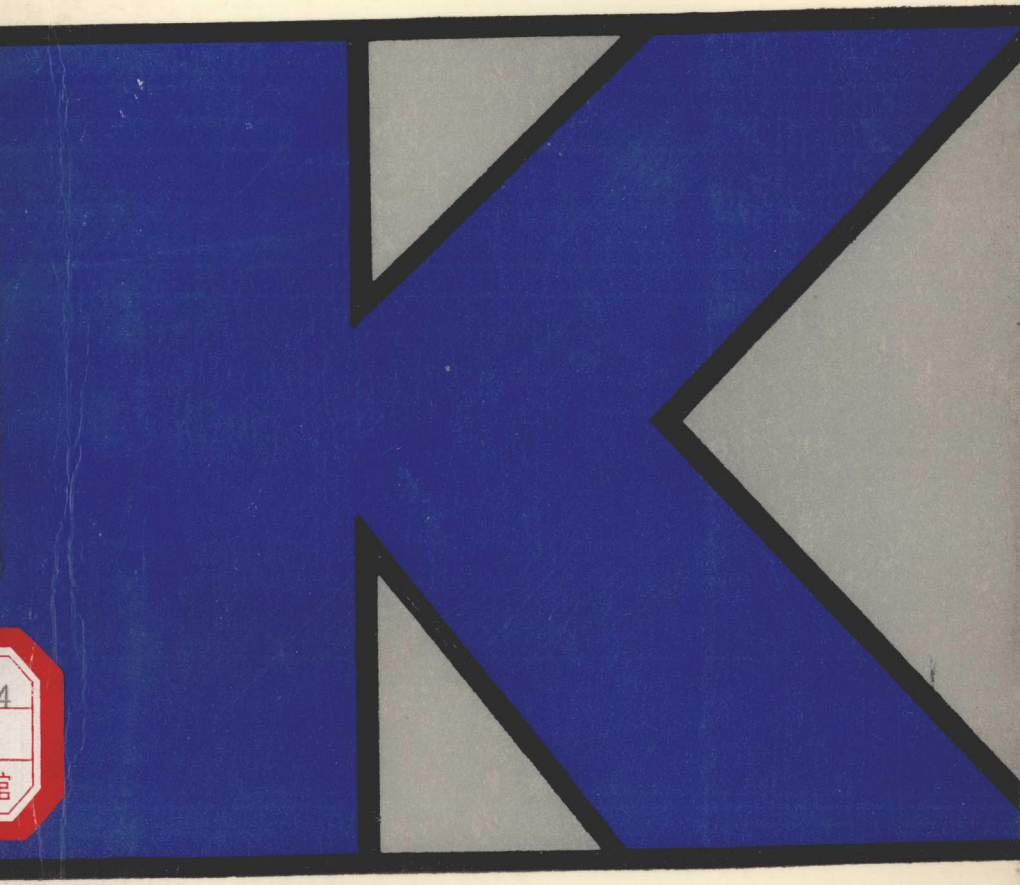
考试的教育测量学基础

国家教育委员会考试管理中心　主编

郑日昌　漆书清　马世晔　编

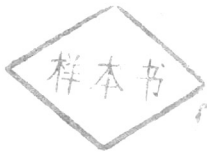


高等教育出版社

考试的教育测量学基础

国家教育委员会考试管理中心主编

郑日昌　漆书清　马世晔编



高等教育出版社

内 容 简 介

本书是国家教育委员会考试管理中心为考试管理干部编写的培训教材，也是一本向各级教育行政干部和教师宣传考试科学化、标准化知识的普及性读物。本书较为系统地介绍了教育测量学的基础知识和基本原理。全书共8章，内容包括：概述，测验的编制，测验的题目分析，测量的误差，测验的质量评估，测验分数的整理，测验分数的解释与应用，题目反应理论。

