



新编21世纪教育学系列教材

教育测量学

Educational Measurement

主 编 朱德全



新编21世纪教育学系列教材

教育测量学

Educational Measurement

主 编 朱德全
副主编 林克松 杨 鸿

中国人民大学出版社
· 北京 ·

图书在版编目 (CIP) 数据

教育测量学/朱德全主编. —北京: 中国人民大学出版社, 2016. 1
新编 21 世纪教育学系列教材
ISBN 978-7-300-22059-8

I. ①教… II. ①朱… III. ①教育测量 IV. ①G40-058. 1

中国版本图书馆 CIP 数据核字 (2015) 第 252237 号

新编 21 世纪教育学系列教材

教育测量学

主 编 朱德全

副主编 林克松 杨 鸿

Jiaoyu Celiangxue

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242 (总编室)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京东方圣雅印刷有限公司

规 格 185 mm×260 mm 16 开本

印 张 12

字 数 238 000

邮政编码 100080

010-62511770 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2016 年 1 月第 1 版

印 次 2016 年 1 月第 1 次印刷

定 价 28.00 元



在大数据时代正在来临的今天，大数据已经渗透进各个行业、各个领域。不论是产业实践者、学术研究者，还是行业管理者，人们都需要搜集数据、处理数据，都需要从海量数据中提取有意义的信息或知识，寻找隐藏在海量数据中的规律、趋势。从某种程度上说，大数据正在改变着我们的思维方式、生产方式和生活方式。在教育领域亦如此，大数据正在改变着我们的教育教学思维、教育教学方式和教育教学生活。在这样的背景下，教育领域的每个人都需要掌握搜集数据、处理数据、分析数据的方法，具备相应的知识与能力。教育测量学作为一门研究如何搜集数据、处理数据、分析数据的学科，无疑具有更加重要的作用与更加突出的地位。

然而，在多年的教学科研过程中，我们发现，许多学习者对教育测量存在认识误区：一是觉得只要不从事相关学科的教学或者科研，教育测量对日常教育教学工作的意义不大，不需要进行系统学习；二是认为教育测量属于理科门类，公式繁多、晦涩难懂，不容易熟悉掌握；三是认为教育测量理论体系繁杂冗余，不便于具体操作。这些不当认识的产生，一方面与学习者自身的认识不清有关，另一方面则与我国教育测量学科发展滞后、课程教材建设质量不高有很大关系。

本书是编者在教育测量理论和实践长期探索的基础上撰写而成的，旨在使读者能够比较系统而又清晰地掌握教育测量的基本理论和操作方法，帮助教育领域的每个人掌握搜集数据、处理数据和分析数据的知识与能力，以在大数据时代背景下更好地胜任教育教学工作。基于此，本书在编写上力求凸显三大特色：

(1) 力求凸显结构性和逻辑性。本书无意囊括教育测量学的所有相关知识，而是尽量尝试用简洁明了的框架，呈现教育测量学的核心知识点和能力点，并依照教育测量的理论和实践逻辑，合理排列这些知识点和能力点。本书共分六章：第一、二章是对教育测量学以及教育测量的本体认识；第三章着重介绍编制教育测验的步骤和方法；第四、五章介绍良好教育测验的表征，第四章从具体项目的层面展开分析，第五章从测验质量的层面进行分析；第六章着力于介绍教育测量结果的统计处理。六章内容在编排上依次推进、前后呼应，体现了知识的结构性和思路的逻辑性。

(2) 力求凸显操作性和技术性。教育测量学属于实践性非常强的应用学科，本书着力于体现这种实践性和应用性。除了在内容上注重对实践运用、操作技术的探讨之外，本书还在第四、五、六章引入普及化高、操作简易的 SPSS 统计分析软件，并以案例为载体，用简洁易懂、清晰明了的语言和操作图例完整地呈现 SPSS 在教育测量项目分析、质量分析和结果处理中的操作技术与操作流程。

(3) 力求凸显可读性和规范性。本书在保证科学性、准确性的基础之上，尽量摒弃烦琐、生僻、冗长的论述，追求清晰明了、精练简洁的行文风格，强化话语的可读性和规范性。

本书由朱德全负责设计框架，主要章节由下列同志完成：第一章，杨鸿、朱德全；第二章，林克松、朱德全；第三章，张媛媛、朱德全；第四章，林克松、朱德全；第五章，王俐嘉、朱德全；第六章，吴乐乐、朱德全。全书由朱德全、林克松、杨鸿负责统稿与定稿。

