



瞿葆奎 主 编

吕 达 副主编

教育科学分支学科丛书

JIAOYU KEXUE FENZHI XUEKE CONGSHU

12

教育测量学

JIAOYU CELIANGXUE

张敏强◎著



PEOPLE'S
EDUCATION
PRESS

人民教育出版社



教育科学分支学科丛书

JIAOYU KEXUE FENZHI XUEKE CONGSHU

- | | | | |
|----|-------|----|-------|
| 1 | 教育哲学 | 11 | 教育技术学 |
| 2 | 教育逻辑学 | 12 | 教育测量学 |
| 3 | 教育社会学 | 13 | 教育统计学 |
| 4 | 教育政治学 | 14 | 教育评价学 |
| 5 | 教育经济学 | 15 | 教育心理学 |
| 6 | 教育文化学 | 16 | 教育史学 |
| 7 | 教育生态学 | 17 | 教育实验学 |
| 8 | 教育卫生学 | 18 | 教学论 |
| 9 | 教育行政学 | 19 | 比较教育学 |
| 10 | 教育信息学 | 20 | 元教育学 |

ISBN 978-7-107-25965-4



9 787107 259654 >



瞿葆奎 主 编

吕 达 副主编

教育科学分支学科丛书

JIAOYU KEXUE FENZHI XUEKE CONGSHU

12

教育测量学

JIAOYU CELIANGXUE

张敏强◎著



PEOPLE'S
EDUCATION
PRESS

人民教育出版社

· 北京 ·

图书在版编目(CIP)数据

教育测量学/张敏强著. —北京:人民教育出版社, 2014. 4

(教育科学分支学科丛书: 函装典藏版/瞿葆奎主编)

ISBN 978 - 7 - 107 - 25965 - 4

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

中国版本图书馆 CIP 数据核字(2014)第 051298 号

人民教育出版社 出版发行

网址: <http://www.pep.com.cn>

北京天宇星印刷厂印装 全国新华书店经销

2014 年 4 月第 1 版 2014 年 5 月第 1 次印刷

开本: 890 毫米 × 1 240 毫米 1/32 印张: 8.25 字数: 209 千字

著作权所有·请勿擅用本书制作各类出版物·违者必究

如发现印、装质量问题,影响阅读,请与本社出版科联系调换。

(联系地址:北京市海淀区中关村南大街 17 号院 1 号楼 邮编:100081)

目 录

绪 论	(1)
第一节 教育测量的性质与功能	(1)
一、教育测量的性质与教育测量的科学	(1)
二、教育测量的功能和地位	(4)
三、正确使用教育测量	(7)
第二节 教育测量的回顾与展望	(9)
一、教育测量学简史	(9)
二、教育测量的现状及发展趋势	(14)
第一章 教育测量的理论基础	(18)
第一节 教育测量的基本概念	(18)
一、几个基本概念	(18)
二、教育测量要素	(21)
三、教育测量的量表	(22)
第二节 教育测验的特征分类	(24)
一、测验目标——成就、能力与能力倾向	(24)
二、测验题目的类型——客观性与主观性	(30)
三、测验分数的解释——常模参照与标准参照	(37)
四、测验标准化——标准化成就测验与教师 自编测验	(49)
第三节 经典测验理论的基础模型及评析	(51)

一、真分数模型	(51)
二、经典测量理论的应用价值	(53)
第二章 教育测验的编制	(55)
第一节 教育测验编制的基本程序	(55)
一、教育目标与测验目标	(55)
二、教育内容与测验内容	(60)
三、测验设计	(63)
四、测验的技术分析与鉴定	(65)
第二节 测验题目编制技术	(70)
一、客观性测验题目的编制	(70)
二、主观性测验题目的编制	(80)
第三节 测验题目统计分析技术	(83)
一、题目难度	(83)
二、题目区分度	(88)
三、综合题目分析	(95)
四、题目分析的特殊问题——标准参照测验 的题目分析	(97)
第三章 教育测验的质量评估	(101)
第一节 测验的误差	(101)
一、测验误差的定义	(101)
二、测验误差随机变量的结构	(102)
三、测验误差的诱因	(102)
第二节 测验的信度	(104)
一、作为一般概念的信度	(105)
二、测验的平行测量——真分数等值模型	(105)
三、信度系数的评估方法	(107)

