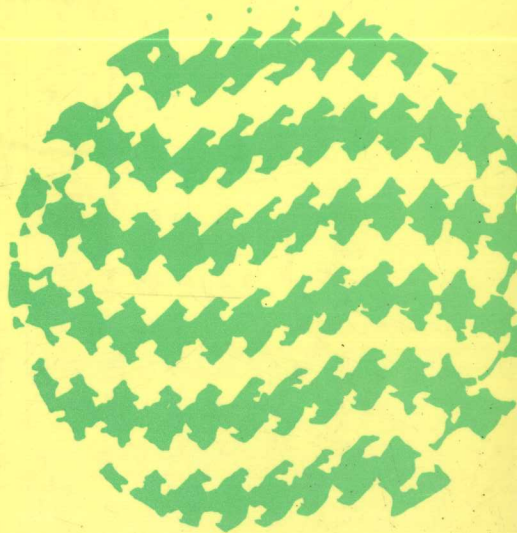


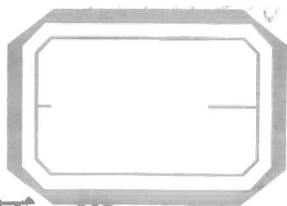
JIAO YU CE LIANG YU PING JIA

教育测量 与评价

张进辅 主编



西南师范大学出版社



教育测量与评价

张进辅主编

西南师范大学出版社

责任编辑:张先金

封面设计:乌 金

教育测量与评价

张进辅 主编

西南师范大学出版社出版、发行

(重庆 北碚)

重庆北碚培萃印刷厂

*

开本:787×1092 1/32 印张:10 字数:220 千

1994 年 11 月 第一版 1997 年 1 月第 2 次印刷

印数:10001—13000

*

ISBN 7-5621-1061-1/G·727

定价:9.80 元

小学教育专业丛书编辑委员会

编 委 名 单

主 任：徐仲林

副主任：李定开 甘大祥 蔡笑岳

编 委：王长楷 李林静 王维城

 泄仲文 李承武 黄培松

 刘乾瑜 沈遂生 周 鸿

 宋乃庆 张进辅 季若霞

说 明

《中国教育改革和发展纲要》指出：“振兴民族的希望在教育，振兴教育的希望在教师。”“要制定教师计划，促进教师特别是中青年教师不断进修提高，使绝大多数中小学教师更好地胜任教育教学工作。到本世纪末，通过师资补充和在职培训，绝大多数中小学教师要达到国家规定的合格学历标准，小学和初中教师中具有专科和本科学历者的比重要逐年提高。”

为了落实《纲要》培养培训师资的决策，我们针对广大小学教师的实际需要和可能，邀请有关理论工作者（专家、教授）、教育管理干部和教育实际工作者，组成三结合的编写队伍，编写了一套小学教育专业的学习丛书，可作为学校教师、教育行政干部业余进修提高和各种在职培训的学习用书，并可作为各级师范学校教师、学生教学或课外参考读物。

这套丛书坚持以马列主义、毛泽东思想及邓小平同志提出的建设有中国特色的社会主义理论为指导，坚持科学性和思想性统一，理论和实际一致的原则，并针对小学教育改革和发展需要，注重实效，讲求学用结合，学以致用，并力求具有大学专科层次标准和较强的针对性、科学性和实践性，以符合业余自学、短期培训、教学参考的要求。

本书由张进辅（西南师范大学）主编，各章撰稿者是：第一、二、四、六章，张进辅；第三、五章，陈红（西南师范大学）；第七、八、九、十、十一章，欧本谷（西南师范大学）；第十二、十三、十五章，乐安中（重庆市教育委员会）；第十四、十六章，黄万琏（重庆市教育科学研究所）。最后主编对全书各章进行了增删、

修改并定稿,其中欧本谷参与了第十四章的修改。重庆市教育科学研究所聂厚德所长曾参加关于本书结构的讨论,四川省高等教育自学考试委员会杨继福主任审阅了全书,在此谨致谢意。