考试的教育测量学基础

国家教育委员会考试管理中心 主编

郑日昌 漆书清 马世晔编

*

高等教育出版社出版

高等教育出版社照排中心照排

新华书店总店北京科技发行所发行

国防工业出版社印刷厂印装

*

开本 850×1168 1/32 印张 7 字数 180 000

1990年 8 月第 1 版 1990年 8 月第 1 次印刷

印数 0001—5 010

ISBN 7-04-003220-1/G·196

定价 1.95 元

考试管理干部培训教材

编审委员会

前　言

中国是考试的故乡。科举是中国对人类文明的一个重大贡献。然而，长期以来，考试与教育、教学的关系却总是困绕着我们，以至出现了片面追求升学率，以取消高考作为“文化大革命”“突破口”之类令人痛心的事情。

1985年，中共中央发出《关于教育体制改革的决定》后，国家教育委员会把考试改革作为教育体制改革的重要问题提到议事日程上，并成立了考试管理中心，统一管理高考等国家教育考试。经国家教育委员会批准，上海进行了普通高中毕业会考试验，广东进行了高考标准化试验，取得明显效果。1989年，国家教育委员会决定在全国各省、自治区、直辖市试行普通高中毕业会考制度，改革高考科目设置与录取方法、高考逐步实现标准化，成人高校招生也试行“资格生”，要求一年考试，三年有效。在短短几年内，中国教育考试从制度、内容，到形式、手段等多方面都开始进行了重要改革。

改革是顺利的，也遇到困难。突出困难是缺乏理论指导，考试工作人员缺乏训练。这里说的理论，不是国外现成的教育统计学、教育测量学等，而是适合当今中国国情的、经过中国考试实践检验，证明是正确的理论，这个理论还应当是全面的，应揭示考试与教育、教学、劳动、人事的关系，阐明命题、考试实施、统计分析等各环节的规律。如上所述，继承我国优秀历史遗产，借鉴国外有益于中国的作法，吸收我国当今丰富实践的成功经验，创造出这样的理论，用这样的理论，在不太长的时间内，通过分批培训，武装考试人员——这是今天考试改革面临的突出问题。可以说，这个问题不解决，考试改革就不能深入，甚至会夭折。正是为了这个目的，我们才下决心编教材，办培训班。这一工作也得到了国家教委主管领导同志和人事司、成人教育司及高

等学校学生管理司的支持。

这套教材包括:《考试的统计分析方法》、《考试的教育测量学基础》、《标准化考试》、《计算机在考试管理中的应用》,以后还准备编写中国考试史、外国考试比较等书。这是适应考试工作人员目前实际水平的培训教材,作为中国考试理论的建设,当然还应有高水平的专著。

我们希望,通过这套教材的学习,使考试工作人员理解教育统计学,教育测量学关于考试的基本知识,掌握考试工作各环节的要领,了解或掌握以计算机为核心的现代化手段,明确传统考试为什么要改革,以及改革的方向,把考试改革深入进行下去。

在编写过程中,我们努力使这套教材应具有科学性、系统性及实用性,并能体现几年来研究与试验工作的成果,同时在表达上力求深入浅出,以适合广大从事具体考试工作同志的需要。目前,这套教材还不成熟,曾只在京、沪两市和广东、江西、吉林三省的培训班上试讲了一轮,作了一些修改补充,应培训急需决定公开出版,不过,我们把教材的编写作为一个过程。希望这套教材能提高学员的理论水平,在提高的基础上,学员能对教材提出修改意见,把国内外考试的新鲜经验充实到教材中去,逐步编写出一套适合中国考试实践需要具有中国特色的高水平的教材。通过这样从实践到理论,又从理论到实践的不断提高,弘扬中华考试之文化,振兴中国考试之事业,使中国的考试,无论从理论上还是实践上,都走到世界之前列。

考试与教学是实现同一教育目标的相辅相成的两个方面。我们也希望广大的教育管理干部、教学研究人员、教师能喜欢这套书,并给我们提出宝贵意见。

《考试的教育测量学基础》较为系统地介绍了教育测量学的基础知识和基本原理,为考试科学化提供理论依据。本书的第一、三章由北京师范大学郑日昌同志编写,第二、六、八章由江西师范大学漆书清同志编写,第七章由国家教育委员会考试管理中心

