

教育测量学

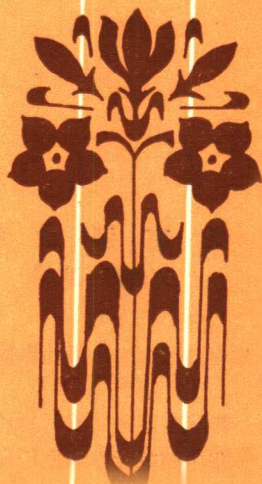
JIAOYU CELIANGXUE

○ 宋岭梅 主编

华中师范大学出版社

教 育 测 量 学

宋岭梅 主编



华中师范大学出版社

教 育 测 量 学

宋岭梅 主编

*

华中师范大学出版社出版发行

(武昌桂子山)

新华书店湖北发行所经销

华中师范大学印刷厂印刷

*

开本850×1168 1/32 印张15.125 字数387千字

1991年7月第1版 1991年7月第1次印刷

ISBN 7-5622-0641-4/G·214

印数 1—4500 定价6.00元

前 言

教育测量学是一门以马克思主义的教育科学理论为依据，运用教育统计学和数学的有关原理和方法，借助现代电子计算技术，研究如何运用各种测试手段和统计方法，测量学业成绩、智力水平、学习能力、教学质量以及教育措施的效益的一门学科。它是改进教学与考试方法、提高教学质量、正确选拔人才、实行教育管理科学化的重要工具。

这本《教育测量学》主要是以宋岭梅教授在华中师范大学教育系讲授教育测量学课程的内容为基础，和几位教师合作编写的。由宋岭梅担任主编，并编写第一、二、三、四、五、六章，徐虹编写第七、八章，苏永华编写第九、十章，孙静娟编写第十一、十二、十三章，附录中的一般电子计算器的使用和利用电子计算机进行数据处理的常用计算程序系苏永华和孙静娟编写的。

本书着重阐述教育测量的基本原理和方法，以及测度未来教育现象的主要预测方法。为了便于教学使用，各章均附有作业题。在编写中力求简明易懂，联系实际，既可作为高等师范院校教育测量学课程的教材，也可供广大教育工作者在教学与科学研究中参考。由于编写时间短促，我们的水平有限，本书难免会有一些缺点和错误，欢迎读者批评指正。

编 者

1990年9月

目 录

第一章 导论	(1)
第一节 教育测量学的研究对象和性质.....	(1)
第二节 教育测量学的作用.....	(7)
第三节 教育测量发展的历史概况.....	(11)
第二章 教育测量的几个基本问题	(22)
第一节 几个基本概念：测量、测验与评价.....	(22)
第二节 教育测量的要素和种类.....	(26)
第三节 良好测验的条件.....	(36)
第三章 测验的信度	(39)
第一节 信度的概念和意义.....	(39)
第二节 信度系数的估计方法.....	(46)
第三节 提高测验信度的方法.....	(73)
第四章 测验的效度	(82)
第一节 效度的概念和意义.....	(82)
第二节 效度的类型与估计方法.....	(87)
第三节 提高测验效度的方法.....	(109)
第五章 测验的难度与区分度	(118)
第一节 难度的概念和计算方法.....	(118)
第二节 区分度的概念与计算方法.....	(127)
第六章 测验试卷的编制	(142)
第一节 确定测验目的.....	(142)
第二节 客观性试题的编制.....	(146)
第三节 论文式试题的编制.....	(168)
第四节 测验的编辑和实施.....	(176)

