JIAO YU CE LIANG YU JIAO XUE PING JIA

侯光文 著

教育测量与教学评价

明天出版社

教育测量与教学评价

侯光文　著

明 天 出 版 社
1991·济南

教育测量与教学评价

侯光文　著

*

明天出版社出版

（济南经九路胜利大街）

山东省新华书店发行　山东新华印刷厂临沂厂印刷

*

850×1168 毫米 32 开本　19.875 印张　442 千字

1991 年 5 月第 1 版　1991 年 5 月第 1 次印刷

印数 1—371

ISBN7—5332—1197—9

G·626　定价：7.90 元

前　言

教育测量与教学评价的理论和方法近年来日益受到人们的关注，这是教育改革向纵深发展的必然结果。广大教师夜以继日辛勤耕耘，自然十分关心自己的劳动成效，希望借助科学的方法及时了解学生的发展变化，以便不断改进教学，提高教育质量；学校为了有的放矢地组织教育教学工作，需要准确掌握教学状况；教研人员在深入探讨教与学的规律时，必须驾驭科学的测评方法；各级教育行政部门为了有效地指导和调控教育工作，离不开科学管理；每年一度的招生考试更是牵动着千百万人的心弦；各行各业也需要有效地开发和利用社会人力资源。这一切都赖以科学合理的教育测量和教学评价。

然而，就目前状况看，我国教育测量与教学评价在理论和方法上，却远不能适应教育发展的需求。因此，迅速提高我们教育测评的科学化水平，建立具有中国特色的现代测量与评价的科学体系，广泛普及有关知识，充分发挥评价功能，便成为当务之急。

基于上述原因的启动，作者萌发了撰写此书的构想，并在编写过程中努力遵循如下原则：(1)从马克思主义哲学立场出发，认识教育本质，确立教育价值，以此为研究基点，揭示教育领域内测量与评价活动的客观规律，指导教育测评实践。(2)既挖掘传统方法的精华，又吸收现代科学技术的新成果，博采多种学科

之长，用以充实教育测量与评价的理论和方法。(3)总结我国历史上和现代教育测量评价理论研究成果和实践经验，并注意合理吸收国外先进理论，为建立具有中国特色的教育测量和教学评价的科学体系铺路奠基。(4)理论阐述与方法介绍相结合。教育测量与教学评价不仅是一种方法或手段，更是一种社会活动，我们应当透过大量现象揭示其固有的客观规律，加强理论研究，与此同时，探求科学的方法和技术。本书努力兼顾二者，并力求形成一个完整体系。

书中有不少地方涉及数学知识，对此，作者作了如下处理：对于重要的数学原理，尽量深入浅出加以介绍。对于过分艰深的数学推导，则采用形象直观的方式描述，或直接给出结论及使用方法，但同时介绍参考书目，为有兴趣的读者作进一步研究提供方便。对于有关方法与技术问题，注意通过实例说明其应用，以便广大教师和干部掌握和使用。一般具有中学数学知识即可阅读。另外，为了充分利用现代化手段，使庞杂冗长的计算得以简化，本书对计算器在这些问题中的使用方法作了介绍，并对运算量特别大的方差分析附有计算(微机)程序。

我国对现代教育测量与教学评价的研究尚处起步阶段，作者深知任务的困难与艰辛。然而，教育乃至社会发展对教育测评理论的急切呼唤，人们为改变这一领域的落后状态而奋力求索精神的感召，敦促作者不揣冒昧决然开笔。编写过程中曾得到不少专家学者的帮助和支持，借鉴了许多中外学者的研究成果和著作，初稿写成后，在有关讲习班多次讲述，不少同志提出了很多宝贵意见，在此一并表示诚挚的谢意！