马世晔同志编写。第四、五章由郑日昌、马世晔共同编写。最后由郑日昌同志负责全书的统稿。本书承我国著名心理学家北京师范学院林传鼎教授，中国教育学会教育统计与测量研究会理事长北京师范大学张厚粲教授审阅，并提出了宝贵的意见，在此深表谢忱。

由于编者水平有限，加之编写时间紧迫，书中不当和错误之处在所难免，恳请广大读者批评指正。

编　者

1990年3月于北京

目　录

第一章 概 述

社会发展离不开教育，教育效果的评估离不开测量。

测量要准确可靠，必须采用科学的方法，必须以系统的理论作指导，于是诞生了教育测量学。

考试是对学习结果的测量，不但是教育工作的一个重要环节，而且在社会上具有举足轻重的作用。在学校里有摸底考试、单元考试、期中考试、期末考试、升级考试、毕业考试、升学考试等等；在社会上有招工考试、晋级考试、自学考试、出国考试、证书考试等等，五花八门，不一而足。考试的触角伸向各行各业，不但关系到社会的发展，而且决定着许多人的命运。

考试的作用如此之大，但由于缺乏科学的理论指导，传统的考试方法弊端甚多，带来了许多不良后果。为了更好地发挥考试的效能，必须以教育测量学为基础，走考试科学化之路。

本书的宗旨在于系统阐述教育测量学的基本原理、基本概念和基本知识，为考试改革提供理论依据。

教育测量学包括的范围很广，能力测量、人格测量、学绩测量均属教育测量的范畴。本书只讨论与学绩测量——考试有关的测量学问题。

第一节 教育测量的历史

要研究教育测量，不可不考察它的发生发展的历史，否则就不能洞察过去，理解现在，展望未来。

一、教育测量在我国的历史与现状

教育测量起源于中国，这是举世公认的。

在我国最早的教育专著《学记》中记载了距今三千多年前西周时期的教育制度和考试制度，分年论述了对学生各种能力与

个性特征考察的情况。

“古之教者,家有塾,党有庠,术有序,国有学。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,强立而不返,谓之大成。”这段话的大意是说:古代的教育制度,20 户人家设一私塾,500 户的县设一学堂,12500 户的行政区设学校,国都设大学。大学每年招收学生,每隔一年考查一次。第一年考查学生分析课文的能力和志趣;第三年考查学生的专业思想是否巩固,同学之间能否相亲相助;第五年考查学生的知识是否广博,对教师是否敬爱;第七年考查学生研究学问的本领和识别朋友的能力,合格的就叫做“小成”。到第九年,学生对于学业已能触类旁通,他们的见解行动已能坚定不移,这就叫做“大成”。

又据《礼记·射义》记载:“古者,天子以射选诸侯、卿、大夫、士”。具体说来,就是天子用试射选拔人才,看其行为是否合乎礼仪,动作是否合乎乐律,射中的次数有多少等,据此择优录用。

春秋时期,我国著名的教育家孔子曾说过:“中人以上,可以语上也;中人以下,不可语上也。”“唯上智与下愚不移”。从孔子的这番话我们不难看出,当时他就根据观察来评定学生,把学生分为了上、中、下三等,这正是现代教育测量理论的萌芽,实际上相当于测量学里的顺序量表。战国时期的孟子曾说过:“权,然后知轻重;度,然后知长短。物皆然,心为甚”。这就明确指出了心理现象与物理现象一样具有可测量性。

汉朝的大学,对考试方法有很大发展,分为“口试”、“策试”、“射策”三种。射策,就是事先出好多道试题并分别抄录下来,由学生随机抽取作答。

三国时期刘劭在《人物志》中提出:“观其感变,以审常度。”意思是根据一个人的行为变化就可以推测他的一般心理特点。并且还提出,可以根据言辞来观察人的智力等。这部书在 1937