第七章 测验的项目分析	(180)
第一节 测验项目分析的意义.....	(180)
第二节 常模参照性测验的项目分析.....	(181)
第三节 目标参照性测验的项目分析.....	(187)
第四节 题目的修订.....	(191)
第五节 潜在特质理论.....	(199)
第八章 标准化测验	(206)
第一节 标准化测验的基本要求.....	(206)
第二节 标准化测验的设计程序.....	(210)
第三节 标准化测验的实施步骤.....	(225)
第九章 学科测验	(230)
第一节 学科测验的意义及编制方法.....	(230)
第二节 语文测验.....	(238)
第三节 数学测验.....	(264)
第十章 智力测验	(275)
第一节 智力测验的意义.....	(275)
第二节 几种重要的智力测验.....	(286)
第三节 智力测验结果的解释及应用.....	(299)
第十一章 教育测量分数的组合与转换	(310)
第一节 教育测量分数的组合.....	(310)
第二节 教育测量分数的转换.....	(316)
第三节 多科测验成绩综合评定.....	(337)
第十二章 回归分析预测法	(348)
第一节 回归分析概述.....	(348)
第二节 一元线性回归预测法.....	(350)
第三节 多元线性回归预测法.....	(363)
第四节 回归分析中的几个问题.....	(382)
第十三章 马尔柯夫预测法	(393)
第一节 马尔柯夫链的基本原理.....	(393)

第二节 马尔柯夫链的转移概率矩阵的估算.....	(403)
第三节 马尔柯夫链在教育中的应用.....	(408)
附录：一般型电子计算器的使用和利用电子计算机进 行数据处理的常用计算程序.....	(430)
附表：	
1. 正态曲线的面积(P)与纵线(Y).....	(452)
2. Fan 的项目分析表.....	(456)
3. 由 ρ 值求 r 对照表.....	(457)
4. 由 R 值求 r 对照表.....	(458)
5. 肯德尔和谐系数显著性考验时 S 的临界值.....	(459)
6. 积差相关系数(r)显著性临界值表.....	(460)
7. t 值表.....	(461)
8. χ^2 值表.....	(462)
9. T 分数的计算辅助表.....	(463)
10. F 值表.....	(464)
主要参考书目.....	(472)

第一章 导 论

第一节 教育测量学的研究对象和性质

一、教育测量学的研究对象

教育测量学的研究对象，应从它的产生和发展的历史去分析认识。教育测量学的产生和发展，也和其他科学一样，是由于社会历史发展的客观需要。

从教育测量发展的历史可以了解到，学校教育的成绩考评，在本世纪20年代以前，世界各国(包括旧中国在内)，基本上采用的是大家所熟知的“旧法考试”。直至第一次世界大战前后，在西方特别是美国，由于科学技术的发展，迫切需要用科学方法甄别人才、选拔人才。作为旧法考试否定物的种种标准化智力测验和教育测验，雨后春笋般地大量编制出来，为促进教育测量这门科学的建立和发展奠定了基础。

我国现行的教育测量方法(包括课堂考试以及升学考试等)不够科学、准确。它主要有以下两个缺点：第一，有主观随意性。首先表现在命题上是凭经验命题，对学生的成绩考核，多是由主讲教师根据教材和自己的经验主观随意地出几道试题，然后根据学生书面答案的质量，评给一个百分制的分数，这实质上就是我国在20年代以前所采用的那一套方法——即旧法考试。其次，还表现在评分上，旧法考试缺乏评定成绩的可靠量具。虽然成绩考核也使用“分”为单位，但因测验的试题不同，教师评卷的标准各异，“分”的值并不相等。如同年级甲校张生

语文成绩为75分，并不一定比乙校李生语文成绩为80分低。分数的价值不同，各分数就没有可比性，这就削弱了教育测量的意义。这种考评办法的弊病早已为教育学专家所熟知。远在几十年前，著名教育家斯太奇（Starch, D.）等曾经做过试验，他把同一份英文试卷，复印后分请142位本学科中学教师评分，结果同一份试卷所得分数竟有35种，从50分到98分，评分高低相差很大。为了研究评分误差情况，广东省教育科学研究所曾桂兴同志曾对1984年高考理科（七门课程）的评分作过调查和实验。调查和实验的方法是：以学科为单位，随机地抽取五名考生的答卷，分别复印十份，分请十位评分者根据既定的统一评分标准进行评分，从而考察评分者给分高低的差异。结果各科都出现不同程度的差异，即评分误差。误差较大的是语文科（平均误差是11.5分）和政治科（平均误差6.5分）；数学、物理、生物三科的平均误差都不低于5分；误差较小的是大量使用客观题（指答案简单或唯一的题型）的外语科和化学科（平均误差不低于1.2分）。为了解论文题（考生可自由应答的题型）的评分误差情况，曾桂兴同志又从语文科中抽出一篇评分误差为10分的作文答卷，再复印20份，请另外20位阅卷者给分，结果对这一篇作文的评分误差竟达20分。可以估计到，由于参加高考的评分者人数增加，评分误差可能还会增大。第二，试题取样不适当，不能有效地代表全部教材内容。教育测验的试题是否能够很好地代表知识总体，是决定测验价值大小的关键。任何学科的教育测量都只能以该学科知识总体中的一个样本对学生进行测验；然后根据学生回答试题（样本）的情况去估计学生掌握知识总体的水平和程度。试题是教材内容的样本，取样是否恰当，影响着测验的效度。旧法考试一般以较少的几个题目进行测量，用较少试题作样本，却要以其测验结果来估计学生掌握全部教材的状况，因而，常常会出偏差。