我们在编写这套丛书过程中,自始至终得到省自考办、西南师大党政领导的大力支持和帮助,我们还参考了一些学者专家的有关专著和论文,在此一并致谢。由于时间紧迫,水平有限,丛书的不足之处在所难免。敬请学者、专家、同行批评指正。

编者

一九九四年十一月

目 录

第一章 教育测量与评价概述	(1)
第一节 教育测量与评价的性质.....	(1)
第二节 教育测量与评价的发展历史	(10)
第三节 教育测量与评价的类型和功能	(17)
第二章 教育测验的编制	(26)
第一节 良好测验的特征	(26)
第二节 编制测验的一般步骤	(29)
第三节 测验题目的编制技术	(40)
第三章 测验的信度	(55)
第一节 测量的误差	(55)
第二节 信度的性质	(60)
第三节 信度的类型与估计	(65)
第四节 信度的特殊问题	(74)
第四章 测验的效度	(77)
第一节 效度的性质	(77)
第二节 效度的类型与估计	(80)
第三节 预测效度的应用	(89)
第五章 测验项目的分析	(94)
第一节 项目难度	(94)
第二节 项目区分度.....	(103)
第六章 测验结果的评分与解释	(111)
第一节 测验结果的评分.....	(111)
第二节 测验结果的解释.....	(118)

第七章 教育评价原则	(131)
第一节 教育评价原则概述.....	(131)
第二节 教育评价的基本原则.....	(135)
第八章 教育评价的组织与程序	(145)
第一节 教育评价的组织.....	(145)
第二节 教育评价的程序.....	(155)
第九章 教育评价的指标体系	(162)
第一节 教育评价指标体系概述.....	(162)
第二节 设计教育评价指标体系的原则.....	(167)
第三节 教育评价指标体系的设计步骤和方法.....	(170)
第十章 教育评价的操作方法	(181)
第一节 教育评价信息的采集整理.....	(181)
第二节 教育评价结果的计量.....	(188)
第三节 教育评价结果的解释和质量分析.....	(197)
第十一章 教育评价的心理与调控	(202)
第一节 教育评价心理与调控概述.....	(202)
第二节 主评者的评价心理与调控.....	(204)
第三节 被评者的评价心理与调控.....	(213)
第十二章 学校办学水平评价	(220)
第一节 学校办学水平评价概述.....	(220)
第二节 学校办学水平评价的指标体系.....	(223)
第三节 学校办学水平评价的实施.....	(235)
第十三章 德育工作评价	(242)
第一节 德育工作评价概述.....	(242)
第二节 各级德育工作评价.....	(245)
第三节 学生品德测评.....	(249)

第十四章	教学评价	(258)
第一节	教学工作评价	(258)
第二节	学生学业成就评价	(270)
第十五章	教师评价	(276)
第一节	教师评价概述	(276)
第二节	教师评价的内容	(284)
第三节	教师评价的实施	(288)
第十六章	学生评价	(293)
第一节	学生评价概述	(293)
第二节	学生评价的指标体系	(299)
第三节	学生评价的实施	(301)
附表 1	正态分布表	(305)
附表 2	X^2 分布数值表	(309)
附表 3	积差相关系数检验表	(311)
附表 4	等级相关系数检验表	(312)
附表 5	肯德尔 W 系数显著性临界值	(313)

第一章 教育测量与评价概述

教育测量与评价是现代教育管理与指导的一种主要手段。随着理论上的深入研究和实践中的广泛应用,教育测量与评价已取得了长足的发展。对于教育测量与评价的理论和方法的研究,已经与教育基础理论研究和教育发展研究并列,成为当代教育科学研究的三大领域之一。