本书既可作为高等院校本、专科以及研究生相关课程的教材，又可作为基础教育管理人员、广大教师以及教育测量爱好者的参考书。

本书的出版得益于中国人民大学出版社的鼎力支持和协助。本书在编写过程中参考、引用了诸多国内外文献资料，在此向出版社和所有作者致以敬意和谢意。由于编者所学有限，书中难免有纰漏之处，恳请同仁和读者批评指正。

编者

2015年7月于西南大学

目 录



第一章 教育测量学概述	1
第一节 教育测量学的历史脉络与当代进展	2
第二节 教育测量学的学科地位与学科贡献	10
第三节 教育测量学的研究范畴与理论体系	15
第二章 教育测量的基本原理	23
第一节 教育测量的基本概念	24
第二节 教育测量的特征与功能	26
第三节 教育测量的要素与种类	30
第四节 教育测量的基本理论	35
第三章 教育测验的编制程序与技术	48
第一节 教育测验的一般编制程序	49
第二节 教育测验的项目编制技术	60
第四章 教育测验的项目分析	73
第一节 难度分析	74
第二节 区分度分析	79
第三节 项目分析的特殊问题	86
第四节 SPSS 在教育测验项目分析中的应用	89
第五章 教育测验的质量评估	106
第一节 测验的误差	107
第二节 测验的信度	110
第三节 测验的效度	121
第四节 SPSS 在教育测验质量评估中的应用	127
第六章 教育测量分数的统计处理	141
第一节 教育测量分数的特征与种类	142
第二节 教育测量分数的整理	146

第三节	教育测量分数的转换	153
第四节	教育测量分数的组合	162
第五节	SPSS 在教育测量分数统计处理中的应用	168
参考文献	180



教育测量学概述

本章提要

本章首先对教育测量学的历史脉络和当代进展进行探讨。教育测量起源于早期学生学业成就和学力水平的检测。教育测量学从早期的萌芽探索发展至今，走过了从无到有、从微不足道到重要学科的发展历程，经历了主观经验化、客观标准化、标准技术化和多元智能化等发展阶段。本章随后对教育测量学的学科地位与学科贡献展开论述。今天，教育测量学有着非常重要的学科地位。它不仅是教育科学体系的重要学科和学校教育教学活动的重要手段，还是教育科学研究的重要工具，更是教育行政部门作出教育科学决策的重要依据。本章最后对教育测量学的研究范畴与理论体系进行研究。与其他教育学科相比，教育测量学有着更为独特的研究范畴与理论体系。它主要应用数量化测定的原理、技术和方法对教育领域的多种对象进行研究，包括测量什么——对象问题、怎么测量——标准及工具问题、测量工具如何——质量问题等范畴，涉及概念、原理、编制和结果处理。

学习目标

- 了解教育测量学的历史发展与特点。
- 说出科举制度在教育测量学发展过程中的重要作用。
- 说出教育测量学客观化、标准化的特征。
- 认识教育测量学的学科地位与学科贡献。
- 掌握教育测量学的研究范畴与理论体系。



教育测量学是教育科学体系中一门重要的学科。从诞生至今，它为教育科学的发展、教育改革的推进作出了重要的贡献。回顾教育测量学的历史脉络和当代进展，探讨教育测量学的学科地位与学科贡献，厘清教育测量学的研究范畴与理论体系，是教育测量学学科建设与发展的重要任务。

第一节 教育测量学的历史脉络与当代进展

现代教育测量起源于古代的考试制度。中国是最早建立考试制度的国家，早在西周奴隶制时期（公元前 1046—前 771 年）就建立了学校教育考试制度，至隋炀帝大业二年（公元 606 年）又兴起“科举取士”考试制度——科举制。中国数千年的考试制度为现代教育测量的产生奠定了坚实的基础。回顾教育测量学的历史脉络，厘清教育测量学的当代进展，是教育测量学学科建设与发展的重要任务。

一、教育测量的萌芽探索：主观经验化

教育测量最早见于早期人们对学生学业成就和学力水平的检测。不论是我国存在约 1 300 年的科举制，还是西方各国广泛采用的口试，早期的这些考试大多凭借主观经验进行，既没有科学的试题编制方法，也没有客观的评分标准。因此，早期人们对教育测量的探索处于主观经验化发展阶段。

