


JIAOYU
CELIANG YU TONGJI

教育

测量与统计


■ 主编 / 于向英

 郑州大学出版社

教育

测量与统计

■ 主编 / 于向英

 郑州大学出版社

图书在版编目(CIP)数据

教育测量与统计/于向英主编, — 郑州: 郑州大学出版社, 2004. 1

ISBN 7 - 81048 - 702 - 7

I. 教… II. 于… III. ①教育测验 - 师范大学 - 教材②教育统计 - 师范大学 - 教材 IV. ①G449②G40 - 051

中国版本图书馆 CIP 数据核字 (2003) 第 096301 号

郑州大学出版社出版发行

郑州市大学路 40 号

出版人: 谷振涛

全国新华书店经销

郑州文华印务有限公司印制

开本: 710 mm × 970 mm

印张: 15.5

字数: 270 千字

版次: 2004 年 1 月第 1 版

邮政编码: 450052

发行部电话: 0371 - 6966070

1/16

印次: 2004 年 1 月第 1 次印刷

书号: ISBN 7 - 81048 - 702 - 7/G · 73 定价: 23.00 元

本书如有印装质量问题, 由承印厂负责调换。

《教育测量与统计》编委会名单

主 编 于向英

副主编 李玉向 张 军

编 委 于向英 李玉向 张 军 李国富

内容提要

基础教育新课程的实施,使得传统教育测量面临前所未有的压力和挑战,教育实践中,直接把测检的结果等同于对教育事实的描述的做法,使得人们认为教育测量将有生命的人当作客体来研究,只用抽象的数字对事物的量的属性做出描述和判断。本书力图明确教育测量与教育评价以及教育统计三者的关系,从测量理论和学校的测量实践中,对教育测量进行准确定位。

在设计中,本书以基本理论、测量编制、质量分析、统计处理、教育运用为基本内容来建构学科体系。尽可能避免传统的教育测量学和教育统计学的那种菜单式、堆砌式的简单摆放。并试图将动态测评的理论引入学科体系,强调统计工具的使用,全面介绍教育测量在各种教育活动中的应用。

本书第一章介绍了教育测量的概念、特点、功能、分类等。同时,还简要介绍了教育评价和教育统计的基本知识,并阐述了教育测量与二者的关系。第二章介绍了教育测量的基本理论:重点介绍了经典测验理论、题目反应理论、动态测评理论等三种理论的基础模型。第三章介绍了教育测量编制的基本程序、编制技术统计分析等基本要素。第四章介绍了教育测验和误差、信度、效度等质量分析的相关要素。第五章介绍了教育测验分数的整理、一般趋势度量、假设检验及统计分析等教育测验结果的统计处理。第六章结合教育实践介绍了教育测量与统计在学科教学测验、智力和非智力因素测验中的具体运用。

序 言

现代教育已发展到 21 世纪,教育改革方兴未艾。世界各国的教育理论家进行了多种多样的理论探索,而各种层次的教育实践家也在不断地进行各种教育尝试,他们都想找到一条最佳的教育发展道路,都想找到一条成功的教育捷径,为本国、本地区的经济发展奠定坚实的基础,为经济的持续发展提供源源不断的人才资源。

什么样的教育是真正的教育?什么样的教育改革是成功的改革?什么样的发展才是人的主动、和谐、全面的发展?等等,这都需要一个严密、完善、科学的检验体系。教育测量是根据一定的原则用数字对教育效果或过程加以确定,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,对教育现状、教育效果、学业成就等方面进行科学的测定。教育统计是利用统计学的方法,将大量、离散的通过调查、实验得到的数据进行科学的推断、处理,使其能真实地反映教育中存在的现象和蕴含的客观规律。教育统计和教育测量都是运用数学方法的研究手段,二者紧密相关,相辅相成。测量是统计的前提,有了测量提供的数据,统计才得以进行;而统计又是测量的基础,从事测量的人员若没有统计的基础知识,也就无法整理、分析测量的结果。

