


# 教育测量

JIAOYU CELIANG

# 与评价

罗黎辉 高翔 编著

云南教育出版社



责任编辑 高建勤  
封面设计 陈俊

书 名	教育测量与评价
编 著	罗黎辉 高翔
出版发行	云南教育出版社
印 装	云南新华印刷厂
规 格	850×1168 毫米 32 开本
印 张	9.75
字 数	220000
版 次	1996 年 2 月第 1 版
印 次	2002 年 2 月第 3 次印刷
印 数	9001 ~ 13000
书 号	ISBN 7 - 5415 - 1124 - 2/G·930
定 价	12.00 元

# 序

据考证，教育测量和评价这种教育现象和活动最早产生在中国。如果把它看成是古典式经验型测评的话，那么现代式科学型的测评，则是上世纪末本世纪初在西方产生的事。无论哪种测评，显示出一种教育规律，真正意义上的教育，须臾离不开教育测评；反过来说，没有教育测评的教育就不成其为教育。因此，历史上多少学者、专家和教师都倾注过不少精力来研究测评的科学化，力求使之更准确、更简捷、更民主、更有力度。远的不说，美国人赖斯、桑代克、泰勒、布卢姆这些有代表性的著名研究者作出的贡献是功不可没的。本世纪以来，国人俞子夷、查良钊、薛鸿志、陈选善、王书林等先生的研究和科普工作，也是令人敬佩的。差不多一个世纪的时光，人们逐渐深化了自己的认识，理解到教育测评像筛选机，能够帮助甄别良莠，遴选优品；教育测评像检测器，能够帮助发现问题，及时调控；教育测评像导航仪，能够帮助找准方向，驶近目标；教育测评还像天平，使客体得到公正的衡量，使赏罚分明。一个功能齐全的事物，总是特别地为人们所青睐和器重。功能完善的现代教育

测评已经被人们用到方方面面，参与到教育的各个领域和范畴，发挥其显著作用。记得有一次我到一所学校访问，听到许多对甲、乙两班看法迥异的意见，校长也莫衷一是，我建议他们搞一次比较周密的评估，结果，不仅摸清了情况，发现了问题，而且找到了症结，统一了认识，大家都心悦诚服，皆大欢喜。

好事多磨，1936年，前苏联对“儿童学”的批判扩大化，殃及池鱼，教育测量与评价的研究成了禁区和“冻土带”。1949年后我国教育纳入前苏联体系，国人因此对这一领域长期不敢问津。一直到了1976年后，中国教育理论工作者才又开始了这方面的研究，并迅速掀起了现代教育测评的学习和应用的热潮。其中，难能可贵的是一批青年后生热忱地投入了这场战斗，有的积极介绍国外信息，开展科普工作；有的进行实地研究，著书立说。作者罗黎辉副教授就是这场战斗中的一员骁将。他早于1983年在华东师范大学教育系攻读硕士学位时，就同伙伴一道开始了对教育目标分类理论的研究，取得可喜的成果。来云南师范大学教育系执教后，在繁忙的系主任工作的同时，坚持教授和研究教育测评这门学科。皇天不负苦心，经过近十年的刻苦钻研，与高翔老师合作，向读者捧出这本《教育测量与评价》力作。出书前他送稿子来，我看了之后很有感触，遂欣然应允，为之作序。

作为一本好教材，首先应当具有创新性特征。这本书在荟萃中外古今教育测评理论精华的基础上，围绕我国教育方针和教育目标，结合教育实践，探讨了一些新问题，突出了一些诸如“学生个体发展”、“办学效果与效益”等方面的论述

和应用。既不囿于一孔之见，又不“言必称希腊”，这是很难得的。

作为一本好教材，还应当便于教，更要便于学。这本书遵循编教科书的要求，各个章节沿着概念——原理——应用这条线路展开，思路明晰，内容翔实，文字流畅，详略得当，既不空泛，又不臃肿，教起来顺手，学起来方便，这也是可贵的。

时代不同了，科学的春天来到了我们身边。如果说 50 年代、60 年代、70 年代，中青年出书还是一个梦的话，那么，80 年代、90 年代就是圆梦的好时光。我衷心祝愿我的朋友、我的同仁，特别是我曾经工作过的云南师大教育系和云南省教育科学研究所一大批中青年教师，抓住机遇，脱颖而出，在充满希望的教育科学田野上，挑起重担，冲出高原，走向全国，迎接 21 世纪教育科学将在世界东方掀起的热潮！

