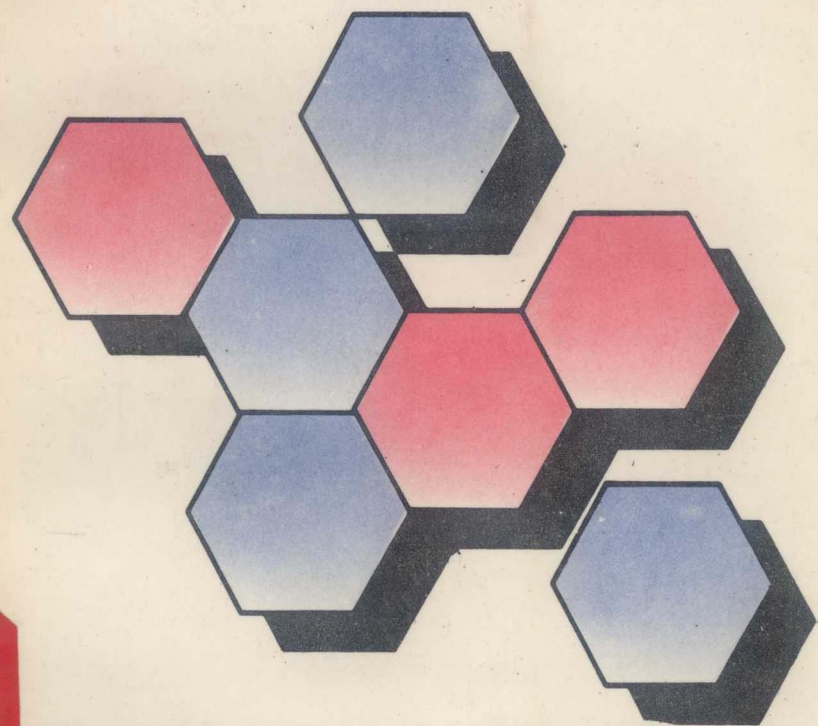


我国普通教育 评价模式研究

宋伏秋 梅克 主编



中国和平出版社

我国普通教育评价模式研究

主 编：宋伏秋 梅克
副主编：张秀媛 顾志跃 黄利群



96023501

中国和平出版社

G62/1

(京)新登字 086 号

图书在版编目(CIP)数据

我国普通教育评价模式研究/宋伏秋等编著. —北京:

中国和平出版社, 1995. 6

ISBN 7-80101-518-5

I. 我… I. 宋… Ⅲ. 普通教育-教育评估-研究-中国
N. G62

中国版本图书馆 CIP 数据核字(95)第 07950 号

※

中国和平出版社出版发行

(北京市西城区百万庄大街 8 号)

邮编: 100037

新华书店经销

河北太行机械厂印刷厂印刷

※

787×1092 1/32 12.5625 印张 267 千字

1995 年 6 月第 1 版 1995 年 6 月第 1 次印刷

印数: 1—2500 册

ISBN 7-80101-518-5/G · 397 定价: 8.00 元

课题组主要成员：

宋伏秋	梅克	张秀媛	顾志跃
黄利群	杨海榕	盛昌兆	李吉会
侯光文	聂厚德	涂恒汉	贾国英
何兆华	吴守垣	陆仁林	杨佐荣
黄云龙	蒋国昌	曾欣然	肖鸣政
李缙仁	陶元红		

序

陶西平

《我国普通教育评价模式研究》一书是“八五”国家教委重点科研课题《普通教育评价基本理论与方法研究》的主要研究成果。这本书是课题组经过几年的共同研究，在实践的基础上依靠集体力量而写成的。

10年来，我国普通教育评价有很大发展，从介绍国外教育评价理论开始，经过科学实验，在基层学校普及评价科学知识，与教育行政管理、督导部门相互配合，广泛地开展了教育评价实践活动。由于大家运用教育评价手段去实现教育目标，促进了教育质量的提高，推动了教育改革。在深化教育改革的过程中，教育评价既是改革的重要内容，又是改革的动力与尺度，它直接影响教育事业的发展，因此教育评价受到普遍关注。

教育评价科研工作者在实践的基础上加强了研究，总结了我国普通教育评价的实践经验，使我国普通教育评价的理论观点、方法技术和评价模式等方面有新的突破。教育评价的范围也从过去限于学生评价扩展到整个教育领域，评价功能也从分等鉴定发展到评价导向、改进、激励、教育功能，通过评价引导大家改进工作，驱动他们的内部活力，调动人们的积极性。随着我国经济体制改革深入发展，计划经济向市场经济转轨，以及教育领域应试教育向素质教育转变的新形势的出现，教育评价如何适应这种新情况，如何针对这个新