教育测量和教育统计的作用主要表现在:第一,检查学习。通过学科测验可以检查和评定学生的学习效果。第二,因材施教。人的智力存在着个体差异,在结构方面也有优势领域的不同。教育应以智力测验和教育测验的结果为依据实施因材施教。第三,教育科研。要科学地评价一种新的教育理论、新的教材、新的教育方法的价值,离开教育测量和统计也就无法得出客观的结论。

本书尝试测量学与统计学的学科整合,并将二者置于课程之中,从课程目标出发,明确教育测量为课程服务,为教育改革服务的基本思想。以教育测量学为核心,运用恰当的统计方法和现实有效的统计工具,以科学结论服务于教育评价,理论结合实践,努力使学习者学有所用,学能所用。

本书的对象是中小学教师,并以此来提高中小学教师的教育测量与统计的

能力,为使他们培养出更多、更合格的未来社会主义事业的建设者服务!

编者

2003. 6

目 录

第一章 教育测量概述	1
第一节 教育测量的涵义	1
一、教育测量的对象	1
二、教育测量的可能	2
三、教育测量的要素	2
四、教育测验的量表	4
五、教育量表的水平	5
第二节 教育测量的特点	6
一、间接性	6
二、相对性	6
三、动态性	6
四、模糊性	7
第三节 教育测量的功能	7
一、检查学习	7
二、因材施教	8
三、反馈激励	8
四、教育科研	8
五、指导就业	8
六、课程开发	9
第四节 教育测量的种类	10
一、按测量目的分类	10
二、按测量的属性分类	11
三、按测验的标准化程度分类	12
四、按记分标准分类	12
五、按同时受测的人数分类	13
六、按测验所用的材料和形式分类	14
第五节 教育测量与教育评价	14
一、教育评价的涵义	14
二、教育评价的价值取向	15

三、教育评价的范式·····	17
四、教育评价的发展趋势·····	18
五、教育评价与教育测量·····	19
第六节 教育测量与教育统计·····	20
一、教育统计的涵义·····	20
二、教育统计的内容·····	20
三、教育测量与教育统计·····	24
第二章 教育测量的基本理论 ·····	25
第一节 经典测验理论的基础模型·····	26
一、真分数模型·····	26
二、经典测量理论的特点·····	27
三、经典测量理论的应用价值·····	29
第二节 题目反应理论的基础模型·····	29
一、题目反应理论的概述·····	29
二、潜在特质与题目反应模型·····	33
三、题目反应理论的应用实例及评析·····	39
第三节 概化理论的基础模型·····	43
一、理论简介·····	43
二、基本原理·····	44
三、单侧面随机设计·····	46
四、双侧面完全随机交叉设计·····	48
五、理论评价·····	49
第三章 教育测验的编制 ·····	51
第一节 教育测验编制的基本程序·····	51
一、确定测验目标·····	51
二、确定测验内容·····	54
三、教育测验的设计·····	56
四、教育测验的技术分析与鉴定·····	57
第二节 测量题目的编制技术·····	60
一、客观性试题的编制·····	60
二、主观性试题的编制·····	64
第三节 测量题目的统计分析·····	67
一、题目难度·····	67
二、题目区分度·····	69

第四节 测量题目的质量分析	73
一、定性分析	73
二、定量分析	74
三、常模参照测验的项目分析	74
四、目标参照测验的项目分析	76
第五节 测验题库的建立	78
一、题库建立的理论与技术	78
二、试题库的管理	79
三、试题库建设的实施	79
四、试题库的使用	80
第四章 教育测验的质量分析	81
第一节 测验的误差	81
一、测验误差的定义	81
二、测验误差随机变量的结构	82
三、测验误差的诱因	82
第二节 测验的信度	84
一、作为一般概念的信度	84
二、测验的平行测量——真分数等值模型	85
三、信度系数的评估方法	87
四、真分数对观测分数的回归	92
五、信度指标的改进	94
六、特殊信度问题——标准参照测验的信度	95
第三节 测验的效度	98
一、效度的概念及其演化	98
二、传统的效度分类及其功能	99
三、效度评估	103
四、效度指标的改进	109
第五章 教育测量结果的统计处理	112
第一节 教育统计的基本概念	112
一、教育统计的涵义	112
二、统计学中的几个基本概念	114
第二节 教育测验分数的整理	116
一、经常性资料	116
二、专题性资料	116