第一节 教育测量与评价的性质

一、教育测量与评价的基本概念

教育测量与教育评价既有密切联系又有区别,要准确理解教育测量与评价的概念,必须分析教育测量和教育评价的含义及其相互之间的关系。

(一) 教育测量的含义

教育测量是测量的一种,要理解教育测量的含义,必须了解一般测量的概念。

测量通常是指人们对客观事物进行某种数量化的测定。测量是人们日常生活中普遍存在的现象。例如,人们用天平、秤测定物体的重量,用米尺测定物体的长度,用温度计测试人体的温度等。人们通过测量,对事物进行数量化的确定,从而更好地认识事物和对事物进行比较和分析。

宇宙间的现象,不论是自然现象,还是社会现象,只要是

客观存在的,就必定有数量方面的属性,而凡有数量的东西,就可以进行测量。随着科学技术的发展,人们既可以对物体的长度、重量、温度以及时间、空间、运动等物理特性作出越来越精确的测量,也可以对人的知识、技能、智力、兴趣、情绪、意志、气质、性格、品德等精神特性进行测量。教育现象也是可以测量的。

教育测量就是根据教育目标的要求,按照一定的规则对教育活动的效果加以数量化测定的过程。它主要用于对学生精神特性的测定。学生的学业成绩有好有差,学生的思想品德、兴趣爱好等也有不同,这些精神现象在程度上的差异,就为教育测量提供了可能和基础。教育测量就是要确定学生各种精神特性发展变化的状况,表明其水平、类别、特点及差异等,并把这些结果用数量描述出来。要完成这一测量任务,必须采取一定的手段并按一定的规则进行。按照教育测量学的原理编制并实施的测验,就是测量人的精神特性的最常用的手段。

测量与测验在含义上具有区别,测量包括物质现象和精神现象的测量,测验只限精神现象的测量。精神现象是难以直接测量的。人的能力看不见摸不着,似乎无法测量,但可以采取提问题、布置任务,让他完成,使其能力表现为物质形式,就可以测量了。这种间接测量的方法就是测验。教育测量主要限于精神现象,并以测验作为主要手段,因而教育测量与教育测验在日常使用中区分并不严格。

(二) 教育评价的含义

评价是指对事物或人在价值上的判定。早在北宋时期,《宋史·戚同文传》中就有“市物不评价,市人知而不欺”的记

载。这里的“评价”意指讨价还价，即评论价格。价格是价值的货币表现，因而，也就是判断价值。在日常生活中，事物和人也有价值问题，都需加以判定。可见，评价一词的本义是评论货物的价格，现在泛指判定人物、事物的作用或价值，并已广泛运用于社会生活的各个领域之中，不仅用于物质生产、科技活动和社会组织，而且也越来越多地用于教育活动的各个方面。

教育评价就是根据教育目标的要求，运用系统方法对教育活动的各个方面在社会效用价值上作出科学判定的过程。其实质在于：根据教育目标对于教育活动效果作出价值上的判断。

广义的教育评价是以教育的全领域为对象，涉及教育的各个方面：教育与社会、教育与经济、教育与政治等。具体地说，包括教育体制、教育目标、课程设置、课程标准、教学内容、教学方法、品德教育、教育管理、学校结构、教师工作、师生关系以及社会影响等各方面。狭义的教育评价是以学生为对象，专指在学生的学习和生活领域中，对学生的智能发展、个性形成（包括思想品德）、身体发育等方面的评价。广义的教育评价范围过于笼统，难以把握，而狭义的教育评价范围过于狭窄，满足不了现代教育活动的要求。本书主要涉及小学教育评价，其范围介于广义概念与狭义概念之间。

（三）教育测量与教育评价的关系

教育测量与教育评价是有区别的。测量是用一定的量度或量表来衡量对象，用数量表示出来，不关心测量对象和测值所具有的价值。评价是一种价值判断，是按一定标准对某种对象和现象的价值进行判定，它不仅要知道其数量多少，更重要的是了解其发展变化过程，它对某人某事进行解释、诊断或价值