我国在西周时期就建立了学校教育制度，设置了学校考试制度，开始了对教育测量的探索。《礼记·学记》记载：“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，强立而不反，谓之大成。”这表明，西周的“国学”已经建立起初步的教育考试制度。它不仅规定了学业考试的时间和步骤，而且规定了学业考试的内容和标准。孟子曾提出：“权，然后知轻重；度，然后知长短；物皆然，心为甚。”这说明教育测量的思想有了早期萌芽。

两汉和魏晋南北朝时期，教育考试在学校教育中的功能愈加稳定，学校教育考试制度基本成型，教育测量在人才选拔中的作用逐步凸显。汉代在考试制度、考试方法和考试功能等方面都取得了重要的进展。在考试制度方面，太学把考试作为考查学生学业和学校管理的一种方法。汉武帝曾制定“岁考制”，“一岁皆辄课”，通一艺者补“文学掌故”，高策者为“郎中”，后改为两年一试。在考试方法方面，开始使用“口试”、“策试”、“射策”三种形式考查学生。“口试”采用问答的形式；“策试”重“章句”和“师法”，首开笔试的先河；“射策”类似于现在的抽签考试。主考者按照题目的难易和范围的大小，将科目



分为甲科、乙科两种。初期将试题写在木牌上，应试者用短矢而射，射中的试题由应试者作答。后来将试题写在帛上或纸上，应试者随意抽取。在考试功能方面，汉代不仅把考试用于督促和检查学生的学习，还将之与太学生的出路联系起来，这也为隋唐时期将学校考试与官职统一奠定了基础。太学生学满两年可以参加考试，以五经为考试内容。通二经者，可任文学掌故；通三经者，授予太子舍人职务；通四经者，授予郎中职务。考取者也可不去做官而继续学习，参加更高一级的考试。甚至已经为官的太毕业生也可参加高一级的考试，考取者可继续升职。

隋唐时期，我国考试制度发生了重大变革，教育测量在人才选拔中的功能得到巩固。隋炀帝大业二年开始设置进士科，这标志着我国科举制的正式兴起。科举制是指普通士人通过逐级的分科考试，朝廷按照考试成绩选取人才，授以相应官职的一种制度。科举制肇始于隋唐废止于清末，作为一种考试制度，它在科目设置、考试内容、考试方式、考试组织等方面都有比较完善的制度，也经历了“随意化—规范化—程式化”的发展阶段。

在科目设置上，隋朝设置了秀才科、明经科和进士科。唐朝早期沿袭了隋朝的科目设置，在中后期开设了众多科目，科目最多时多达 16 种。除了秀才科、明经科和进士科等综合型科目外，还有五经科、四经科、三礼科、三传科等衍生科，也有明法科、明书科、明算科、道科、开元礼科等实用型科目，一些科目是根据需要而临时设置的。宋朝的科目设置经历了由繁复到简明的过程，以北宋神宗熙宁四年（1071 年）为界，繁复期多达 11 科，简明期仅以进士一科取士。自此，进士科成为科举考试的必考科目，并且衍生了如经义进士科、诗赋进士科和策论进士科等更具体的进士科。辽、金在保留进士科为必考科目的同时，也开设了律科，但其地位无法匹及进士科。元、明、清三代的考试科目设置比较单一。元代将经义进士科、诗赋进士科和策论进士科整合为德行明经科，对考试科目进行了规范、统一的设定。明代、清代沿袭元代开设德行明经科，清代同时为八旗子弟开设了翻译科，但影响甚小，应试者寥寥。

在考试内容上，从隋唐到北宋前期，由于科目设置繁复，考试范围也比较庞杂。从北宋后期开始，科目设置比较简明，考试范围也更加明确。明经类科目主要有四经科、五经科、九经科、明二经、明三经、学究科、三传科和三礼科等。主要的考试范围是儒家经典，主要包括：《诗经》、《尚书》、《周易》、《论语》、《孟子》、《礼记》、《周礼》、《仪礼》、《孝经》、《尔雅》、《春秋穀梁传》、《春秋左氏传》、《春秋公羊传》等。科目不同，所考查的儒家经典也不尽相同。每一科目一般会考“十三经”中的一部或几部，这个考试范围在不同朝代有不同的规定。每个科目要求的儒家经典的数量不一甚至有些随意，单一化、随意化的特点成为科举制度暴露出来的问题。纵观科举约 1300 年的发展，尽管不同朝代、不同科目的考试范围不尽相同，但以儒家经学为主要内容的方向却一直未曾改变，甚至得到不断强化。