年被译成英文在美国发表了，书名叫《人类能力的研究》。

隋炀帝大业二年（公元606年），创立科举制度，通过分科考试，选拔任命官员。至唐朝科举制逐步完善，经宋、元、明定型，到清光绪三十一年（公元1905年）废除，科举制在我国整整延续了一千三百年。

科举的帖经、墨义、策论、诗赋等，是当代填空、简答、论述、作文等题型的源流。科举考试历经几朝，制度已相当完备，考场规定也愈加严格，如考试完毕后，要“弥封”、“易书”等。所谓“弥封”即在考后将考生的名字糊上，以防考官认出，待评卷结束定出等级后再行开视。所谓“易书”则是在弥封后不马上直接评定，而是送到誊录院，由专人抄录成副本，考官根据副本来评定成绩。

科举制度在教育测量史上具有重要地位。法国大革命时期的资产阶级启蒙思想家伏尔泰（Voltaire）曾经对中国的科举制度备加赞扬：“人类精神，肯定想象不出比这样的政府更好的政府。在这个政府里，重要的衙门彼此统属，任何事情都在那里决定，而其成员，都是先经过几场严格的考试的。”欧美各国采用考试的方法选拔官吏，是十八世纪末、十九世纪初从我国学去的。

教育测量虽发端于我国，但我国古代的考试还称不上是科学的教育测量。辛亥革命后，盛行于西方的心理与教育测验开始传入我国。

我国近代的测验运动大约始于1914年前后，当时有人在广东对500名儿童的理解和记忆进行了测验。1918年俞子夷编制了“小学国文毛笔书法量表”，这是我国最早的教育测验。1920年廖世承和陈鹤琴在北京高等师范学校和南京高等师范学校分别开设了测验课，并用测验试测投考该校的学生，1921年他们的《心理测验法》正式出版。1922年，中国教育改进社聘请美国测验专家麦柯尔（W · A · Mc Call）来中国讲学，在他的指导下，北京师范大学、北京大学、燕京大学、清华大学、东南大学的师生

编制了 40 多种测验。据麦柯尔说:当时中国的各种测验几乎达到了美国的水平。1923 年在教育改进社支持下,对全国 22 个城市和 11 个乡镇的 92000 名小学生进行了一次大规模的测验,引起了当时教育界的注意。1931 年中国测验学会成立,并于第二年创办了名为《测验》的杂志,嗣后全国各大学教育系和中等师范学校相继开设了教育测量学课程。1938 年当时的教育部曾宣布师范院校实行统一招生考试,考试科目、命题标准及录取标准一律由招生委员会规定,考试分为笔试和口试两种。由于抗日战争爆发,致使我国教育测量的发展工作中断 。

解放以后,虽然在 1952 年成立了全国招生委员会,实行了全国高等学校的统一招生考试。但直到 1979 年心理测验正式在我国恢复地位之前,三十多年来学校里停开了测量学课,对于测量学的应用研究更是无人问津。

党的十一届三中全会后,测验才在科学的春天里复苏,全国各高等师范院校的心理系、教育系相继开设了心理和教育测量课,此后全国的测验队伍逐渐扩大,研究也逐渐展开。1980 年以来,北京师范大学率先运用教育测量学的理论对高考试卷进行了一系列科学分析,全面考察了高考命题、施测、评分、录取各个环节,对我国高等学校的入学考试改革提出了建设性的意见。1985 年,国家教育委员会正式成立了考试管理中心,并先后在广东、上海两地进行了高考标准化和高中毕业会考的试验。近年来,围绕高考和各级各类学校考试中存在的问题,全国各地的教育统计、教育测量工作者,在高考、会 考、自学考试以及医学院校统考等方面进行了大量研究,取得了可喜的成果,出版了一批有关考试和教育测量方面的专著。