综上所述，可以看出我国一般学校现行的成绩考评办法，存在着不科学、不准确的缺点，严重影响有效地培养人才和正确地

选拔人才。社会在发展，科学技术在进步，教育科学也有许多新理论、新课程、新的教学方法和新的教学设施不断产生，必须有与之相适应的新的教育测量方法，才能正确评定各种教学活动的价值。

因此，教育测量学应以怎样运用各种测试和统计方法分析试卷质量、编制各种测验、评定学生成绩、检查教育效果、转换测量成绩等作为自己的客观的研究对象。

教育测量学从本世纪20年代建立起，在短短的60多年时间，已获得迅速发展。我们应在辩证唯物主义的思想指导下，借鉴国外教育测量学的最近发展趋势，汲取其中某些已被实践验证的理论、方法，结合我国社会主义教育实际，在各种教学实践中不断探索编制测验与成绩考评的新经验，以促进我国教育测量学这门学科的建设和发展。

二、教育测量学的学科性质

教育是根据国家制定的教育目的要求而进行的一种有计划的活动。教育是否达到既定的目的要求，需要运用一定的测量工具，对教育效果进行考察，才能作出结论。教育测量的任务主要是对学生的基本知识和技能的掌握程度，及其应用、分析和综合能力的发展水平，以及思想品德的数量化进行测定。这些学业成就和心理品质是看不见摸不着的行为属性。

对学生的学业成就和心理品质的测量比对物体的物理特性的测量要困难得多。因为学生是活生生的人，常常会受到兴趣、爱好、情感、意志等以及社会环境的影响，时刻在变化着，不易控制。对物质方面的测量，由来已久，“权然后知轻重，度然后知长短”。在日常生活中，人们总是对各种物体不断地进行测量。如售货员用秤来权衡货物的重量；木工用米尺度量木料的长短；医生用温度计测定病人的体温等，这一切都是十分明显的测量过程。对物体的物理特性进行准确测量是人所共知的。但是对学

生的心理品质的测量，就不能用一架天平或一把米尺把学生的智力和数学的运算能力测量出来。然而，这并不是说，学生的学业成就、学习能力和思想品德等就不能测量。这就涉及到教育测量的学科性质，也是教育测量的理论依据。现分以下几点来说明：

(一) 凡物之存在必有其数量，凡有数量的东西都可以被测量。

“凡物之存在必有其数量”是美国心理学家桑代克(Thorndike, E.L.)于1918年提出来的一条重要原则。世界上任何现象，只要有质的存在，总有量的表示。根据辩证唯物主义观点，客观事物的质和量是辩证统一的，质总是具有一定量的质，量也总是有一定的质的量，没有一定的质和一定的量的事物是不存在的。正如毛泽东同志所指出的：“任何质量都表现为一定的数量，没有数量也就没有质量。”^①学生的学业成就和心理品质也不例外，只要有质的存在，总有量的表示。我们都承认人们的聪明才智有高低之分，计算的熟练程度有差别，语文水平有优差之分，这说明教育方面的一切现象，都有程度上的不同。程度上的不同，即数量上的差异。