实用类科目主要是为选拔某一特定领域、特定行业的人才而设置的科目。唐宋时期都设置了明法科、开元礼科等科目。明法科以当朝施行的刑法律令为主要考试内容，开元礼科主要考试内容为《大唐开元礼》（150卷）。术技科目主要有明算科、明书科、医科等。明算科的考试范围为历代算学经典，如《五经算术》、《九章算术》、《缀术》、《孙子算经》、《五曹算经》、《周髀算经》、《缉古算经》、《夏侯阳算经》等。明书科则主要考试《说文》、《字林》两部字书，兼试书法。由于科举制度旨在培养后备官员，实用类科目的考生仅能胜任某一领域的具体工作，故在后期这种科目逐渐为综合性科目所取代。

进士科是科举制度中最受重视的科目。不仅考生众多，而且进士出身的考生多被授予高官。在隋代及唐初，进士科多考时务策。经过唐太宗、唐高宗时期的改革，进士科的考试内容增至三种，包括经学、文学、时务策。宋、辽、金三朝又对进士科进一步划分，北宋将进士科分为经义进士科和诗赋进士科，金代在此基础上又开设策论进士科。至元、明、清，由进士科发展起来的德行明经科成为标准科目，文学和时务成为次要内容，而经学的考试范围则被局限于四书五经。四书包括《大学》、《中庸》、《论语》和《孟子》，五经包括《周易》、《尚书》、《诗经》、《礼记》、《春秋》。科举制度在科目设置和考试内容上逐渐趋于规范化和标准化，这虽然是科举制自身发展的需要和专制政治的需要，但也使科举制度逐渐失去活力与生命力。

在考试方式上，主要有贴经、经义、策问、诗赋等。明经科目考试多采用贴经的方式，类似于今天的填空题。这种方式可以考查应试者对儒家经典的熟悉程度，考查内容配有规范标准的答案，这在一定程度上提高了考试的规范性和操作性，也保证了考试的公平性和公正性，但很难检验出应试者对经典的认知和理解水平，因此仅适合于考查记忆能力的内容。

经义是一种注重考查应试者能力的方式，要求应试者根据规定的文句展开议论，阐发义理，类似于现在的议论文。经义有口义、墨义之分。口义相当于口试，由考官现场出题，应试者随即回答，考官根据其回答评定成绩。墨义相当于笔试，它可以在同一时间内组织众多应试者参考，再由考官统一阅卷。墨义便于组织、省时省力，也是一种充分体现公平、公正的考试方式。

从隋唐到宋代，贴经和经义一直并存使用。尤其在北宋之前，贴经是科举考试采用的主要方式，从北宋开始，经义（主要是墨义）开始逐渐替代贴经，甚至成为儒家经学的唯一方式。元、明、清主要以经义、各种文体和时务策为主要方式，对经义作了进一步改造，逐渐形成程式化的八股文。

在考试组织上，科举制也有着严密的等级与严格的要求。在隋代及唐初时期，科举制只有省试一个级别。唐代的规范化程度远远高于隋代。在省试之前，唐代有统一、规范的“解送试”制度（由国子学及各州府按照分配的名额，举行相应的考试选拔一定数量的士



子，再推荐到朝廷参加尚书省的考试)；在省试之后，要再次进行“覆试”(对省试的检查和认定)。宋、辽、金三代在省试基础上增加了殿试。殿试由皇帝亲自主持，皇帝直接掌握选拔和任命官员的权力。殿试虽由唐代武则天创设，但在唐代并未形成定制。直到北宋开宝五年(公元972年)和开宝六年(公元973年)，宋太祖连续亲自覆试省试合格进士，并授予官职，殿试才形成建制。辽、金的科举制度也主要由省试(或会试)、殿试构成，省试、殿试的实质是一个级别、两个步骤。当然，省试之前的解送试也是必不可少的。