二、教育测量在西方的产生与发展

“工业革命”成功后,为了满足社会对人才的需求,欧美资产阶级学者对考试进行了积极的研究和改革。

1702 年，英国的剑桥大学开始采用笔试。

1791 年，法国建立了文官考试制度。

1845 年，美国波士顿市第一次进行了全城范围的书面考试，从而拉开了现代标准化测验运动的序幕。同年，美国大教育家迈恩 (Horace Mann)在论述口试与笔试的利弊时提出:1)划一的笔试比个别的口试公平些，因为应试儿童可接受同样的试题，不至有难易的不均。2)划一的笔试比个别的口试可靠些，因为笔试题多，受偶然因素影响小。3)划一的笔试比个别的口试在时间上经济些。4)口试容易引起临场的慌乱。这些论述虽然对于测验运动的发展没有产生太大影响，但与后来标准化测验的观念极为吻合。

1864 年，美国有一位名叫费舍 (George Fisher)的教师，曾广泛收集学生的书法、拼字、算术、文法、作文、历史、自然、图画、法文等作业样本，编成量表，作为评量各科成绩的标准。书中备有各科学生作品的不同水平的样本，并为每一样本评定一种分数，以示优劣。在评定某学生某科作品时，可将其作品与量表集中的各样本互相比较，以求得与该学生作品优劣相等的样本，此样本的分数即为该学生应得的分数。由于在评定分数时有标准可循，因而不致漫无依据。费舍编制量表集时仅凭个人的主观判断来评定样本的分数，在手续上当然不客观、不精密，但是他采用的方法与后来的书法量表、作文量表编造的方法大体是相同的，可惜的是费舍的量表集没有得到当时教育界的注意，因此对于教育测验的发展并没有很大影响。

19 世纪末，美国教育界有一场争论。当时有人主张对学校里只注重练习与背诵的教学方法进行改革，增加实用学科，遭到守旧派的反对。后者认为新的功课一加入，学生就没有功夫学习旧有的基本科目了。1894 年，来斯 (J · M · Rice)选定 50 个字作为拼法测验，测量各校学生的拼字能力，并调查各校每周讲授拼法的时数。结果表明，讲授时间的多少与成绩优劣没有多大关系。8 年之中每天用 15 分钟学习拼法的学生，其成绩并不次于每天

用40分钟学习拼法的学生。来斯的工作虽然受到许多人的怀疑，但也引起了少数有思想的教育家的注意和赞同。他采用客观方法来研究教育问题，对测验运动的贡献是不可磨灭的。

来斯并没有编造过现在所谓的标准化测验。教育测验运动的中心人物当推美国心理学家桑代克(E · L · Thorndike)。爱恩斯(L · P · Ayres)曾说过："我们既称来斯为教育测验的创始者，则对于桑代克应称之为教育测验运动的鼻祖"。1904年，桑代克出版了《心理与社会测量》一书，介绍了心理统计方法及编造测验的基本原理。这是世界上第一本社会科学方面的测量学专著。1908年，在桑代克指导下，斯东(C · W · Stune)编造一个算术推理测验，这是一种最早的标准化测验。1909年，桑代克发表书法量表，这是世界上第一个用科学方法制成的教育测量工具，是测验运动中极重要的事件。自此以后，各种标准化测验和量表日渐增多，由单科测验发展到成套的一般学绩测验；由常模参照测验发展到标准参照测验；由小学扩展到中学、大学；由用于调查和选人发展到用于诊断和促进教学；由对知识能力的测量扩展到对学习态度、兴趣以及品德、性格等方面的测量；由单纯对学生学习成绩的评估，发展到对课程设置、教材、教改方案以及教师、学校乃至一个地区和国家教育效果的评估。

时至今日，欧美一些发达国家，不仅在学校教育工作中大量使用测验，而且在军队、政府部门、工业、商业、交通、体育等行业，以及律师、医生、警察、会计、理发等职业，也都使用测验来选拔人才、分配工作和鉴定工作效率。

目前，世界上出现了许多从事教育测量研究与测验编制的专业机构，如美国教育测验中心(ETS)、美国大学测验中心(ACT)、英国伦敦职业考试中心、日本大学入学考试国家中心以及香港考试局等。其中美国教育测验中心是目前世界上最大的测验研制机构，总部设在新泽西州普林斯顿，占地400多公顷，有近3000名工作人员，多为心理学、教育学、心理与教育测量学、统计