“凡有数量的东西都可以被测量”，是美国教育测量学家麦柯尔(Mccall, W.A.)于1922年提出来的，这条原则在理论上是无可置疑的。对这个问题，麦柯尔作了最好的解释，他指出：没有一种数量是不能测量质的，也没有一种质是不能被测量的。世界上一切现象，一定有其特质，如高低、美丑、智愚等。而各种特质，总不能一样，必有程度的不同，所谓程度的不同，就是数量的不同。所以，一切现象只要有质的存在，就会有数量的现象，都是可以被测量的。这即是说，世界上存在的各种事物现象，都有程度的不同，程度上的不同也就是数量上的差异，既然有数量上的差异，就可以进行数量化的测定。从教育测量的内容看，在

^① 毛泽东：《党委会的工作方法》，《毛泽东著作选读》（下册）人民出版社，1986年8月，第一版，第671页。

测量学生学业方面，已有多种测验，如成绩测验、学能测验、诊断性测验、常模参照性测验、目标参照性测验等等，已经取得了大量成果。对心理品质方面如：智力、性格、兴趣等，也已经有多种测量工具，如各种测验量表，如观察法、问卷法、自陈法、人物推定法及投射法等。当然这些测量工具还不完善，测量程序也比较复杂，在测量过程中容易受其它因素的影响，不能得出十分确切的结果。但是，在教育测量的科学化方面已取得了一定的成绩，绝不是不可测量的。

(二) 教育测量是一种间接测量

教育测量与物质测量有着根本的区别，教育测量有它自己独具的特点。物质测量可以运用工具测量物体的本身，如一个物体的重量、长度、温度，可以用天平、尺子、温度计直接测量，一般采用的是直接测量的方法，来认识物体的物理属性。而教育测量所要测量的对象是学生对知识的理解程度、智力发展水平及才能等，这些都是触及不到的心理品质，属于行为属性。我们不可能运用某种测量工具直接测量这种行为属性，只能通过教育影响使受教育者从行为上表现出来，运用推理和判断的方法来辨别学生的知识水平、智力高低和品德的好坏。所以，除了对学生的身体素质和发育情况（如身高、体重、握力、视力、听力等）可以直接测量外，其它的教育测量都是采用间接测量的方法。正因为教育测量是间接的，测量的对象是人，不是物，这就大大增加了测量工作的困难。首先是在测量过程中必然会受到许多无关因素的影响，产生测量误差，影响测量结果的准确性。再就是对测量的内容，如要测量学生的智力，但对“智力”的理解，现代心理学家是众说纷纭，至今仍然没有统一的认识。又如布卢姆的学习目标分类，它反映了学习结果由低级到高级的层次，对教育工作者有很大启示，但不少教师对认知领域中知识、理解、应用、分析、综合、评价这六个层次，究竟应如何理解，如何测量，也没有统一的认识。由此可见，从教育测量是一种间接测量这一特点，就决

定了教育测量比物理测量要复杂得多。

(三) 教育测量尺度具有较大程度的不确定性

一般所说的测量，是一个与选定的标准物相比较，以对被测量物体赋值的过程。因此，测量所依据的标准物对测量的质量有较大的影响。对于物理测量来说，它的标准物是具体的、有形的。人们用天平的砝码作为称重的标准物，用米尺作为度量长短的标准物。这种标准物看得见、摸得着，因而，它具有较大的客观性。它可以在较大精确度的范围内，保证两次以上测量的同一性。但是，对于教育测量来说，情况就比较复杂了。从多数学校所用传统考试方法对学生学业成绩的评定，以及学习能力、思想品德的评定来看，其测量的标准存在于测量者的头脑之中，这些标准物一般都是无形的、抽象的，因而，在很大程度上是不确定的。

