代表性的思想如下。

（1）泰勒的学生布鲁姆（Bloom）于1956年发表《教育目标分类学——认知领域》。它对于完善教育目标和教学理论起了重要作用，也为泰勒的教育评价模式提供了理论基础。

（2）1963年，克龙巴赫（Cronbach）发表题为《通过评价改进课程》的论文，对泰勒评价模式提出不同见解。他认为：①评价人不仅应关心教育的目标，检验教育目标达到的程度，更应关心教育的决策；②评价的重点应放在教育过程之中，而不是在教育过程结束之后；③评价不是竞争决定优劣，而要作为一个收集和反馈信息的过程，这些

信息对改进课程有重要意义。克龙巴赫的论断，经后来实践检验，证明是正确的。

(3) 1966年，斯塔弗尔比姆(Stufflebeam)提出以决策为中心的CIPP评价模型。CIPP是背景(Context)评价、输入(Input)评价、过程(Process)评价、结果(Product)评价的第一个字母。斯塔弗尔比姆同样对泰勒评价模式提出异议。他们认为“评价不应局限于评判决策者所确定教育目标所达到预期效果的程度，而应该收集有关教育方案实施全过程及其成果的资料，为决策提供信息的过程”。为此，它把决策分为四类，组成一个评价体系。这就是：①预期结果的决策，为此提供信息的是背景评价；②预期方法的决策，为此提供信息的是输入评价；③实际方法的决策，为此提供信息的是过程评价；④实际结果的决策，为此提供信息的是结果评价。其中背景评价是周期性进行的，而后三种评价则在背景评价基础上提出需要、发现问题，并在要求予以解决时进行。在CIPP评价模式中作出评价结论的依据不是决策者预定的教育目标，而是教育活动参与者的意图，可以反映评价对象的全貌，因而它具有民主性。

(4) 1967年，斯克里芬(Scriven)发表《评价方法论》一文，文中明确提出以下几种评价概念的区别：①形成性评价和终结性评价；②专业性评价和业余性评价；③比较性评价和非比较性评价；④目标达到程度的评价和目标是否值得达到的评价。其中在教学过程中进行的，目的在于形成适合于教育对象的教学过程的“形成性评价”，已成为当今学校教育评价的重要内容。

(5) 1975年比贝(Beeby)把评价定义为“系统地收集信息和解释证据的过程，在此基础上作出价值判断，目的在于行动。”他强调：①评价要收集系统而非零散的资料；②资料必须经过精心整理、精心解释，引入评判性的思考；③思考的结果要对教育活动的价值作出判断，包括对教育目标本身作出判断；④价值判断的结果应有助于决策的科学化并付诸今后的行动，亦即评价应对实际工作有指导意义。

(6) 在20世纪70年代中期，由斯塔克(Stake)提出并由古柏(Guba)、林肯(Lincoln)等人进一步发展而成型的“应答式评价模

型”。这种评价模型要求：①评价人要与被评价活动有关的各种人员接触，了解他们的愿望、看法，并做实地考察；②根据由此获得的信息确定评价范围、设计评价方案、选择收集信息的方法；③对所收集的信息进行分析判断并按需要回答的问题分类；④将分类评价结果写成报告分发各有关人员；⑤在分类评价的基础上对教育方案作出全面评价。应答式评价模式在思维上是发散而非收敛的，在价值观念上是多重而非单一的，在方法上是自然主义而非科学主义的，在判断结论上是定性分析的而不是定量分析的。它与西方整个社会的意识相吻合，具有比 CIPP 模式更广泛的民主性，因而一经提出就受到重视和欢迎。

(7) 美国的高等教育鉴定从 20 世纪 50 年代至今，已稳定地发展了 50 年，自 1949 年成立国家鉴定委员会 (NCA) 以后，1964 年成立高等教育地区鉴定委员会联合会 (FRACHE)，1975 年 NCA 和 FRACHE 合并成中学后教育鉴定委员会 (COPA)，它是一个得到教育部承认的民间组织，至今一直在主持全美的中学后教育鉴定工作，包括高等院校的鉴定和各类专业教育计划的鉴定。它对促进美国高等教育的进步，改进现有高等教育计划和发展未来高等教育计划起着重要作用。

纵观上面列举的以美国为代表的西方学校教育评价发展的一百几十年的历史，可认为它经过以下历程。

(1) 从评价的对象分析。它从评价学生个人学力，发展到评价教学中的课程，进一步再发展到评价以教育计划为主的学校教育活动的各个方面。

(2) 从评价的目的分析。它通过教育测验来选拔适合教育的学生，发展到评判教育活动是否达到预定的教育目标，进一步再发展到通过评价来改善教育决策、提高教育质量、促进教育改革、推动教育进步。