随着社会政治、经济、文化的发展，科举考试的选拔功能更加突出，参加科举考试的考生也越来越多。省试(或会试)、殿试这样的一级考试制度已经不能适应社会的需要。元代将解送试改为乡试，各地对乡试的重视程度也大大提高。明代将乡试进一步规范与明确，清代乡试更加严格。清代乡试之后要进行覆试，再次对乡试结果进行核定，以保障乡试的公平、公正。乡试制度的成型，意味着科举制度变得更加规范与统一。

在考试时间上，北宋之前基本都是一年一届。宋代早期的省试、殿试也是一年一届，但随着参考人数和选拔人数的增加，考试组织实施出现了疲于应付的问题，常常会采用隔年或者相隔数年才举行一届考试的做法。至北宋英宗治平三年(1066年)才形成三年一考的定制。辽、金基本沿袭了宋代三年一周期的做法，并将具体时间进行了规定。元、明、清也对乡试、会试及殿试的具体时间作了规定。

科举制从隋朝大业二年到清朝光绪三十一年(1905年)举行最后一科进士考试为止，存在了1300年左右。它的科目设置、考试内容、考试方式、考试组织等制度大多借助经验性思维，依靠统治者的主观需求与维护统治的需要而建立。这种制度的形成经历了由不固定(不明确)到相对固定(明确)，由相对固定(明确)到高度一致的过程，这种高度程式化和规范化的做法，也直接加速了科举制度退出历史舞台的步伐。但科举制度仍然发挥过重要的作用，它不仅打破了大官僚贵族依靠门荫把持选官权力的积弊，为庶族中小地主与出身寒微的平民开辟了入仕途径，大大扩展了封建统治的社会基础，更促成了教育制度和选官制度的一体化，为统治者加强中央集权和皇权提供了教育基础和人才基础。

二、教育测量的学科定型：客观标准化

18—19世纪，就在我国科举制度逐渐走向程式化、僵化之时，欧美各国在工业革命的推动和教育普及的浪潮之下，受西方现代心理学、教育学、统计学等学科理论发展的影响，教育测量呈现出生机勃勃的发展态势，并作为一门学科正式产生。

18世纪之前，欧美学校的考试大都采用口试。随着教育规模的迅速发展，学生人数迅猛激增，口试无法适应发展的需要，笔试逐渐登上历史舞台。1702年，英国剑桥大学首先采用笔试代替口试，成为西方教育测量发展史上重要的里程碑。1845年，在教育家霍勒斯·曼(Horace Mann)的倡导下，美国波士顿市开始以笔试考查毕业生，掀开了以

统一的试卷测量众多学生的历史新篇章。为进一步克服命题与评卷的主观性、随意性，提高考试的可靠性和客观性，研究者们致力于开发各种量表。1864年，英国教师乔治·费希尔（George Fisher）收集了不同水平学生的语法、作文、拼字、书法、算术、历史、自然、图画等学科成绩样本，编制成《量表集》，为评价学生各科成绩提供了标准，这是客观标准化测量的萌芽。1894年，约瑟夫·赖斯（Joseph Rice）选定50个字对20所学校的16 000多名学生进行测验，设计了拼字测验。

19世纪末20世纪初，心理学的发展与实验心理学的兴起对教育测量产生了很大的影响。1883年，高尔顿（Galton）出版了《人类才能及其发展的研究》，首先提出“测验”、“心理测验”两个术语。1904年，桑代克（Thorndike）出版了《精神与社会测量导论》，该著作系统介绍了统计方法及编制测验的基本原理，提出了“凡物之存在必有其数量”的论断，为测量学奠定了理论基础。因此，桑代克被称为教育测量学的鼻祖。为巩固测验理论，桑代克还编制了一些测验与量表，运用统计学上的“等距原理”发明了量表编制的单位，编写了图画量表、书法量表、拼字量表、作文量表等标准测量工具。美国测量学家麦考尔（W. A. McCall）提出了T·B·C·F制（T: total ability; B: brightness; C: classification; F: effort），1922年又提出“凡有数量的东西都可以测量”。在桑代克、麦考尔等人的大力倡导和创新下，美国编制了各种标准化测验，设立了测验研究部，组织教育测量专家进行大规模教育测验，教育测量走上了科学化的道路。