教育测量的不确定性表现在两个方面：第一，测量尺度的不同一性。所谓同一性是指对于一项科学的测量来说，作为标准物的尺度应在整个测量过程中是一致的、稳定的。就测量重量来说，不能时而规定一公斤等于995克，时而又规定一公斤等于1005克。测量标准物的一致性与稳定性品质，是测量结果能为大家公认的基础。然而，对于教育测量来说，测量尺度一般很少具有这种品质。事实上在教育测量中，对同一被试者的测量，不同的测评者使用的尺度往往不相一致，在前面曾举出不少这方面的例子。例如，我国教育科学研究人员，曾作过这样的调查，先请一位教师给40名学生数学试卷评分，隔一小时后，再请这位教师给已评过的原40份试卷重新评分。结果两次评定的成绩，完全一致的占百分之六十，不一致的竟占有百分之四十。数学试卷的评定尚且如此，论文式试卷这种易受主观性影响的评定，其差距可想而知。之所以产生这种误差，测量尺度的不同一性是其主要原因。这种现象在教育测量活动中，是屡见不鲜的，它必然会影响到测量的科学性。第二，教育测量活动的不易复验性。对物体的物理测量是可以复验的，在日常生活中，人们可以根据需要对某一测量

活动进行重复和检验。例如,我们可以用售货员使用的米尺对买的衣料进行再次丈量,以检验售货员是否有故意克扣的行为,这种测量是可以复验的。但是,对于教育测量来说,一般地就不具有这种品质。由于在教育测量中,教师作为学生学习质量的测评者,他所使用的尺度存在于他的头脑之中,我们很难用他所使用的尺度对他的测量进行检验。因此,我们就无法对测量的质量进行检查。教育测量活动的这一特点,决定了教育测量比物理测量要困难得多。

了解这些特点,对我们认识教育测量学的学科性质,科学地运用测量方法都是很有帮助的。

综上所述,我们已知学生的学业成绩、学习能力及思想品德等方面是可以测量的。但由于它采用的是间接测量的方法,同时教育测量的尺度又存在着较大程度的不确定性,因此,要使教育测量准确、可靠,能真实地反映考生的成绩,得到较科学的结论,必须对测量的方法认真地进行探讨。

教育测量学是一门以马克思主义的教育科学理论为依据,运用教育统计学和数学的有关原理和方法,借助现代电子计算技术,研究如何运用各种测试手段和统计方法,测量学业成绩、智力水平、学习能力、教学质量以及教育措施的效益的一门学科。它是改进教学与考试方法、提高教学质量、正确选拔人才、实行教育管理科学化的重要工具。

第二节 教育测量学的作用

本节主要阐述教育测量学的作用以及学习教育测量学的目的和方法。

一、教育测量学的作用

教育测量在教育过程中有着重要的地位与作用,它是改进教学、实行科学化教育管理和教育研究的重要工具。

(一) 教育测量是为实现教育目的服务的

教育测量是为改进教育工作，提高教学质量，为更好地实现教育目的服务的。教育测量对整个教育事业有着极为广泛的影响。各级各类学校都根据教育目的制定自己的培养目标，并根据这些目标确定相应的教学计划、课程、大纲、教材并选择教法。这一系列的教育措施是否能使预定的教学目标实现，达到教育目的，必须通过教育和教学效果来反映，而教育和教学效果又往往要通过测验来衡量。因此，教育测量的各种测验量表的制定，都必须根据教育目的的要求，以各种教材和教育内容为依据，以便更合理地组织教育活动，安排教学内容，选择教学方法。运用教育测验这一工具，可向教育部门提供可靠信息，以评价教育目标的贯彻、大纲的执行及教材教法改革的情况。那种不顾教育目的的要求，随意地制定测验量表，任意地进行测量、评价，都是不符合教育测量要求的。在考查学生学业与能力方面，教育测验有以下意义：

1. 检查学生掌握学科知识的程度

检查学生掌握学科知识的水平，是教学过程中的一个重要环节。通过测验检查学生的学习效果，将学生学习的结果与教学目标相对照，借以确定所达到的程度。这就要采用科学化的测试手段，以便客观、公平、合理地评定学生的学习成绩，以求得学生的真实成绩。目前不少国家积极改进考试办法，逐步采用绝对评分与相对评分相结合的方法，评定学生的测验成绩。它是指评定学生成绩，不仅以考生对测验所要求的全部知识内容究竟掌握多少作为评分的依据，而且还以考生测验成绩相互比较，按其在团体学生成绩中居于什么地位作为评分的依据，以正确、客观地检查学生掌握知识的程度。