冯春林

1995 年 12 月 25 日

# 前 言

教育测量与评价是 20 世纪教育科学领域中发展最为跌宕，而又最显生命力的学科之一。发展至今，它的作用和意义已远远超出对学生学习水平的衡量与判断。它开拓了教育工作者的视野，提供了研究教育、改革教育的新理论与新方法，使人们对教育过程获得了新的认识；使课程设计、教学模式与方法的研究获得突破性的进展；使教育管理从经验型走向了科学化，改变了教育管理的面貌……，所有这些，集中起来，促进了一所学校乃至一个国家总体教育质量的提高。教育测量与评价不仅检验质量，而且创造质量，被喻为提高教育质量的“杠杆”，受到各个国家的广泛重视，成为当前国际教育研究的三大热点领域之一。

在我国，质量作为深层次的问题，已经成为教育改革与发展的重心。全面贯彻教育方针，全面提高教育质量，客观要求在全各级各类教育中积极开展教育测量与评价工作。然而，由于种种原因，目前教育界，无论是教师，还是管理者，真正掌握教育测量与评价理论及方法者甚少，不少教育工作者，即使是教育测量与评价的实践者，对它也不甚明

了。这方面的专著也不易买到。为使广大教育工作者掌握和研究教育测量与评价的理论、方法和技术，以利于指导实践，提高教育质量，我们编写了这本拙著。

本书力图不囿于一说，洋为中用，厚近薄远，广泛吸收国内外新的研究成果，突出应用性理论和理论的应用。在此基础上，对教育测量与评价的类型、原理、方法、技术、手段及历史沿革作了系统的阐述，并针对基础教育中主要的测评对象进行了测评变量、测评方法及测评心理的系统分析。

我们衷心希望本书对推动教育测量与评价的实践能起到积极作用。但是，由于时间和作者水平所限，虽“焚膏油以继晷”，错误疏漏之处仍在所难免，祈望同行、学友不吝赐教。

承蒙沙毓英教授捉暇审阅书稿，冯春林教授为本书泼墨作序，李红女士皱袖绘图，以及许多同仁为本书作了默默工作，谨此致谢。

作 者

1995年11月6日

# 目 录

## 序 前 言

<b>第一章 教育测量与评价概述</b>	1
第一节 什么是教育测量与评价	1
第二节 古代的教育测量与评价	2
第三节 现代教育测量与评价在西方的兴起与发展	8
第四节 我国现代的教育测量与评价	21
第五节 教育测量与评价分类	25
第六节 教育测量与评价的功能、作用与目的	32
<b>第二章 教学目标分类理论与应用</b>	35
第一节 认知领域的教育目标分类理论	36
第二节 情感领域的教育目标分类理论	44
第三节 动作技能领域的教育目标分类理论	49
第四节 教育目标分类理论在教育测评中的应用	63



---

<b>第三章 教师自编学力测试</b>	67
第一节 学力测试的基本方式	68
第二节 测试计划和测试目标	72
第三节 试题类型	78
第四节 试题质量分析	90
第五节 组卷与试卷质量分析	97
<b>第四章 学力的标准化测试</b>	101
第一节 标准化测试概述	101
第二节 学力标准化测试的一般程序	104
第三节 标准化测试的选用	117
<b>第五章 学力评价</b>	120
第一节 终结性学力评价	120
第二节 形成性学力评价	132
第三节 诊断性学力评价	138
<b>第六章 分项测评的原理与方法</b>	145
第一节 分项测评法概说	145
第二节 指标内容与指标结构的制定	147
第三节 指标权重的制定	163
第四节 指标评语的制定	170
第五节 综合评判的数学模型	174
<b>第七章 学生个体发展综合测评</b>	179
第一节 学生思想品德测评	180