判断的目的在于改进缺点和不足,以求得最佳效果。

但教育测量与教育评价又有密切联系。测量是为评价提供证据和素材的方法之一或来源之一。一般而言,测量工具的制定、测试和记分属于测量过程,而对结果的解释,便不可避免地要和评价相联系。实际上,二者是同一过程不同阶段的任务。因而,教育测量与教育评价常常又合称为教育测量与评价。

教育测量与评价就是对学生的学习成绩、学习能力、兴趣爱好、思想品德等学习结果以及教育管理、教育教学等工作的效益进行科学测量与评价的过程。

二、教育测量与评价的特点

(一) 教育测量的特点

1. 教育测量的要素和水平

无论是物理特性的测量,还是精神特性的测量,都必须具有两个要素:参照点和单位。

参照点是计算事物数量的起点(即零点),参照点不同,测量结果便无法相互比较。参照点分为两种:一种是绝对零点,如各种度量衡器上的零点,测量长度以“恰恰没有一点长度”作为起点,测量重量以“恰恰没有一点重量”作为起点;另一种是人定零点,即以相对的零点作为计算的起点,如经纬度的计算以格林威治天文台和赤道作为零点,温度的测量以冰点作为零点。在教育测量中,所应用的参照点几乎都属于人定的零点,因为学生的智力、学业成绩和思想品德等的绝对零点难以找到。人定零点的最大限制是,从参照点开始计算的分数,不能以倍数的方式解释。

单位即计算数量的单位。例如,长度以米、厘米等为单位,

重量以吨、千克等为单位,时间以年、月、小时、分、秒等为单位。缺少单位,数量的多少便无法表示,数量化的分析也就无法进行。好的测量单位须具备两个条件:一是有确定的意义;二是有相等的价值。“有确定的意义”是指同一单位在人们的理解中有同样的意义。“有相等的价值”是指单位与单位之间的距离相等。在物理测量中,所有的单位均符合这两个条件。而在教育测量中,所使用的单位严格地讲不是绝对等值的。例如以百分制为单位测量学生的学习成绩,59分与60分之间的价值差异,与100分与99分之间的价值差异很难说完全相等。当然,教育测量必须具有单位,教育测量的单位也是能够进一步研究改进的。

有了参照点和单位,就可以制定量表。所谓量表就是具有参照点和单位的用以测量某种事物以确定其数量的连续体。它是测量的工具,是表示量数的方法。如尺子是度量长短的量表,天平是权衡重量的量表。教育测量所使用的量表则是以文字试题、图形、符号、操作等形式出现的。

量表可分为四级水平:

① 类别量表。这种量表是最简单最基础的量表,它只用数字来代表事物或对事物进行分类,数字没有任何数量大小的含义,当然也没有绝对零点和相等单位。如学生的学号,准考证号;又如男生用1表示,女生用2表示。这些数字只能作为标记和分类,具有区分性,但不具有序列性、等距性和可加性,不能作数量化的分析,也不能进行加、减、乘、除。

② 等级量表。这种量表不仅指明类别,而且也表明不同类别的大小。它依照某种标准对事物排列顺序,如对学生按成绩排列顺序。它无绝对零点也无相等单位,其数字表示等级顺

序,有大小之分,但各数字之间的距离不一定相等。因此,等级量表也不能用加减乘除来处理。

③ 等距量表。这种量表除了能区分类别和排列顺序,各个单位之间的距离也是相等的。其数值可以做加减,但由于没有绝对零点而不能做乘除运算。典型的例子是温度计。等距量表的观察值,可以同时加减乘除同一个常数,而数值之间的关系不变,这样,就可以将一个等距量表上所得到的测值转换到另一个参照点和单位不同的等距量表上去。