从本质上来说，教育测量与心理测量是相互交融、互为发展的。1905年，比内（A. Binet）、西蒙（T. Simon）发表了《诊断异常儿童智力的新方法》一文，介绍了第一个智力量表——比内-西蒙量表。1908年，他们将年龄作为测量智力的准则，采用智力年龄的方法计算成绩，又发表了修订量表。这种方法解决了不同年龄相互比较的标准问题，增强了测量结果的明确性与可比性。1916年，斯坦福大学教授特曼（L. M. Terman）修订了比内-西蒙量表。修订后的量表首先引入了智力商数，使测量结果更加明确；量表增添了新的测试内容，也使测验的适用范围更加广泛。戴维·韦克斯勒（David Wechsler）编制的儿童、成人智力量表被广泛地应用，他摒弃以智龄为测量单位的做法，提出用离差智商代替比率智商。随着测验编制技术的不断成熟，智力测验越来越完善，各种能力倾向测验、操作测验、人格测验等也不断涌现。1918年后，教育测量的应用从小学延伸到中学甚至大学，一些大学因此设立了教育测量学课程。这些成就大大地推动了教育测量的发展，教育测量表现出客观化、标准化的显著特征。

首先，测验编制趋向客观化，命题、评分、成绩分析尽量避免主观经验性和片面性，试题类型尽量减少或避免主观性试题，以确保测验编制的准确性与客观性。其次，测验的测试范围比较大，测试对象比较多。最后，测验的标准客观化，测验大都建立了常模，形成了客观的标准。随着教育测量的应用不断扩大，这种客观化测验的弱点也逐渐显现。测



验编制的素材主要来源于教科书，多数测验重点考查学生对教材内容的记忆，对学生的兴趣、情感、态度、创造等缺乏关注，这不利于学生的全面发展。

我国辛亥革命后，教育测量学逐渐被引进，中国学者开始了对教育测量的研究与教学工作。1918年，俞子夷在借鉴桑代克书法量表的基础上，编制了“小学国文毛笔书法量表”，开创了我国现代教育测量的先河。1919年，廖世承和陈鹤琴开始在南京高等师范学校设置测验课程，开创了我国高校设置测验课程的先例。一些教育机构开始成立教育测验组织，对测验进行宣传和推广。1922年，中华教育改进社邀请麦考尔来中国讲学，并邀请他主持编制各种测验。1923年，中华教育改进社组织开展了全国小学教育调查。1924年，廖世承和陈鹤琴出版《测验概论》。1925年左右，全国编制了几十种中小学教育测验，这些测验涉及小学、中学等不同学段，测验内容包括语法、默读、国文、常识、化学、物理、地理、历史、算术、英语等科目。1931年，在艾伟、陈鹤琴、肖孝嵘、陆志书等人的倡议下，中国测验学会在南京成立，这标志着中国教育测验的发展进入了新的历史时期。1932年，中国测验学会会刊《测验》杂志创刊。中国测验学会通过举行年会和创办杂志来促进学术交流与量表编制，推动教育测量在我国的发展。20世纪30年代至40年代，我国陆续编制了一批测验并应用到实际教学中。一些学者出版了测验专著和教材，并在高等教育学校和中等师范学校中开设教育测量或心理与教育测量的课程。可以说，在此阶段我国教育测量的理论研究与实践应用都得到了很好的发展。

三、教育测量的深度发展：标准技术化

1929年，美国爆发经济危机。经济危机迫使人们对教育进行全面反思，也开始对考试评价进行改革。1934年，为了促进人才的全面发展，以美国俄亥俄州立大学教授泰勒(R. W. Tyler)为首的美国进步教育协会(PEA)教育研究委员会，在300所大学、学院和30余所中学开始了历时8年的课程与评价研究。泰勒试图将社会的要求、学生的需要整合到课程与测验中，他提出了一套以教育目标为核心的课程和测验编制原理。这在一定程度上促进了教育测量学的深入发展，一些专业的教育测量机构也陆续建立起来。1947年，美国成立了教育考试服务中心(Educational Testing Service, ETS)，它是世界上最大的非营利教育测试和评估组织，专门从事测验编制和组织标准化考试。

知识拓展：

20世纪30年代，资本主义国家经济危机的爆发导致美国经济开始衰退，引起人们对学校教育的反思，对中学的教育目标、课程设置、教学大纲和课程评价等提出了质疑。在此背景下，美国进步教育协会发起了一项历时8年(1934—1942年)的教育实验，目的在