问题加强研究，是当前我国教育评价面临的主要挑战。为此教育评价科研工作者要提高认识，加强科学实验，通过理论研究，指导评价实践，开拓评价的新局面。

“八五”国家教委级重点科研课题《普通教育评价基本理论与方法研究》，在“七五”课题研究的基础上，着力于理论探讨，力图使理论成为评价实践的先导，回答实践中提出的问题。课题组通过对80年代我国教育评价的回顾，认真总结经验，理清研究思路，同时也借鉴国外教育评价理论，了解国外教育评价发展的新动向与新趋势，坚持理论与实践结合，把握我国普通教育评价发展的新走向。课题组从理论研究入手，重点开展了多种评价模式的实验研究，通过几年的努力，突破了单一评价模式的束缚，为创立符合中国国情的多种评价模式并存的局面，打下了理论基础。

这本书的编写过程，是在马克思主义指导下，从中国的实际出发，立足于建立具有中国特色的教育评价体系与模式，因此本书有了一些新的特色与深度。本书的主要内容有我国普通教育评价模式的理论基础和基本观点、构建评价模式的原则、结构、方法与类型，以及国内外教育评价模式分析与举例等。它是大量实践经验的升华，也是我国教育评价科研工作者、教育工作者和教师的智慧结晶。

这本书的意义不仅在于评价模式本身，还在于为我国教育评价的发展找到一个新的生长点，为教育评价研究向纵深发展开辟了新的思路，而且它也具体地回答了新形势下我国普通教育评价理论与实践的问题。这本书有较强的理论性、实用性和可操作性，无论对教育科研部门，还是对教育行政管理、督导部门和学校都是有实用价值的参考书。

目 录

第一篇 评价模式研究概述	宋伏秋(1)
研究教育评价模式的背景与意义	(1)
一、我国教育评价的由来与发展	(1)
二、我国普通教育评价现状分析	(15)
三、研究教育评价模式的意义	(32)
教育评价模式的基本概念	(39)
一、教育评价的概念	(39)
二、教育评价模式的定义	(46)
三、教育评价模式的基本特征	(48)
四、教育评价模式的构成与类型	(52)
我国的国情与教育评价模式	(62)
一、我国的教育体制与教育评价模式	(62)
二、我国各地发展不平衡与教育评价模式	(71)
三、我国教育评价实践与教育评价模式	(76)
教育评价模式的构建	(83)
一、教育评价模式构建的指导思想与原则	(83)
二、教育评价模式构建的方法	(92)
三、教育评价模式构建的程序	(114)
第二篇 模式基本理论与方法探索	(121)
教育评价模式的理论探索	顾志跃(121)
教育评价模式方法研究	侯光文(142)

第三篇 我国评价实践中形成不同的模式举例…………… (172)

教育型目标调控模式…………… 梅 克(172)

指标量化、评语描述模式…………… 宋伏秋(210)

学校评价“两体一制”模式…………… 涂恒汉(354)

代表性品德行为整体测评模式…………… 聂厚德(375)

第一篇 评价模式研究概述

宋伏秋

研究教育评价模式的背景与意义

一、我国教育评价的由来与发展

(一) 教育评价的历史发展

1. 中国古代的考试制度

考试主要指采用口头或书面回答问题的方法来检查受试者的知识与能力。这是历代统治阶级最常用的选拔人才的方法。早在西周时代,我国就采用以射选士,《礼记·射义》中说:“古者,天子以射选诸侯、卿、大夫、士。射者,男子之事也,因而饰之以礼乐也。”又说:“天子之制,诸侯岁贡于天子,天子试之于射宫。”从秦汉至隋唐,封建王朝选拔官吏的方法基本上也是以选为主,以考为辅,先乡举里选,再由朝廷考试,然后任官。自从隋炀帝大业二年(公元606年)始置进士科,试策问,这就是以考为主科举制度的开端。封建科举制度是以封建主义的价值观为准绳,把政治审查、学力测评结合在一起形成的一套独特的评价制度。这套选拔人才