三、数据的种类	118
四、统计表	119
五、统计图	121
第三节 教育测验分数的转换	125
一、百分等级分数	125
二、标准分数	128
三、T分数	130
第四节 教育测验的一般趋势度量	131
一、集中趋势的度量	131
二、离中趋势的度量	136
第五节 教育测验分数的假设检验	140
一、假设检验的基本原理	140
二、抽样分布	142
三、总体平均数的估计	143
四、总体平均数的显著性检验	145
第六节 教育测验分数的统计分析	151
一、相关分析	151
二、方差分析	160
三、回归分析	168
第六章 教育测量在学校中的运用	173
第一节 学科教学测验	173
一、基本涵义	173
二、编制原则	174
三、结果解释	175
第二节 智力因素测验	177
一、智力涵义	177
二、智力结构	179
三、智力测验	181
四、智力测验的种类	181
五、智力测验的功能	182
六、智力测验的编制	183
七、智力测验实施	185
八、智力水平的表达	186
九、智商解释	188

第三节 人格测量	189
一、基本涵义	189
二、人格测量的功能	190
三、人格测量的方法	191
四、自陈量表法	193
五、优化策略	195
第四节 卡特尔 16 人格因素量表(16PF)	196
一、理论背景	196
二、卡特尔人格特质因素	197
三、16PF 的施测	198
四、16PF 的因素解释	198
第五节 投射测验法	201
一、投射测验的假设	202
二、投射测验的特点	202
三、投射测验的分类	203
四、罗夏墨迹测验	203
五、默瑞主题统觉测验(TAT)	205
六、投射方法的评价	207
统计用表	208
后 记	233
参考文献	234



第一章

教育测量概述

第一节 教育测量的涵义

一、教育测量的对象

从广义来讲,教育测量是根据一定的客观标准对教育领域内的事物或现象进行考核,并依据特定的规则将考核的结果进行数量化描述的过程。从狭义来讲,教育测量是指对学生在经过学科学习和训练之后,所获得的知识、技能的测量。在教育活动中,无论是科学研究的需要,还是改革教学方法、提高教育质量的需要,或者对学生的思想品德、学习成绩、健康状况的测量,对教师的教学效果的测量,对教学经费、物资设备以及行政管理效率



的测量,都属于教育测量的范畴。

教育测量主要是对人的知识、技能、智力、兴趣、气质、性格、品德、人际关系等心理特性的测量,因此,教育测量是心理测量在教育活动中的具体运用。教育测量又称成就测量、学业成绩测量或学科测量,它是按教育测验的规则,对学生掌握某些学科的知识、技能的程度或水平进行数量化的描述过程。当然,广义的教育测量也包括了物理属性的测量,如学生的身高、体重等。

二、教育测量的可能

日常生活中的度、量、衡和自然科学中的各种测试等物理测量,由于可以直接测量,一般能够获得比较准确和可靠的结果。与物理测量相比,教育测量无法进行直接测量,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、主试人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。教育测量具有间接性、多元性和随机性的特点,它比物理测量要复杂、困难得多。

尽管心理属性具有抽象、主观的特点,表面看起来不易捉摸,要想实现直观的测量比较困难,但是它的可测性是毋庸置疑的。任何现象,只要是存在的总有数量,凡有数量的东西都可以进行测量。心理是脑的机能,是客观事物在大脑中的反映,它也有数量上的差异。比如,人的智力有高低,学习成绩有优劣,反应速度有快慢等。虽然这些差异不能用尺子来量、由秤来称,但是它总是会反映在人的各种活动中或表现在人的某种行为中,这样我们就可以通过对人的外显行为的测量来推测他的某种心理属性。