如果通过此书能够为建立适合我们国情的教育测量与教学评价理论体系尽“引玉”之微功，作者便得到很大欣慰。由于水平

所限，不当之处在所难免，恳请指正。

作　者

1989 年 3 月初稿

1990 年 5 月修定稿

目　录

第一章　教育测量与评价的历史沿革

研究教育测量与评价，首先应了解它的历史进程，掌握它的发展规律，以便有效地分析现状、展望未来，进而向深层探索和发掘，在实践中充实它的理论，改进其手段与方法。

第一节　教育测量与评价的渊源在中国

正规而系统的教育测量与评价活动起源于中国，这已被世人所公认。二十世纪之前，我国考试制度的发展历程大体经历了两个阶段：

一、科举前期

早在西周（公元前 1100—前 771 年）时期的“国学”中的大学，就设有定期的学业考查制度。《礼记·学记》中记载：“比年入学，中年考核，一年视离经辨志，三年视敬业乐群，五年视博习亲师，七年视论学取友，谓之小成；九年知类通达，强立而不返，谓之大成。”可以看出学生入学之后，每隔一年就考查一次，并且具体规定每次考查的内容和要求，如果在第七年的考试中达到了标准，就是“小成”，若在所有的考试中达到了标准，称为“大成”，

这的确是一套有系统有目的的考核办法。汉代的太学，曾规定“一年辄课”制，即一年考一次，以后又改为“二岁一试”制。太学生只要达到考试的要求，就可以毕业，学校没有学习年限的规定。考试的方法有“口试”、“笔试”、“射策”三种。其中“射策”是由学生从事先写在竹筒上的不同试题中，随机抽取作答的一种考评方法，可见当时考试形式已相当丰富多采。

教育测量与评价的发展历程，总是与人才测评活动的发展密切相关的。一方面，人才德识的测评结果，正是对教育效果的衡量；另一方面，它的内容、方法和手段又直接左右着教育活动的发展。在我国，对人的德识的测量与评价活动不仅在学校教育中有所进展，而且在人才选拔任用上也创立了一整套标准与方法。从这个意义上讲，我国的教育测评活动可以追溯到更远的年代。

远在我国原始社会末期的尧舜时代（约公元前2200年），在对人才的选用上就有了测量评价的因素，当时的执法官皋陶从德、识、才能诸方面提出了选才任职的九条标准，合称“九德”。部落联盟各首领都根据这九条标准选拔和任用，而且每任三年便对他们的任职情况进行一次考核，三次考核之后，提拔贤能者，贬黜失职者。西周时期，虽然高中层官职由奴隶主阶级世袭相传，但是低级官职则是从士中由乡里逐级推选，称为“乡举里选”。选拔标准规定为德行和道艺。其中德行包括六德和六行。“六德”为明于事理、热爱人和万物、通达有见识、善于决断、言行一致和刚柔相济；“六行”：即孝敬父母、友爱兄弟、和睦家庭、善处外亲、信于朋友和赈济贫乏。道艺包括“六艺”即礼仪、音乐、射技、驾术、书法和算术。后来的春秋战国时期，又出现了士游说制，即君主先颁布诏令，然后从应聘的士中根据他们对问题的分

析情况授予官职。这些都是我国古代人才评价、选拔制度的雏型。

运用考试手段选拔人才始于汉代察举，具体方法是：朝廷根据政治、经济和军事等方面的实际需要，临时制定出科目，规定举选人的资格和被举选的人数及标准，并对被推荐举选上来的人进行考试。公元前165年，汉文帝出题考试，从士子和现任官员中选"贤良方正能直言极谏者"，他把试题写在竹简上，称为"策题"，试题分为四个策目"朕之不德、吏之不平、政之不宣、民之不宁"，这是我国取士考试的开端。皇帝用策题考士称"策问"，应试者在竹简上作答，写出答案称"对策"。策问和对策的创立，为测量评价人才学识优劣、才能高下提供了一种好方法，这是我国首次将笔试用于考试，它比欧美各国的笔试早得多，若按1702年英国剑桥大学在西方国家率先使用笔试计算，我国的笔试较欧美国家早一千八百年。