2. 诊断学生学习上的缺点和难点

教育测验还可以起到诊断作用。一般学校的班级教学，对一个班级中的所有学生虽然曾经过慎密的选择，希望程度整齐。然

而，实际上学生成绩的优差仍很显著。多数学生可能属于中等，有少数优等生，学习上没有什么大的困难，总常有少数学生在学习上，难免有这样或那样的困难。教师有责任随时了解学生学习的情况，对优等和中等学生负责指引他们前进，对于在学习上感觉有困难的学生，有责任设法帮助克服困难，也能共同前进。测验可以使教师及时评定学生学习的效果，分辨优差，对学习成绩差的学生，应诊断困难之所在，以便矫正错误，弥补缺陷。比如，语文学习中，对学习发生困难的学生，教师在评定学习效果后，便可诊断出个别学生感觉困难的根源。有的可能是由于对某抽象的词汇不理解其意，有的可能是由于所掌握的词汇不够，而不易理解课文中的涵义，有的可能是由于眼睛视力较差，以致影响阅读的速度等。这一类困难，都会影响学生的语文学习效果。通过测验，教师可以及时了解学生中感到困难的部分，以便调整教学程序，有的放矢地进行教学。

3. 便于识别人才，因材施教

成绩测验、能力倾向测验、智力测验等可以较好地预测学生的学习能力和兴趣，可以为不同的学生安排适当的学校、专业，有利于因材施教，发挥学习潜在能力。智能测验在于区别智愚的等级，便于早期发现超常学生与智力有缺陷的儿童。识别超常的学生，是开发人才资源的重要手段，可以特别加以培植；对于智力有缺陷的儿童，可以及时采取措施予以矫正。对于具有某些特殊素质和才能（如音乐、绘图与体育等）的儿童，可早期发现，早期进行培养，这对智力开发和人才培养大有好处。测验也是编级、转学和插班、将学生安排在恰当班级的重要手段，使他们与以前所受的教育、现有的学业、能力水平相衔接，获得正常的发展。

4. 使教学从“开放系统”变成“闭合回路”，起反馈作用

教学工作不是一次完成，需要反复深入，螺旋前进，应根据收获改进耕耘，只顾耕耘不问收获，决不能提高产量。现代教育

理论证明教学效果的反馈和强化,使教学从“开放系统”,变成“闭合回路”,正是提高教学质量的关键所在。

学业成绩的检查与评定是整个教学过程中的有机组成部分。通过对学业成绩公正、合理地考核,不仅可以使学生及时了解自己学习的优缺点和对各学科达到学习目标的差距,对发生的错误随时加以矫正,还可以自动调节努力方向,充分发挥学习的主观能动作用。对教师来说,测验可为教师的教育、教学工作提供反馈信息,了解自己的教学效果,及时调整教学要求,改进教学方法,以保证教学任务的完成。

(二) 教育测量是教育管理的重要手段

我国多年来在教育管理方面,如对学校教育质量的评价,教师水平的评定,教育经费的使用,教育措施的效益等等,多是凭人们的经验来确定,缺乏科学的依据。为使教育管理科学化,可应用教育测量的理论和方法,根据一定的目的,编制各种量表,进行广泛的测量,对测量结果作出科学的数量分析与评价。它对于正确地评价学校教育质量、教师水平,合理地使用经费,科学地选拔、培养和安置人才,决定教育政策,进行教育规划等,都起着重要作用。

(三) 教育测量是教育研究的重要工具

随着社会的迅速发展,新的教育理论的建立,要改革教材、教法与教育措施,为了判断它是否正确,是否有价值,都需进行试验,并通过教育测量、检验才能得出结论。当前教育实践活动中有许多问题需要研究,如学校教育质量的评价问题,学生不同层次学习能力的培养与检查问题,改革考试制度问题等等。近来,在改革考试制度方面,已经引起教育行政部门的重视,国家教委已成立专门机构,积极开展改革高等学校招生考试的研究活动,已先后在广东、广西、辽宁、江西、山东等省进行标准化考试的试验,并把计算机等现代化手段引进考试,最大限度地控制评分误差。这项试验活动正在积极进行着。