在 20 世纪初,美国心理学家桑代克(Thorndike E. L.) 在 1918 年提出“凡物的存在必有其数量”;美国测验学者麦柯尔(McCall, W. A.) 在 1923 年提出:“凡有数量的东西都可以测量”。

三、教育测量的要素

测量包含三个基本的要素:事物的属性、数字、规则。教育测量的对象是人



的心理属性,如学生的智力、个性、品德、知识、技能、态度、人际关系等。它的存在形式比较抽象,不像物体的长度、重量、体积、时间、温度等事物的物理属性那样可以被人的感觉器官直接把握。但是,它和事物的物理属性一样,在性质和形式上同数字系统的特征具有较高的类似性,如果我们能够合理地用数字来描述事物的属性,赋予它量的意义,就可以通过对数的运算并根据运算结果推测事物的属性。怎样合理的用数字来描述事物的属性对教育测量是至关重要的,这需要以特定的准则和方法给事物的属性分派数字。分派数字所依据的准则和方法就是测量的规则。当测量的其他条件相同,使用不同的规则时会产生不同的测量效果。规则在数字与事物属性之间的对应性或一致性十分重要,它可以体现测量水平的高低。人的心理特性与数字的完全对应是很少的,给所要描述的事物属性分派数字的规则越具有意义,那么事物的属性与数字的一致性也就越高。见图 1.1。

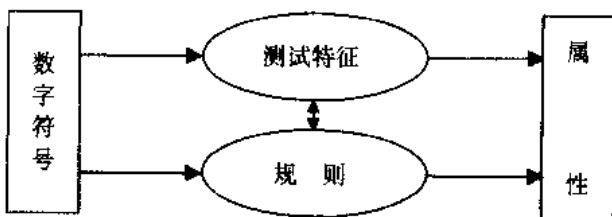


图 1.1 测量要素之间的关系

在教育活动中,教师对学生的学业成就进行评定时,可以用等级评分法,将学业成就最高者评为 5,最低者评为 1,其他依据学业成就的不同程度评为 4、3、2。这里,学生的“学业成就”就是教师所要测量的属性,“1、2、3、4、5”五个等级是描述学生学业成就高低的数字,教师给予学业成就高低不同的学生以不同的等级的这种主观的规定就是规则。在实践中,不同的教师评定的规则总会有着一定的差距,因此,同样的学业表现未必能得到相同的评定等级。

从广义而言,测量是根据法则给事物分派数字。

——史蒂文斯(Stevens, S. S)



四、教育测验的量表

量表是教育测量的基本工具,是表示量数的方法。比如,尺子是度量长短的量表,天平是权衡重量的量表。教育测量中使用的量表多以文字、图形、符号、操作要求等形式出现。

使用数的系统属性来描述事物的属性所达到的程度是不同的,表现为测量的不同水平,每一种测量水平都有对应的测量量表。教育测量常使用的量表有四种:名称量表、等级量表、等距量表、比率量表。

名称测量也叫类别量表,是测量中最简单的形式。它对被测的事物进行分类,并分派给不同的数字或符号,用来描述各类事物的名称。它只具有相同与不同的特性,没有数量大小的含义。用来描述事物的数字仅仅是代表事物的符号,可以区分事物的类别,而不能表示事物程度上的差别,因此,不能将名称量表的数据进行加减乘除四则运算。例如,将学生按性别进行分类,男生用1表示,女生用2表示。如果同时对某门学科喜欢和不喜欢进行分类,喜欢用1表示,不喜欢用0表示,那么,男生喜欢者可表示为11;男生不喜欢者可表示为10,女生喜欢者可表示为21,女生不喜欢者可表示为20。在统计上,名称量表只能计算每个类别的次数(频数)。使用的统计方法属于次数统计,如计算百分比或列联相关,进行卡方检验等。