的制度在封建统治阶级笼络和控制知识分子上起到了重要作用，到1905年袁世凯奏请废止，沿续了1300年之久。

科举制度实际上是一种各地预选与全国统考结合的制度。唐代曾规定，考生须经地方考试合格才有资格参加全国统考。宋代也规定，考生须经“州长吏先校试合格，始听取解”。清代则更为严格，读书人未经考试者称“童生”，经县试、府试、院试合格者称“生员”（亦称“秀才”），经乡试合格称“举人”，再经会试合格称“贡士”，最后经殿试合格称“进士”。由此可见，历代科举都须经一系列考试，它是一套严格的地方预选与统考结合的制度。

科举不仅是对考生学习、文化水平的测评，同时也有很高的政治要求。如唐宪宗时规定，报名者如“迹涉流狂，兼亏礼教，或曾任州府小吏有一事不合清流者，虽薄有辞艺，不得申送。”明嘉靖年间规定，考生如有“离经叛道，诡辞邪说”的，即使考中也要“黜为民”。除此之外，对考生的身体状况也都有一定要求，如元代就规定“患废疾”者“不许应试”。所以科举制度是一套政治条件与文化水平并重，兼及身体状况的以综合测评为基础的人才选拔制度。

从评价的观点看，我国的考试制度早于西方1000年，而且有较完善的规定，考试方法大体有：一是帖经，将儒家经典掩其两端，中间唯开一行，再裁纸贴三字，让考生写出读出这三字，类似默写或填空。二是墨义，取儒家经典中的句子让考生对答或要求对答本句含义，或要求对答下一句，或要求对答注疏，类似今天的名词解释或简答题。三是诗赋或杂文，是选用儒家经典中的思想、前人的诗句或景物为题，规定韵脚，让考生按声韵格律要求作诗填词，类似命题作文。四

是经义，以儒家经典一段、一句或不同章节同一主题的句子为题，让考生作文阐述自己的认识和见解。五是策论，也称策问，有经史策和时务策两种，以历史典章、重大理论问题或时务问题为题，让考生分析议论，提出自己的主张和解决办法。以上考试方法，唐代宋初重诗赋，兼以帖经、墨义和策论，宋朝王安石改革科举制度，废止诗赋、墨义，以经义为决定考取的关键，以后各代虽恢复诗赋等，但仍以经义为主，直到废科举。可见封建科举考试始终以儒家经典为主要考试内容，特别重视文化素养和写作能力，并把考查分析政策理论问题的能力作为选拔官吏的重要条件。而且为了避免评卷者的主观性和舞弊，自唐代起就采取了“糊名”、“誊录”等办法。科举考试经过历代积累，形成了一整套取士的考试制度，在教育评价发展史上曾居于领先地位，特别是这种制度所形成的口头策问与论文式笔试的考试方法，对我国和世界其它国家的教育发展都产生过重大影响。

2. 西方教育测量运动

传统的考试制度从选拔人才上看有一定的作用，但主要弊端是考察的知识面窄，信度与效度差，难以考出学生的真实成绩，不能达到预期的考试目的。同时造成猜题、押题等投机心理，影响正常教学的进行。还有评分的主观随意性，波动性大，不同的评卷人评同一份试卷就会有不同的评分结果，甚至同一个评卷人在不同的时间对同一试卷也会有不同的看法。为了改革这种不科学、不公正的考试办法，西方国家自20世纪初，就掀起了一个科学的测验运动，这一运动的核心是追求测量的客观化，因此也叫客观化的测验运动。

西方各国古代教育，在漫长的历史时期里，教育上一直

偏重于口头背诵和冗长的论文式测验。随着社会经济的发展，当手工业经济逐步发展为新兴的资本主义经济时，对教育提出了更高的要求，不仅要求办更多的初等和中等教育，而且要求培养大量的适应社会需要的掌握读、写、算技能的人才。1845年美国波士顿正式出现了以学生成绩评价学校成就的倡议，在贺拉斯·曼的推动下，把笔试引入了波士顿文法学校。因为学生日益增多，笔试取代口试是不可避免的，当然笔试也有很多局限，虽同一试题可以同时考核一个集体，却存在考查知识范围很窄，考查能力也很难的问题，而且没有统一的标准，在评分过程中也存在主观性与片面性的问题。为了克服这一弊端，追求测量的客观性与科学性就成为一种趋势。英国科学家高尔顿 1882 年在伦敦设立人类学测验实验室，搜集人类身体特征和心理特征的资料，他对人类个别差异寄予关心，设计了问卷和处理统计方法。这些问卷和统计方法应用于教育，对教育测量的形成影响很大。他发现考试分数表现为正态分布曲线，即大多数有向平均数或正态分布曲线中心集中的趋势，离开中心而趋向两端的分数愈来愈少。因此，他认为人的能力也是正态分布，正确的测验分数必然表现为正态分布曲线。他还提出了统计学上重要的量数之一，相关概念，而决定测验的信度与效度及因素分析等都是这一发现的直接后果。其后美国心理学家卡特尔收集了大量数据，进行了一系列实验研究，更加强调数量法与等级法等。