---

第二节	学生体质测评	192
第三节	学生审美测评	202
第四节	学生劳动技能测评	203
第五节	学生个体发展综合评价	204
<b>第八章</b>	<b>教师教学工作测评</b>	<b>211</b>
第一节	教师教学工作测评概述	211
第二节	影响教师教学工作质量的变量分析	223
第三节	教师教学工作测评应注意的问题	237
<b>第九章</b>	<b>普通学校测评</b>	<b>241</b>
第一节	普通学校测评概述	241
第二节	学校办学人员测评	247
第三节	学校办学物力条件的测评	256
第四节	学校办学过程测评	259
第五节	学校办学效果与效益测评	269
<b>第十章</b>	<b>教育测评心理与调适</b>	<b>274</b>
第一节	教育测评心理概述	274
第二节	测评者心理与调适	276
第三节	被测评者心理和诱导	284
<b>附 录</b>		<b>291</b>
一、	统计用表	291
二、	主要参考书目	298

## 第一章

# 教育测量与评价概述

## 第一节 什么是教育测量与评价

教育测量是以一定的方式将被测教育客体与一选定的参照系进行比较、衡量，以确定客体被测属性的量是多少的活动过程。

教育评价则是以一定的价值准则对系统收集起来的关于某教育客体的各方面信息进行分析、综合，对客体做出价值判断的活动过程。

二者相互区别。首先，教育测量关心的是量，关心的是量的获得，关心的是量的可靠性、准确性、客观性；教育评价关心的是质，关心的是所获得事实及其量的价值，关心的是质的联系与区别、质的意义；其次，虽然教育测量是教育评价的基础，但并非一切教育评价都以教育测量为基础，有的教育评价，如主观性教育评价便不在测量的基础上进行；第三，教育测量排斥一切离开某一参照系的主观意见，而有的教育评价则恰恰不可避免地依赖于评价者的主观经验；第四，教育测量活动相对比较单一，教育评价活动则是一种综合性较强的活动。

二者又互相联系。形式上，教育测量过程与教育评价过程紧

密相联，有时甚至相互交织在一起；实质上，教育测量是教育评价的基础和手段，多数教育评价往往在教育测量的基础上进行，评价是测量的目的，离开这一目的任何教育测量是无意义的。

## 第二节 古代的教育测量与评价

与许多人类文明一样，教育测量与评价的产生是一种多源头现象，它不仅仅产生于一个国家，在许多国家都可以找到它的源头。

### 一、中国古代的教育测量与评价

据可考证的材料来看，教育测量与评价发生的最早源头在中国。人类迄今发现的最早的教育专著：《礼记·学记》便记载了中国西周（公元前1100—前771年）奴隶制社会“国学”中的一种古朴的教育评价制度：“比年入学，中年考校，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，强立而不反，谓之大成。”意思是：学生入学，隔一年进行一次考试。第一年考查学生能否分析经义和章句，辨别并确定自己的志愿。第三年考查他们能否专心致志地从事学业，与大家团结友爱。第五年考查他们学习是否广博，对师长是否亲善。第七年考查他们对所学的知识能否形成自己的观点和看法，能否选择良好品德的人交朋友。达到标准，称之为“小成”。到了第九年，则考查他们推理论事，触类旁通的本领，以及是否具有坚定不移的志向，不再违背师长的教导。如果达到了标准，称之为“大成”。从这段记述中可以看出，此时的学校教育不仅有了对学生的学力评价，有了对学生思想品德的评价，而且两方面的评价是结合在一起进行的；不仅对评价内容做了系统的

安排，而且形成了定时进行评价的思想和制度。（可见，早在西周奴隶制社会，一套教育评价制度在中国已经确立起来。）

（进入封建社会后，教育测量与评价在制度、方式与方法上推陈出新。入汉，武帝创立太学，太学既是一个最高学府，又是一个国家考试机关，它汲取了西周教育评价制度的精华，并加以发展，实行了“一岁皆辄课”，即一年考试一次的制度。考试的形式除了口试外，已经使用笔试。考试方法采用“设科射策”，就是事先命多道试题，抄录在竹简上，根据试题的难度分为甲乙两个档次，由考生随机抽取作答。至东汉桓帝时对此又进行改造，取消甲乙科，代之以以上、中、下三等，并将课试改为“二岁一试”制。汉代的教育评价在我国古代教育测量与评价的思想、方法、制度发展上有着重要的地位。）