四、真分数对观测分数的回归	(114)
五、信度指标的改进	(117)
六、特殊信度问题——标准参照测验的信度	(119)
第三节 测验的效度	(122)
一、效度的概念及其演化	(122)
二、效度的类型及其功能	(123)
三、效度评估	(128)
四、效度指标的改进	(136)
第四章 现代教育测验理论——题目反应理论	(139)
第一节 测验理论的发展概述	(139)
一、经典测验理论与题目反应理论的特点比较	(139)
二、经典测验理论与题目反应理论在题目分析上的区别与联系	(143)
第二节 潜在特质与题目反应模型	(147)
一、潜在特质评述	(147)
二、题目反应模型	(151)
第三节 题目参数与能力参数的估计	(163)
一、题目参数估计	(163)
二、能力参数估计	(170)
第四节 信息函数及其应用	(176)
一、信息函数的评述	(176)
二、测验题目的信息函数	(178)
三、测验的信息函数	(184)
第五节 应用实例及评析	(188)
一、题目反应理论的应用实例	(188)
二、实用问题的评析	(195)

第五章 教育测验的等值	(198)
第一节 测验等值的构想	(198)
一、测验等值的概念与条件	(198)
二、测验等值的设计	(201)
第二节 经典测验理论的等值	(205)
一、经典测验等值的定义	(205)
二、经典测验等值的方法	(206)
第三节 题目反应理论的等值	(212)
一、真分数等值法	(212)
二、回归等值法	(213)
三、平均数和 Sigma 等值法	(214)
四、强平均数和 Sigma 等值法	(214)
五、特征曲线等值法	(216)
第四节 应用实例及评析	(219)
一、题目反应理论等值的应用实例	(219)
二、实用问题的评析	(227)
第六章 教育测验的题库	(229)
第一节 题库建立的理论与方法	(229)
一、题库与经典测验理论	(230)
二、题库与题目反应理论	(232)
三、题库与教育理论	(235)
第二节 题库的管理	(236)
一、题库管理的要素	(236)
二、题库建设的程序	(237)
第三节 题库的使用	(239)
一、题目反应理论的题库应用准则	(239)

二、自适应测验·····	(244)
三、自适应测验的编制和测验实施·····	(246)
主要参考文献·····	(250)
后 记·····	(251)

绪 论

第一节 教育测量的性质与功能

随着科学技术的发展和进步,人们不但对物理测量作出了越来越精确的测量,同时,人们也不断地尝试对人的知识、思维、想象和能力、学术水平、成就等心理特征进行测量。现代教育理论的发展,更加注意强调人的素质教育,强调发挥人的主观能动性,强调因材施教,要检验教育的效果,离不开对被教育者的评价,其中最重要的一环就是采用教育测量的方法来检测教育效果。什么是教育测量呢?简单地讲,教育测量就是根据一定的法则用数字对教育效果或过程加以确定,教育测量学即是以现代教育学、心理学和统计学作为基础,运用各种测试方法和手段,运用计算机技术及相关的技术方法,对教育现状、教育效果和学业成就、能力、品格、学术能力倾向等等方面进行科学的测定的教育科学的一门分支。

一、教育测量的性质与教育测量的科学

人们较易理解的测量是物理测量。如日常生活中的度、量、衡都是司空见惯的;自然科学中的各种测试,由于测量技术和测量工具的进步及改善,可以进行直接的测量。物理测量一般能够获得比较准确和可靠的结果,也正缘于其直接测量的特征。

与物理测量相比较,教育测量的对象是复杂得多的人。正如世上没有两片完全相同的树叶,世上也没有两个完全相同的人。所谓

千人千面，人的内在的心理物质亦是千差万别。人与人之间存在的差异，在心理学中被称作个别差异。教育测量永远无法直接测量，它只能通过检测心理现象的外显行为或外在表现特征来推知个体的心理能力和个性特点等。同时，教育测量很难排除一些无关因素的影响，诸如知识水平、教学条件、师资水平、情绪、健康状况、主试导向等多方面因素都或多或少地影响到教育测量的结果，使之出现随机性或误差。