④ 比率量表。它是最高水平的量表,既有绝对零点又有相等单位。这种量表在物理测量中常见。其数值可做加、减、乘、除运算。

那么,教育测量属于哪一级水平的测量呢?一般而言,由于教育测量中很难确定绝对零点和获得相等单位,因而教育测量的分数都属于等级量表。从理论上讲,教育测量的统计只应使用等级量表适用的方法,但这样又会使许多有力的统计方法不能运用。于是,在教育测量中常采用变通的方式:① 假设量表上的单位是相等或近似相等,其前提是常态分布的假设,同时在编制量表时严格谨慎。② 采用统计方法把测验分数转换到一个有相等单位的量表上。最常用的方法是把原始分数转换成标准分数,而标准分数则是一个等距量表。所以,教育测量严格讲虽属于等级量表,却常使用等距量表的统计方法。

2. 教育测量的特点

教育测量与一般的物理测量相比,既有共同点也有区别,教育测量的特点有:

① 教育测量是受多种因素影响的。由于教育测量主要是

对人的知识、能力、个性等精神特性的测量,而引起人的精神特性变化的因素很多且不稳定,不同的时间、环境、人的身体状况、情绪等都可能引起人的知识、能力和个性的变化,这就给教育测量带来很大困难。

② 教育测量是间接的。物理测量主要是直接测量,而教育测量中人的行为变化却无法用量具直接测出。教育测量只能通过一个人在活动中(具体讲就是对测验题目)的反应来间接推断人的某种精神特性的状况。

③ 教育测量是相对的。物理测量的结果有相对的,也有绝对的。但教育测量因为没有绝对的零点,便没有绝对的标准,其结果都是相对的。教育测量的结果的等级、高低、强弱、大小等都是与所在群体大多数人的行为或某种人为确定的标准相比较而言的。例如,一个学生的测验成绩是70分,孤立地看这个分数不能说明其意义和成绩高低,只有将其放在群体中与群体的平均水平比较,或者与教学计划规定的教学内容、目标比较,才能确定分数的含意。

④ 教育测量是为教育目的服务的。教育测量是为改进教育工作、提高教育质量、更好地实现教育目的服务的。因而它不能脱离教育目的和教学大纲、教材的要求,随意地制定测验量表。对测量结果的评价,也应依照教育目的来进行。

教育测量与物理测量相比,困难较大,但教育测评的数量化却是科学发展的必然趋势。我们要认清教育测量的特点,掌握教育测量的理论和技术,从而不断提高教育测量的准确性。

(二) 教育评价的特点

1. 教育评价的要素

从系统论的观点看,组成教育评价系统的要素包括价值

目标、人员组织、实施程序、方法技术与质量保证,它们之间的有机组合构成了教育评价系统的基本结构。

价值目标是对评价方向的规定,包括价值、目标、指标三个基本因素。价值是指教育能够满足人的发展和社会进步的需要的程度,包括促进人身心发展的培养价值和促进社会进步的社会价值。目标是指一定的社会对教育价值的具体表现形式的要求,对教育有直接的规范与导向作用。指标是教育评价中体现教育价值与目标的范畴、标准以及各要素间的相互关系。

人员组织是对评价构成的规定,包括机构、角色、职能等基本因素。机构是指评价人员的组成形式,其性质又决定着评价人员的角色地位与职能,并影响整个评价的功能。

实施程序是对评价基本过程的规定,包括评价的方式、步骤和分工。评价方式是评价的实施形式,直接制约评价步骤和各部分人员的分工。

方法技术是对评价的认知过程与工作方式的规定,包括方法论、具体方法与操作技术三个层次。方法论是指教育评价中使用的各种方法技术的理论依据和指导思想。具体方法是指评价中使用的方法与技术路线以及它们之间的相互关系,主要包括评价活动管理方法与评价信息管理方法。操作技术指评价中各个环节的具体操作技术与各种工具的使用。

质量保证是对评价科学性的规定,包括技术方面的再评价与人员方面的心理调控。教育评价是一项技术性很强的活动,需保证每一个环节都顺利完成、保证评价的可靠性和有效性,并在评价过程中不断加以反馈矫正。同时教育评价是一项以人的交往、信息交换为主要形式的活动。受人的心理状态影