（唐代统治者重视教育，教育评价也受到重视，得到进一步发展。唐初，官学与私学大发展，学校成为人才的重要来源。为保障人才质量，国家以法令规定，学官当先重考试，并把学生考试成绩作为考勤学官的重要标准；学校实行定时考试制度，按旬、月、季、岁进行考试。）旬试：一个月分三旬，一旬为十天，每旬进行一次考试，考试一旬来所讲习的内容，考试方式为帖经及问义。月试：每月第三旬末，考试一月来所讲习的内容。（一开始，学校旬试与月试并举，使学生与博士成天围着考试转，十分紧张，后来放弃了旬试。元和元年（公元806年）规定国学“每月一度试”，月试遂成为一种固定的考试。）季试：每季度将终时，对一季度学业情况进行的考试。季试比月试意义更大，因此，有的学校的考试由国家官吏来主持进行。岁试：年终前对学生全年学业进行的总考试，考试方法主要是问义，衡量标准分为上、中、下三等。《新唐书·选举志》载：“岁终，通一之业，口问大义十条，通八条为上，六为中，五为下。”岁试的目的不仅是对一

年学业水平的总甄别，考试成绩还将作为升留和奖惩的依据。因此，岁试通常都是由国家的主管官吏亲自主持。由此可见，唐代的绩效评价，不仅在制度方面有所发展，形成了旬、月、季、岁相结合的定时考试制度，而且在范围方面也有所扩展，把学业测评的结果与对教师的“考勤”联系起来，开始出现了对教师教学的评价；绩效评价不仅服务于教学的温故知新，提高学习质量，而且服务于教学中的良莠甄别，加强教学管理，为我国的教育测量与评价史书下了有意义的一页。

唐宋以后，由于政治的需要，统治者越来越重视科举，学校为其左右，日益成为科举的附庸，学校教育的教学内容与方法，必须适应科举的需要，考试制度更是如此，尤其考试方法，几乎完全依照科举考试。尽管如此，在1300年的时间里，学校的考试制度还是取得了一定的改进和发展。特别要提到的是宋朝学校考试中出现的积分法，以及至元、明时期进一步发展形成的积分制：明朝国子监规定，在高年级的孟、仲、季月分别进行一次考试，每次试卷分三等，文理双优者得一分，理优文劣者得半分，文理俱劣者不得分，一年内积分达到八分者为及格。开始使用积分来计量学业成绩，并以积分形式确立及格概念，这是我国教育从单纯定性评价走向定量测量与评价的开始，在我国古代教育测量与评价的发展上无疑是一个重要的进步。

在中国，除了教育测量与评价在学校体系中的自身发展以外，不同历史时期的各种选拔人才制度的推行，对教育测量与评价有着特殊的影响，客观上促进了我国古代教育测量与评价的发展，并对欧、亚许多国家产生过较大影响。

当我国还处在原始社会末期的尧舜时代（约公元前2200-前2140年），在人才的选拔任用上就有了评价的因素。据载，当时执法官皋陶提出选才任用的九条标准，称“九德”。根据这九条标

准来选拔部落联盟各首领。而且，每任满三年便对他们的任职进行一次考核，根据“三载考绩”，“黜陟幽明”，拔贤能，贬庸禄。

西周时期，低层官位实行“乡举里选”制，选拔人才的标准为六德、六行和六艺。六德，即知、仁、圣、义、忠、和。六行，即孝、友、睦、姻、任、恤。六艺，即礼、乐、射、御、书、数。并有“诸侯月试其国，州伯时试其部，四试而一考。天子岁试天下，三试而一考；前后三考而黜陟”<sup>①</sup>的做法。

汉代取士的主要方式为征辟和察举。征辟是皇帝或官府直接聘任官员的选士办法。察举则是由丞相、列侯、刺史、守相等官员根据考察，把认为品德高尚，才能出众的贤人推荐给朝廷，经过考试，任以官职的办法。公元前165年，汉文帝出题考试，“举贤良方正能直言极谏者”授予官职。试题写在竹简上，称为“策题”，分四个策目：“朕之不德、吏之不平、政之不宣、民之不宁”<sup>②</sup>。皇帝以策题考士称为“策问”，应试者在竹简上作答，称为“对策”。“策问”与“对策”既是我国取士考试的开端，也是世界笔试考试的开端，对教育测量方法的发展具有重要的历史意义。