总的说来，这一时期，我国的考试制度还很不正规，就是汉代的察举也仍是以推荐为主，考试还不是选拔人才的主要手段。

二、科举的兴衰

东汉之后，我国的选士制度，又经历了魏晋的"九品中正制"的过渡，到隋炀帝大业二年（公元606年）开设进士科，开始了我国历史上的科举制度。

科举制度是采取举选和考试相结合而以考试为主要方式的选拔人才的制度。科举考试分设若干科目，按照考试成绩、依据才学取士，打破了在选拔人才时的门第观念。它作为九品中正制的否定物而形成，远源于汉察举，又比汉察举前进了一大步。科举制到了清代，更加突出了考试在选用人才中的地位和作用。

科举制对人才的选拔是逐级进行的。唐朝的科举考试分州

县试和礼部试两级，宋代又增殿试一级，到了明清时考试增为四级：童试，为乡试前的预选性考试，在各府举行，及格者为秀才；乡试，基本上以行省为单位，由秀才参加，及格者为举人，举人是乙级功名，有资格入仕。或为知县和教职，或赴吏部铨选；会试，在乡试的第二年举行，也是三年一次，由举人参加，是殿试的前奏，属甲级考试，及格者为贡士；殿试，由贡士参加，这是会试的继续和完成，也属甲级考试，以皇帝亲试的名义举行，由八名读卷大臣代替皇帝评卷，评卷成绩分为五等，根据八名大臣阅卷时划分的等级，进行统计、排列名次，考试及格者为进士，前三名分别称为状元、榜眼、探花。进士及第后，授给一定官职。

科举考试中的科目主要有进士、明经、明法、明书、明算五种，随科目的不同，考试的项目和内容也有不同的侧重，进士科主要考杂文诗赋和策论，其他科的考试分别以五经、律令、书法、算术为主。

科举的考试方法有口试、帖经、墨义、策论和诗赋。帖经是用纸遮盖经书中的一页、仅留一行，再帖上三、五字，令考生按原文填写。墨义是根据经书出题，要求考生用原文回答。策论是按题目的要求，写出论证文章自行发挥作出解答。诗赋须按题目的要求和规定的格律进行创作。这些方法正是今天的填空、简答、论述和作文等考试的源流。

科举制的创立，在测量与评价史上具有划时代的意义，它的进步性在于建立了完备的考试制度，创造了一整套行之有效的考试方法。当然科举制度也存在着严重的缺陷，其中最突出的就是内容陈腐，它一直沿用儒家经学作为考试内容。诚然，儒家学说其经典中的确保存着许多富有价值的思想文化，但是，随着时间的推移，其中若干内容已变为阻碍社会进步的教条，变为影响

新的思想文化产生和发展的桎梏。由于考试内容年久不变，而科名又被士子无限推崇到了如痴如狂的地步。于是，在当时的文人中间，出现了儒学之外没有学问，九州之外没有世界的荒诞信条，什么科学技术，发明创造，统统不屑一顾，视为邪说。愚蠢的井蛙之见，使我们同欧美的距离拉大了。科举制度发展到后来，已经严重地阻碍了社会的进步，终于在大清行将灭亡前的光绪三十一年(1905年)颁布诏书，废止科举，经历了1300年的科举制自此结束。但是它的考试方法作为测量评价人才的手段，是富有生命力的，不仅保留下来，而且在欧美等国得到了进一步发展。1583年，葡萄牙修士胡安·贡萨雷斯·德万多萨的《伟大的中国》出版，书中介绍了明代科举考试的内容和方法，此书被译成了多国文字而流传，从1570年到1870年间用英文出版的有关明清政治制度的书籍达70多种，都把科举制度当作重要的内容加以介绍。十八世纪法国资产阶级启蒙思想家伏尔泰和孟德斯鸠也对这种制度作过介绍评论。1866年北京同文馆馆长马丁在波士顿作《中国的竞争考试》的报告，开始向美国人介绍科举制度。孙中山说："现在各国的考试制度，差不多都是学英国的。穷流溯源，英国的考试制度，原来还是从我们中国学去的。"① 但是，由于种种原因，在近代我国对教育测量、评价的研究，一直进展不大，而我国的考试制度传到欧美之后，却得到了长足发展，形成了一整套现代教育测量与评价的理论与方法。