1895 年，法国比纳与亨利设计了一套智力测验方法，主要是为了鉴别学习差的儿童，以便采取补救措施。1905 年比纳与西蒙共同发展了智力测验，用比较复杂的方法测量高级心理能力，编制了儿童智力发展量表，经过 1908 年、1909 年

两次修订，创制了智力测验量表，对教育测量影响是巨大的。

1904年，卡特尔的学生桑代克发表了《心理与社会测量学导论》，后人认为这本书是教育测量史上划时代的巨著。在这本书中，桑代克系统地介绍了统计方法以及编制测验的基本原理，并提出了一个著名的论断：“凡是存在的东西都有数量，凡有数量的东西都可测量。”这一论断对教育测量的发展起了很大的推动作用。

在教育测量的影响下，教育上开始采用客观化测试方法，以考查学生的学习成就。1908年美国教育家斯通第一个提出客观测验算术的方法，全部12问，从易到难排列，时间15分钟，对象为小学6年级学生。1910年桑代克提出书法量表。1915年查布设计语文测验。由于比纳、西蒙的智力测验传入美国，1916年由斯坦福大学推孟修订，形成了适合美国儿童的智力测验，同时又将测验结果量化，并引用智力商数加以表示。1912年弗纳尔德着手制作人格测验，到1924年还组织了人格教育调查会，进行人格测验工具的研究。教育测量发展很快，从试题的编制规则、难度、区分度的检验，测验的信度、效度的计算，一直到评分，统计分析等都有完整的精确的办法。1923年出版的斯坦福成绩测验是可以综合使用于全部课程的第一个主要的标准化成绩测验。它从有代表性的1500名儿童测试中得出了各种年龄儿童的国家标准，以便于任何个人同平均数进行比较。其后，又做了多次修改。

20世纪20年代是美国教育测量运动的高潮，使教育测量进一步发展，不仅使考试客观化、标准化，而且把学习能力、个别差异变成测定的对象，获得量的数据，这是教育测量取得的重大成绩。但由于教育现象本身的复杂性和多样性，

教育测验很难测量出教育的全部效果，比如社会态度、创造精神、兴趣、鉴赏力等，教育测验就不能充分测定。这样教育测验只能用于能够量化的对象，对于难以量化的对象，往往被忽略或被删掉，最后只偏重于容易量化的知识和记忆、理解之类的对象，因此教育测验就不能适应教育发展的需要，就出现了对教育测量的批判。

3. 现代教育评价的产生与发展

科学的教育测量运动主要追求的是使测量客观化，经过近百年的发展，它已形成一套比较完整的精确的测量与统计方法，它适用于竞争性的选拔考试。但后来逐渐暴露出它的弱点，因为它以正态分布的理论为基础，测验结果的分数必呈正态分布曲线，造成学生之间的激烈竞争。人们在批判教育测量运动中，也孕育了对教育效果新的判断方法，教育评价日益被大家所关注。

20 世纪 30 年代，美国在批判教育测量的同时，开始了从测验向评价的转变。美国杰出的教育家泰勒教授提出了现代教育评价的思想。他对课程和评价提出了广泛而有创新意义的概念体系，并进行了实验研究。他认为，一份好的测验，不应只是从教科书中挑选一部分内容改写成问题形式，而是应以一定的教育目标为指导，教师首先应研究课程，然后根据他们所希望学生应掌握的内容，用行为将课程目标表达出来，制定出教育目标。这样泰勒首次把测验与课程目标和教学方案联系起来，把测验的目的从着重揭示个人差异转变到努力寻找和说明造成差异的教学、课程上的原因上来。这就是现代教育评价思想的表述。这种思想的形成是有其历史背景的，由于 1929 年世界性经济危机波及美国许多学校，造成学校关