(3) 从评价的手段分析。它从推崇各种标准化、客观化的教育测量，发展到提倡观察、调查等手段的定性分析，进一步再发展到广泛收集信息、进行解析论证、作出价值判断的一种定量与定性相结合的方法。

(4) 从评价的形式分析。它从成果性评价，发展到形成性评价，进一步再发展到背景评价、输入评价、过程评价、成果评价等多种

评价。

(5) 从评价的程序分析。它从只按照教育测量结论衡量教育工作，发展到依靠评价人观察教育效果是否符合教育决策者制定的教育目标，进一步再发展到评价人和教育活动参与人广泛接触，共同确定评价内容和实施方案，以便对教育决策作出修正，对教育方案作出判断，对多数人意愿作出应答，使教育满足各种人的需要。

总之，学校教育评价的范围愈来愈宽广；目的愈来愈明确；手段愈来愈科学；形式愈来愈多样；学校教育评价也愈来愈注重民主化、大众化，为各种从事学校教育工作的人所接受、应用，成为推动学校教育进步的一种科学管理方法。

第二节 高等教育评估的概念

一、教育评价及其实质

教育评价起源于学生学力测验的客观化和标准化。在 20 世纪 20 年代曾形成一种教育测验运动，那时，人们并没有教育评价的概念。教育测验，包括学力测验和后来发展的智力测验和品格测验，只是用一种教育测量的方法对个人教育成就进行考查；而教育测量则是对这种成就的直接的或非直接的衡量。无论教育测验或教育测量都不能对教育活动进行教育评价，因为它们无法衡量教育活动满足两个根本的教育目标的程度——教育满足社会需求的程度以及教育满足人的自身发展的程度。因此，教育测验运动逐步过渡到教育评价。

教育评价是什么？这个问题从泰勒提出教育评价的概念起在西方一直有不同的见解。

泰勒认为：“在本质上，教育评价过程乃是一种测量课程和教学方案在多大程度上达到教育目标的过程。”在这里，泰勒把教育评价的实质理解为对教育或教学活动达到教育目标程度的一种衡量。

而克龙巴赫则把教育评价理解为：“为作出关于教育方案的决策收集和使用信息。”后来斯塔弗尔比姆也提出评价“不是为了去证明，而

是为了去鼓进”的论断，认为教育评价是“满足决策和教育效能核定的需要，描述、获得和运用有关客体的目的、计划、过程和成果价值的过程”。他们的共同理解是，教育评价实质上是一个为决策者搜集、描述、使用信息价值的过程。

泰勒和克龙巴赫、斯塔弗尔比姆论点中的差异在于：前者把教育评价看成一种符合教育目标程度的判断；后者则把教育评价视作一种为教育决策提供叙述性和判断性信息的过程。他们都未提及“价值判断”，而看作是教育决策者的事情。约十年后，比贝对教育评价提出了有意义的见解，他认为教育评价是“系统地收集信息和解释证据的过程，在此基础上作出价值判断，目的在于行动”。在这里，比贝表述了教育评价是一种目的在于行动的价值判断。

笔者认为，评价和认识不同。对教育的认识，是人脑反映教育客体（事物或人物，下同）属性和联系的一种心理活动；它着重在体现教育客体的规律和尺度，是客体性认识的过程和结果。而教育的评价，则是作为评价主体的评价人，在头脑中反映教育客体的属性和联系的基础上，依据主体需要，所建立的对教育客体有关价值的评论；是主体性认识的过程和结果。简言之，对教育的认识是人对教育客体的客观反映，反映是人认识客体的必然；而教育的评价是以评价人对教育客体的主体价值观来衡量的判断，进行价值判断是人评价客体的必然。反过来说，评价总是包含着对一定价值关系可能后果的预见和推断，它不可能只是对信息的简单描述，没有价值判断的描述只能称为认识，不称其为评价。

那么，教育评价中的主体——评价人，能不能在收集信息的基础上对教育客体进行价值判断呢？回答显然是肯定的。但由于评价必然带有主体性；而人们应力求评价的结论尽可能多地符合客观实际，减少主观色彩。因而，只有参与收集、分析、综合信息的评价人，才能够得到比较切合实际的判断；只有当评价人能够以反映社会和人民群众根本利益的价值观作为评价的依据，才能够得到比较符合社会和人民需要的评价。由此可见，在教育评价过程中，评价人不仅必然地要进行价值判断，而且应努力寻求，较好地进行价值判断。