显然，教育测量的间接性、多元性和随机性注定它比物理测量要复杂困难得多。但是，美国心理与教育测量学家桑代克（Thorndike, E. L.）和麦柯尔（McCall, W. A.）早就提出一个假说：“凡是存在必有数量，既有数量即可测量。”^① 这就是说，事物的质可以转化为量来计算。教育测量的存在既然是必要的，我们自然希望其亦能够合理存在，即能够尽量减少和排除内在的或外在的无关因素的影响，提高测量的准确性。这就要求教育测量必须在技术和工具上狠下功夫，才有可能与物理测量相媲美，加强人们的心理认同感。

进行测量必须要有相应的测量工具。显然，测量工具的性能好，测量的效果必然也好。物理测量借助于量尺、温度计、计时器、磅秤等等工具，不同的工具又有不同的量纲单位，也就是说，它们各有不同的参照标准，这样才能使其测量结果具有可比性或可加性等性质。教育测量的主要工具则是测验。测验旨在对教育效果进行科学的测量，因此，掌握教育测量的科学就成为必需。教育测量学研究的主要内容就是测验的编制和使用以及测验结果的评价。

教育测量所用的测验总是由一组题目组成，题目是构成测验的元素。好的测验必须是优良的题目的组合。比如一个用于选拔性目

^① [美] 桑代克, R. L., 哈根, E. P. 著, 叶佩华、邹有华、刘蔚成主译:《心理与教育的测量和评价》, 人民教育出版社 1985 年版, “译者的话”第 3 页。

的测验就应当可以把具有不同学业水平的考生区分开来,而若该测验中某一道题目所有的考生都得满分或都不得分,这一道题就失去了区分不同学业水平考生的效用。可见,选好题目是教育进行科学的测量的一项重要工作。

测量应具有可靠性、客观性。比如物理测量中长度的度量:用一把尺子度量某一物体的长度,那么,无论何人、何时、何地进行这项测量工作,其结果应该是一致的,唯有如此,这把尺子才是可靠的;而若不同的人使用,或同一个人在不同时间、地点使用,测量的结果并不一样,这把尺子便是不可靠的。教育测量亦同此理。一个好的测验应当在不同的人使用、不同的时间、地点施测同样具有一致性。要想确保这样的一致性,无论是物理测量,还是教育测量,都必须进行科学的测量。对教育测量而言,我们就要研究和讨论测验的可靠性和客观性,这就是测验的信度问题;我们还要研究和讨论测验反映测量目的的有效性,这就是测验的效度问题。这两个评价指标缺一不可,可靠的测验并不一定有效。因此,寻求和编制具有优良信度和效度的测验是教育进行科学的测量的当然工作。

测验的结果一般都以分数或等级来表示。因此,测验分数的评定及比较、等级的划分以至对各个测验分数的解释等等问题,也是教育测量学不可忽视的问题。

学校、社会对测验的要求是多种多样的,如高考、升学考试、招技术工种、招行政管理人员、军人入伍、公务员选拔、研究生入学考试、成人高考、高中毕业会考等等。不同的测验目的有不同的测验要求,不同的测验又有不同的编制要求和不同的分数评定体系等。上述这些问题的研究亦需要科学的教育测量学。

教育测量学是教育科学中的一项重要分支,尤其是教育测量的对象是人,这就使我们更为注重要作科学的测量,而要作科学的测量,就有必要掌握教育测量的科学。

二、教育测量的功能和地位

当今的世界是充满决策的世界，如果学校能对学生的心理和教育进行全面系统的测量，根据测量所得的结果反馈于教学，那么，必可在实际的教学、教育决策中发挥很大的功用。教育测量的功能，摘要简述如下。