门，学生减少，教师被解聘，使教育界出现了危机。与此同时，危机也带来另一严重后果，工厂倒闭，一大批青年因无工作只能涌向中学。而美国中学的课程是为大学服务的，不适应失业青年的需要。这就使学生的需要与学校课程之间发生了矛盾。这种状态促使许多教育者力图革新教育，美国俄亥俄州立大学泰勒教授受卡内基金会的资助，受命进行课程与评价研究，于30年代中至40年代初，历经8年进行系统研究。通过这场研究，泰勒得出这样的结论：当时的课程与测验是教科书中心主义，测验的内容只是记诵教材的知识内容，是极为片面的，它不能真正反映学生全面的发展。为此他提出了一套以教育目标为核心、为依据的课程编制和测验编制的原则，试图以此把社会的要求、学生的需要反映在课程与测验之中。为了把这一思想与早期测验区别开来，泰勒等人正式提出了教育评价的概念。他认为教育评价就是衡量实际活动达到教育目标的程度，测验是它的手段。根据这一理论进行的评价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。同时它对学校鉴定这类活动也产生了重大影响。使学校评价的内容发生了重要变化。按照泰勒提出的评价理论实施教育评价，必须分析教育应达到的目标，再用这个目标来评价教育效果，并运用评价促进教育向理想目标逼近。

第二次世界大战以后，现代教育评价很快传播到西欧、日本等国，特别是由于世界范围的经济竞争和技术竞争引起的教育竞争，各国在强调开发智力资源、重视智力投资的背景下，普遍加强了教育质量与教育效果的检测和评定，推动了教育评价的发展。特别是在发展泰勒的行为目标模式方面开

展了大量研究，取得了许多进展。泰勒在别人帮助下建立了教育考试服务中心，设计标准化测验试卷，加强了对测验的管理。同时教育学家为了明确地表述教育目标，加强了对教育目标分类的研究，1948年美国心理学会与大学考试的专家们在波士顿召开心理学年会，正式提出建立教育目标分类体系的研究课题。其后布卢姆等人于1956年完成了认知领域的目标分类体系，1964年克拉斯沃尔完成了情感领域的目标分类体系。

1957年苏联卫星上天，美国朝野震动，一致认为美国科技所以落后，原因之一是美国教育落后。翌年，1958年美国联邦议会通过了“国防教育法案”，拨款近3亿美元来加强学校的数学、科学和外语教学，强调教学内容和手段的现代化，重视实验，培养天才儿童。既然大量投资于教育，就产生了投资效益问题，即投资的结果能否提高数学、科学和外语的教学质量，提高学生的教育成就。这样以测定教育投资的效益为背景，教育评价又受到普遍重视。政府聘请专家对新课程目标与效果进行评价，但这些评价活动并没有收到预期效果，而且回答不了改革中提出的问题。为此有人重新审视了当时占统治地位的评价理论与方法，认为评价以目标为中心，那么目标的合理性又根据什么来判断，况且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，那么这种非预期效果又如何评价。还有人认为，教育乃是个人自我实现的过程，用统一的目标模式去统一个人的自由发展，从根本上是不可接受的。

1963年，克龙巴赫在《通过评价改进教程》一文中，提出这种评价是以陈旧的思想习惯和确立已久的技术，是不足

以指导教程改革的，缺乏适当性与实用性。为此，他呼吁建立新的评价概念。

1965年，美国国会通过《初等及中等教育法案》，授权联邦政府补助各地中小学校，对文化差的学生实施补偿教育。法案规定，凡接受补助的均必须经过评价。接受补助的资格之一是地方教育机构先提出评价计划，州教育机构提出总结报告，报告中既要讲明经费预算的分配情况，又要有方案实施后学生进步状况的评价。这一系列措施使得评价在教育中的地位更加上升，吸引了许多著名学者致力于评价理论与方法的研究。从而一些学者提出新的评价理论，出现了众多的评价流派。1967年斯克里文创造了形成性评价与总结性评价两个术语，把评价引进了教育活动的过程中。他提出通过形成性，反馈矫正，使教育活动高质量完成，这就使评价产生了改进与完善事物发展的功能。

70年代教育评价进入了多元研究的新阶段，逐步形成了现代教育评价的新观念。各种评价理论相继问世，在西方出现的各种评价模式已达40余种。教育评价渗透到教育各个领域，成为80年代到90年代世界各国教育科研三大领域之一。教育评价的发展更加向专业化方向发展，对评价的认识再不能把它当作纯技术性工作，也不是现象的客观叙述，而是受价值取向所制约，是一种价值判断。评价时既要作量的描述，也要作质的描述；既要过去所作所为进行评定，也要为将来如何决策提供咨询信息。而且教育评价除对教育现象中人的特质进行评价，还要对课程、教学计划、行政管理、政策法规、教育改革方案等各种现象都应进行评价。这种教育评价新观念，赋予评价新的角色与功能，使教育评价成为一种