等级量表也叫位次量表。这种量表将事物的差距按大小、轻重等程度分成若干等级,然后规定某种等级中包含的某种属性,用以测定事物应属的等级。按照某个标准对事物的属性进行分类时,分派给各个类别的数字,不仅具有区分性,还具有等级性(位次性),这些数字能表示事物大小的位次关系。用这样的数字表示的量表叫等级量表或位次量表。例如,将学生的语言表达能力分成甲、乙、丙三个等级。甲等用3表示,乙等用2表示,丙等用1表示。对学生语言表达能力的评定就构成了 $3 > 2 > 1$ 的位次关系。这些数字确定的是事物相等或不等的关系,在不等的情况下,确定的是大于或小于的关系。但是,3与2和2与1所代表的事物之间的差距是不相等的,因此对于量表上的这些数字不能进行加减乘除的运算。这种水平的测量适用的统计方法,除了次数统计外仅限于中位数、百分位数、等级相关系数、肯德尔相关系数等。

有相等单位和人定参照点的量表叫等距量表。它对事物的划分是等距的,即它们的测量单位是等值的,它的参照点是人为指定的,只具有相对意义,没有绝对零点。因此,这类量表上的数值可以作加减运算,却不可以进行乘除运算。



例如,在用摄氏温度计测量温度时,以冰点作为人定参照点,摄氏零度并不意味着没有温度。 9°C 与 6°C 之差等于 6°C 与 3°C 之差,但是,这并不意味着 9°C 是 3°C 的三倍。等距量表使用的统计方法有均数、变差、积距相关系数、T检验和F检验等

比率量表也叫等比量表,是有相等单位和绝对零点的量表。它是最高水平的量表,同时具有以上3种量表的性质以外,还具有等比性。这类量表测量的结果既可以确定一个事物比另一事物大多少,又可以确定大多少倍。因此量表可以进行加减乘除的四则运算。在物理测量中,长度、重量、开氏温度量表(绝对温度量表)都属于比率量表。例如,甲生身高144 cm,乙生身高120 cm,可以说甲生比乙生高24 cm,也可以说甲生身高是乙生的1.2倍。这种水平的测量,除了适用上述一切统计方法之外,还可以使用几何平均数和相对差异量。

五、教育量表的水平

在教育测量中,以上四种量表都会用到。教育测验的分数多属于等级、位次量表。因为测验分数之间只能表明哪个大,哪个小,不能表明大多少,小多少。例如,一次数学测验的平均分数为70分,在这里,60分与70分之间和90分与100分之间,虽然都相差10分,但是它们的差异是不相等的。显然,90分与100分之差要比60分与70分之差难度大得多。另外,测验得零分的学生,并不意味着他在所测验的知识、技能方面为零

由于大量统计方法适用于等距量表,教育测量中,常采取办法实现等级量表数值向等距量表数值的变通。把测验分数作为等距量表来处理也是可行的,这是因为:第一,在统计上可以将总体呈正态分布的测验分数转换成单位相等的标准分数;第二,如果测验编造得较好,特别是对于测验结果两极端分数的微小差异可能反映着巨大差异的这一现象给予注意的情况下,就可以使本来属于等级量表的分数作为等距量表来处理所造成的误差减少到最小程度;第三,如果测验的编程序能使测验分数接近于等距量表,而且对测验分数当作等距量表处理时所得到的结果也确实是有意义的。尽管有许多人对教育测量中运用等距量表的假定有所争论,但人们仍在广泛地运用它,很多标准化测量量表都是承认这个假定的

然而,教育测量不可过分地追求绝对的测量水平。在数据处理过程中,要注意不能把归于类别量数的资料当作等级量数,或把针对等级量数的资料运用适合于等距量数的统计方法。