（一）因材施教

教育的一条基本原则就是“因材施教”。实际的教育过程欲体现这一原则，则要求教师必须了解自己的学生。了解的依据一般有二：一是基于经验之上的主观直觉，一是借助于测验的客观测量。前者是前科学的方法，有时并不可靠，常常是很大程度上受到主体状态的影响。例如对学生能力、知识水平等方面的测量，主观直感可以在一定范围内使教师了解学生的特点，但这并不完全靠得住，很容易受到教师的好恶、成见、情绪等多方面因素的影响。因此，为了更准确、更客观地了解自己的学生，使用测验并依据测验结果来了解学生是必不可少的。唯有如此，教师才能够针对学生的具体情况作出相应的合理安排，依据学生能力和已有知识水平的个别差异作出适当的教学决策，比如说编班、分组、开特殊课、课后个别辅导等等。所谓有的放矢，其效必佳。

（二）选拔人才

教育测量是一种选拔人才的手段。为了确保教育质量，确保培养出高层次高水平的人才，并最高效益地使用好有限的教育经费和教育设施，必须甄选出最有可能成功的学生。随着社会与科技的发展，凭借个人经验的选拔已无法满足实际的需要，高效、准确的决策辅助工具——测验——便尤其必要。比如：高中升学考试，全国高校入学考试，全国研究生入学考试等。只要科学地利用测验考试这一工具，将会提高选拔人才的可靠性和科学性。

随着社会化大生产的发展，人类的分工越来越精细，各种特殊

能力测验应运而生，并应用于不同部门和不同领域，使人员与工作之间建立最佳搭配，大大提高了人才选拔和职业训练的效率。

（三）诊断补救

始于 20 世纪初的测验运动的最直接的原因，在于了解那些心理上有缺陷的儿童、对他们作出早期诊断并进行早期矫治方面的特殊功效。

除临床外，测验的诊断功能亦体现在教育工作之中。学生在校的适应不良和学习困难缘于何故？可能因素很多，比如缺乏某方面知识或某种能力，具有某种不良性格或习惯，等等。测验能够将人的行为进行多方比较，从而确定一个人的相对的长处和短处，找出问题症结之所在，为有效解决问题提供较可靠的信息。教师可以依据这些信息，有针对性地采取补救措施，对症下药，才能事半功倍，达到最佳教学效果。

（四）评价教学

测验的评价功能在教学评估中既可面向教师与教学方法，亦可面向学生与学业成就。测验可以选拔与评定学校管理人员与教师，提高师资水平；还可以评价和鉴定教材与教学方法，从而提高教学质量。根据单元考试，期中与期末的综合考试等，可以对学生的学业成就作出评定，决定学生的等级成绩，使教师得以了解学生，亦使学生得以自我了解和自我评价。

（五）就业指导

在校的学生面临着一个共同的选择：以何种职业迈入社会？个人的能力倾向、知识范围、兴趣特长、职业价值观、个性特征等多方面因素的综合决定一个人最适合和应该选择何种职业。教师可以利用测验对学生进行全面评估，从而为其就业提供有效的指导；学生则可以根据各种测验结果帮助自我决策，以使自己能在社会上发挥最大潜力和作用；等等。

科学的教育测量要求每一个教育工作者都应该掌握教育测量的

科学，但由于许多实际的原因，教育测量学对相当多的教育工作者来说还是一门不够熟悉的学科。许多教育工作者会出考试试题，会组织各类测试，但还缺乏教育测量理论的学习和了解，因而在提高测试质量和改善测试等方面则工作就难以进行了。这可以说是教育界急需改善的状况。因此，教育测量学在教育中的地位是不可忽视的。

首先，教育测量是教育科学研究和教育教学改革的基本工具。要想发展教育事业，就必须适应科学技术的发展来改进教学体制，并引入新的教材和采用新的教法。但是，好的体制、教材和教法要通过教育效果来体现，而教育效果通常又由学生的测验成绩来说明。例如，全国有多套九年义务教育教材，分别为“人教版”、“华东版”、“沿海版”、“上海版”等等。这些教材既要执行全国统一的教学大纲与要求，又要结合本地的实际，还要反映地方的教育水平。而判别和了解教材的优劣和质量高低，则须通过相当长时间的教学实验和对学生进行考核，并尽量排除各种干扰因素，最终对不同的教材质量作出评定。