此外，汉代董仲舒还最早使用“考试”一词，后来《三国志·魏王昶传》把“考试”训为准绳，“考试犹准绳也，未有舍准绳而意正曲直，废黜陟而空论能否也”，从而从功能上对“考试”概念作出了界定：考试乃是衡量政绩所必需的一把尺子。这一界定虽然与现在的考试一词在内涵方面有很大差别，却是今天考试概念的来源。

到魏晋南北朝时期，取士以“九品中正”制为主。这种制度将

①《春秋繁露》第二十一。

②辞海编辑委员会：《辞海·教育、心理分册》，上海辞书出版社1980年版，第61页。

统治者分为九个等级，即九品。形式上，选拔工作由中正官负责，小中正官品第者，送大中正官核实后送司徒，司徒再核，然后由副尚书选用。但实际上，“九品中正”制是一种按门第高低取士的制度，前三品仅限由世族豪门担任，称为上品；四品以下为下品，从寒门选出，下品不能升入上品。这种世袭性质的选士制度，造成了“上品无寒门，下品无世族”，酒囊饭袋，身居要职的局面，势在必改。当然，这一时期的取士也不尽然都是实行“九品中正”制，北魏曹丕黄初五年（公元 224 年）制定了“五经课试”法，规定每两年考试一经，五次考试都及第者任官。晋代曾实行察举，但孝廉、秀才一律以考试选拔，孝廉试经，秀才试策，这种选士考试对隋唐科举制产生了直接的影响作用。

隋唐时期，我国的选士制度，经过两汉、三国、魏晋南北朝的历史积淀，到隋炀帝大业二年（公元 606 年）发展成为科举制，完成了从杂乱的考试现象到规范化考试的过渡，从考绩与学力衡量相分离或偶尔结合到相结合的过渡，形成了一整套稳定的制度。

科举制即分科考试，根据成绩选拔人才的制度。炀帝始置进士科。唐代在进士科基础上增置秀才、明法、明书、明算、一史、三史、开元礼、童子、道举等科，后武则天亲行殿试，又增设武举科…… 唐代先后共设有 50 余科。实行分科考试，并用制度加以保障，这在考试史上有着重要的意义，对教育测量与评价来说，是一种积极的启进。

科举考试采用了两种考试方法：口试与笔试；采用了帖经、墨义、策论和诗赋等题型。尤其必须提到是在考试的管理方法上，武则天为科举考试的试卷管理创造了“弥封”的方法，即将答卷上应试者的名字装订起来，使阅卷者不知道答卷者为何人，以达到防止作弊的目的。这一方法沿用至今。



科举考试是一种分级考试。唐朝科举一般分县试与礼部试两级进行，宋代增设殿试，明清时增至四级：童试——乡试前的预选性考试，在各府进行，及格者为秀才；乡试——基本上在行省进行，秀才方有资格参加，及格者为举人，举人为乙级功名，有资格入仕，可做知县或任教职，也可赴吏部参加铨选；会试——在乡试的第二年进行，三年一次，属殿试的预备性考试，举人参加，及格者为贡士；殿试——在国都举行，贡士有资格参加，以皇帝亲试的名义举行，由八位读卷大臣代替皇帝评卷，按五等确定成绩，根据八位大臣阅卷评判的成绩，排列名次，及格者为进士，前三名分别冠之以状元、榜眼和探花，分别授予一定的官职。

科举考试制度，始于隋，兴盛、完备于唐，废于清末，历时1300年，是我国实行时间最长的一种选士制度，对我国乃至对世界的考试制度有重要的影响。孙中山先生说：“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。”<sup>①</sup>

〔纵观两千多年的选士制度史，不难发现它对教育测量与评价所产生的多方面影响。它不仅首开了笔试之先河，独创了帖经、墨义、策问与诗赋等题型，而且创造了开卷考、闭卷考、“糊名”、“弥封”、“誊录”、“墨卷”与“朱卷”、“磨勘”、复试、分科考试、分级考试、全国统一考试等方法；不仅在考试内容、考试时间与考试程序方面，不同朝代形成不同的定制，而且在命题、考试、阅卷等方面形成了一整套制度，对至今为止的教育测量与评价都不无影响。〕

<sup>①</sup>盛奇秀：《中国古代考试制度》，山东教育出版社1988年版，第128页。