于通过实验加强中学与大学的合作关系，实现中学升学与就业的双重目标，验证或确立进步教育协会提出的教育目标，同时建立新型的中等学校模式。实验在新成立的中学与大学关系委员会的指导下，从200所中学中挑选了不同地区、不同规模、不同水平的30所中学参与实验研究。同时，实验也得到了300所学院和大学的积极配合。参加实验的300所大学和学院对30所实验学校放弃原有的高考制度，根据中学校长的推荐和学生的在校表现来综合评价学生，以决定是否招收学生。为了更好地培养本校学生，参与实验的中学拥有足够的自主权来改革学校的教学大纲和教学方法。研究提出，学生应该把学习视为对人生意义的一种持久的探索，而不仅仅是积累学分；要让学生渴求学习，不断进取，并勇于探索新的领域；要使学生懂得如何安排时间，如何更好读书，如何更有效地运用学到的知识，对所面对的生活更富有经验。整个实验分为两个阶段：第一阶段从1934年到1936年，参与研究的学校在指导委员会及其分支委员会、进步教育家和专业人士的指导下，开始在内部实施改革，制定学校教学计划，编制课程，实施民主管理，鼓励教师参与实验研究，与学生合作开展实验活动。第二阶段从1936年到1942年，参与实验的大学和学院依据合作中学校长的推荐和学生的综合表现录取学生，由大学跟踪委员会对学生四年的大学生活开展跟踪，将参与实验的学生与没有参与实验的学生分成实验组和对照组，按照大学成功标准对实验组和对照组进行评价和比较，得出实验结论。1942年，中学与大学关系委员会发表了实验研究报告。尽管由于外在环境的变化，实验研究没能得以继续进行与深化应用，但还是取得了丰硕的成果。研究不仅得出“传统中学单纯强调学术性，为升学准备的课程不符合社会和学生发展的需要”这样的结论，还开创了现代教育实验的范式，确立了制定教育目标的原则，奠定了现代课程理论的基础，同时形成了现代教育评价理论。

随着教育测量的深入发展，各种测验更加标准化、客观化，教育测量也由过多重视测量学生的知识，转为重视对学生的智力、思想品德、能力等的测量；由追求单一答案的求同式思维，发展为重视多种答案的求异式思维；由偏重测量学生的学习成绩，发展到关注课程、教材、教育教学改革等各个方面；由传统单一的常模参照性测验，发展到常模参照性测验与目标参照性测验相结合的形式。

20世纪40年代到50年代，计算机的诞生为教育测量带来了一片新天地。计算机技术和教育测量思想的融合带来了测验内容、测验方法和测验形式的变革。它不仅可以改变传统纸笔形式，还可对测验内容进行分析、比较，大大减少测验的数据分析工作量，提升数据处理精确度，更可以实现测验的标准化，提高测验的客观性和科学性。同时，教育测量的理论研究与技术开发不断完善，提出了测验信度、效度、项目的难度与区分度等指标及其经典的统计分析方法，经典测验理论逐渐成形，为教育测量的发展奠定了理论基石。

四、教育测量的当代进展：多元智能化

随着经济社会的发展和教育改革的推进，教育测量在教育教学中发挥着越来越重要的作用，教育测量在人才培养和人才选拔中占据着举足轻重的地位。

自新中国成立后到 1978 年的 30 年，我国由于种种原因，教育测量学处于停滞状态。党的十一届三中全会以后，教育测量学得到了恢复和发展。一是发表出版了一系列的研究成果。张述祖在 1979 年第 5 期《教育研究》上发表《论教育测量的重要性和教育测量的一些基本概念》，这是建议重视教育测量的第一篇论文。1982 年，以叶佩华、邹有华为代表的学者翻译了《心理与教育的测验和评价》（1977 年第四版），这是教育测量恢复发展后第一本有影响的译著。1987 年，王汉澜教授主编的《教育测量学》出版。同时还有张厚燾的《标准化考试简介》、张敏强的《标准化考试》、郑日昌的《教育测量学的基础》等。众多学者开始翻译、修订、自编智力测验、人格测验和学习能力测验等，并着力推动我国高考的标准化改革，取得了一系列的成绩。