① 转引自盛奇秀：《中国古代考试制度》(山东教育出版社)第128页。

第二节　西方教育测量的历史发展

就在我国的科举制度停滞不前，并逐渐走向消亡的时候，在十九世纪末，二十世纪初，西方的教育测量伴随着心理测验理论的发展，出现了生机勃勃的局面，形成了历史上的测验运动，并且对二十世纪三十年代以后的教育评价的发展产生了深刻的影响。

在西方，教育测量的发展大体经历了三个时期：

一、先验期

这是指1904年以前的漫长阶段。十八世纪以前，西方学校教育测验主要是采用口试的方法。1702年，英国的剑桥大学首先以笔试替代口试，开西方学校笔试之先河。美国的笔试则开始于十九世纪中期，当时美国的教育发展较快，初等教育普及，学生人数骤增，口试已远远不能适应需要。1845年，在美国著名教育家贺拉斯·曼(Horace Mann)的倡导下，波士顿市教育委员会率先以笔试考查该市毕业生。

笔试的引入，使考试方法的客观性和可靠性都有较大的提高。但是，很快就有人发现，当时的笔试无论从命题上看还是从评分的过程分析，主观随意性都很大，不能很好地反映和评定学生的学力水平，而且也不便于比较。为了提高测验的客观性和可靠性，1864年出现了第一个教育测量的量表，这就是英国格林威治医院的一位教师乔治·费舍(George Fisher)所编制成的《量表集》。他搜集了各种学生的书法、拼字、算术、语法、作文、历史、自然、图画、法文等学科成绩样本，并为每一样本评定一种分数，

以示优劣，以此作为度量教育成绩和学生作品的标准。如果要评定学生某一方面的成绩，就可以将学生的作业与量表中的样本加以比较，找出与学生作业水平相当的样本，这时，样本的分数，即为某生应得的分数。当然，《量表集》中对各样本的评分，是仅凭他的主观认识和经验而定的，其客观性受到一定局限。但是，费舍的工作无疑是具有开创性的，这是用科学方法研究教育测量问题的最新尝试。不过费舍这一工作由于种种原因，没有得到当时教育当局的注意，因而它的量表在当时没有产生多大的影响。

在教育测验问题上，引起人们极大关注的应推著名的莱斯(Joseph Rice)拼字测验。这与当时的教育背景有很大关系。在十九世纪末，美国教育界发生了一场争论，争论的一方主张改革当时的教育，加入实用学科；而另一方则持反对态度，他们认为一旦加入新的课程，就会挤掉学生学习旧有的基本科目的时间，他们坚守学科训练的学说，只注重练习和背诵的教学方法。为了从客观上解决这场争论，1894 年莱斯设计了一种测验，他选定了五十个字来测量二十个学校的一万六千名学生的拼字能力，同时对各校每周教授拼法的课时数进行了调查。测验结果表明，八年中每天用四十五分钟学习拼法的学生，较之每天花十五分钟学习拼法的学生成绩没有多大差别，莱斯的这一结论当时引起了强烈反响，遭到了不少人的反对，不过这项研究引起了人们对测验问题的广泛关注。而且他所运用的客观方法，对后来教育测量的发展产生了深远的影响。

这一时期教育测验之所以引起了人们的普遍兴趣，与实验心理学和心理测量理论的发展也是密切相关的。这是因为教育测验与心理测验虽然所要测量的内容不尽相同，但是它们依据