其次，教育测量是教育过程中的一个重要环节。教育是一个完整的系统，从教育控制论的角度看，教育过程可以被看作教师对学生学习进程的影响和控制的过程，而测验考试为这一过程提供了反馈信息。教师可以利用反馈信息来改进教学、提高教学质量，学生可以利用反馈信息来改善学习方法和效率，抓住重点和解决难题。取消教育测量，就等于取消教育质量检查，而没有教育质量检查的反馈，就失去了对教育过程的控制。可见，教育过程是离不开测验考试的。

最后，教育测量是相关学科的基础，并成为社会科学研究的一种重要工具。在教育学科中与教育测量密切相关的学科有教育统计学、教育评价学等，教育测量的结果成为这些相关学科的主要数据与资料来源。而在教育学、社会学、政治学、法学、经济学等许多

领域中,测验亦成为一种重要的研究工具,其结果被作为定量化研究的指标。没有教育测量,那么再完善的统计方法、评价技术亦无用武之地;而若教育测量没有进行科学的测量,其结果必然会有偏差,这些不准确、不可靠的数据又会不可避免地导致错误的结论、错误的解释,最终会导致错误的决策。可见,教育测量的科学性在此有至关重要的作用。

三、正确使用教育测量

在上述中我们可知,尽管教育测量在教育科学中具有相当的作用和地位,但是人们还是对测验的使用及其作用表现出了种种的担心,对各种考试考查方法提出了批评。但在未能找到替代方法之前,教育测量学中的考试方法一直在沿用着,而且由于批评,又促进了教育测量理论及方法的进一步完善及发展。但是,为了使教育测量能发挥最大的功能,把不良影响减少到最低限度,我们提出以下几条正确使用教育测验的建议。

(一) 测验的选择要慎重

每一种测验都有其特定目的、功能和适用范围,因此,对被试采用何种测验,应慎重考虑。而且,所选测验必须符合教育测量学的要求,即必须是标准化的,有可接受的信度、效度水平的测验。

(二) 测验使用者要具备一定的资格

一个合格的测验者必须经过严格的教育测量学的训练,具备普通心理学的基本知识。测验的选择、实施、记分以及测验分数的解释等,只有训练有素的教育测验工作者、教育工作者或医生才能胜任。

(三) 测验的内容要保密

对于大多数心理与教育测验来说,公开或泄露测验内容,可能会使测验失效。例如,有的教师为了让他的学生在测验中取得好分数,就预先泄露一些类似题目给学生练习,结果学生的测验成绩普

遍提高，可是测验本身作为预测或诊断工具的效度却相应降低了。当然，这种不负责任的教师只是极少数，但是，在一般的科普读物上或公开发行的书籍上（包括某些教科书）也常公开发表一些测验题目或以测验题举例，这样一来，测验内容的失密不可避免地导致测验结果的失效。因此，应严加控制测验的出版和发行。

（四）测验的施测过程要严格控制误差

任何测验实施过程中的操作，需要按照该测验严格规定的要求进行，严格控制这一过程中可能出现的各种无关因素。

（五）正确解释和看待测验结果

测验分数的解释人员必须经过专门的训练，懂得所使用的测验及测验结果的解释原则。对测验结果作出合理而恰当的解释，必须根据效度资料加以说明。并且，由于测验的信度不可能完美，所以任何时候都要把所测得的分数看成是一个范围而不是一个确定的点，在解释分数时，应提供一个可能的参考范围或者是一个最佳估计。不同的测验分数不能直接进行比较，即使两个名称相同的测验，由于所包含的具体内容不同，样本的组成不同，量表单位的各异，其分数便不具可比性。在宣布和分析测验结果时，主试不应简单宣布一个分数，而应给予合理的解释。

此外，测验分数并不是存在于真空之中，分数的意义是和与分数相配合的信息群的完整性及广泛性成正比的。这些信息群包括：个人文化教养方面的信息、家庭情况、个人历史、生理与健康状况，以及其他许多方面的情况。因此，教育测验的结果只是部分说明了学生在校学习上可能或已经达到的知识水平。教育者对任何一项决策的制定，除了考虑测验分数以外，还应充分并正确地把一切可以利用的个人资料都加以综合、权衡和整理。而且，不能一劳永逸地看待一次测验分数。在教育过程中必须利用测验不断地检查学生的学习进展情况，并在此基础上迅速调整指导方法，作出新的但同样并非一劳永逸的决策。