二是成立专业研究组织机构，建立教育测量学研讨平台。1981 年，成立了中国教育学会教育统计研究会，并组织了多次全国性学术研讨会。1988 年，正式将教育统计研究会更名为中国教育学会教育统计与测量研究会，开始组织国际性的学术研讨会。同年，我国加入国际教育成就评价协会。

三是着力推进教育测量学的学科建设，并做好人才培养。逐渐地，全国很多高等院校将教育测量学列为教育系的必修课，作为本科生、研究生的课程之一。一些教师培训也将教育测量及相关课程作为培训内容，提高教师们教育教学的应用能力。

20 世纪 70 年代到 80 年代，美国兴起基本能力测验运动，该运动的目的是弥补高中毕业生所缺乏的进入社会必需的技能与能力。同时，教育测量还被应用到教师入职中，教师在获得教师资格前必须通过一系列的测验，经检验具备了成为一名教师的基本能力，才能获得教师资格。至今，ETS 开发出了若干种标准化考试，同时还在超过 180 个国家、9 000 个地区管理着包括托福（Test of English as a Foreign Language, TOEFL）、托业（Test of English for International Communication, TOEIC）、美国研究生入学考试（Graduate Record Examinations, GRE）以及 Praxis 等在内的国际考试。此外还有各种职业性考试，律师、医生、会计、交通、军队等行业，都用教育测量来选拔人才、分配工作和鉴定工作效率。

在此基础上，教育测量的理论得到进一步丰富，项目反应理论在 20 世纪 70 年代得到逐步扩展，并在计算机技术的支持下逐渐应用到测量实践中，认知诊断成为测验发展的新的追求。随着计算机应用功能的逐渐强大与应用范围的扩大，科学化、标准化的测验得到极大发展，尤其是计算机化测验和题库得以迅速兴起。计算机化测验包括基于计算机的测验（computer-based test, CBT）和计算机自适应测验（computerized adaptive testing,

CAT)。CBT 是一种以计算机为载体的测验方式。目前，一些学业成就测验、智力测验、人格测验开始采用这种方式的测验。CAT 从测验的编制与组合、测验实施、测验的评分到结果的解释与呈现，都由计算机来实现。它克服了传统纸笔测验的不足，采用灵活的“因人施测”策略，提高了测验的信度、效度与效率。目前，ETS 推行的工商管理类研究生入学考试（graduate management admission test, GMAT）就是采用 CAT 这种测验形式。题库是为了适应考试事业大规模发展，更好地满足社会发展和标准化考试的需要而设立的。它既可以大大提高测验编制的效率与质量，真正实现平行试卷的编制，又能确保试题的保密性，实现考试的公平、公正。

随着计算机科学、认知科学及数理统计学等学科渗入教育测量学领域，教育测量学的多元智能化趋势还将进一步凸显，在教育实践中的地位和作用还会进一步提高。

第二节 教育测量学的学科地位与学科贡献

教育测量从 20 世纪诞生至今，为教育改革与发展发挥了积极作用，也受到各国教育研究者、管理者和教师们的青睐与重视。探讨教育测量学的学科地位与学科贡献，深入分析教育测量学发挥的作用，是我国深化教育改革与发展的内在需求，也是教育测量学学科发展的必然要求。

一、教育测量学是教育科学体系的重要学科

20 世纪以来，随着我国教育科学的学科建设不断推进，教育科学这个大家庭的成员越来越多，体系越来越完善，并继续朝着分化与综合这两条路径纵深发展。一方面，教育学的学科分化趋势更加突出，分支学科不断涌现；另一方面，教育学的学科综合趋势十分明显，交叉学科、综合学科、复合学科、边缘学科层出不穷。例如，教育哲学、教育逻辑学、教育社会学、教育管理学、教育法学、教育文化学、教育生态学、教育经济学、教育心理学、教育统计学、教育测量学、教育评价学、教育实验学、教育技术学、教育信息学等分支学科，就是各种学科不断交叉、相互渗透的结果，由此形成了一个庞大的教育科学体系。在这个体系中，教育测量学作为一门“研究如何运用方法来分析教育活动”^①的学科，显现出较强的工具性、应用性、交叉性等特点。

教育测量学是一门工具性学科。在社会科学研究中，教育测量是研究者广泛收集数据、资料、信息的重要手段，教育测量的结果成为定量研究的关键指标。在教育科学体系

^① 张敏强：《教育测量学》，代序 18 页，北京，人民教育出版社，1998。