

教育测量与评价

胡中锋 李方 编著

JIAO YU CE LIANG YU PING JIA



广东高等教育出版社

教育测量与评价

胡中锋 李 方 编著

广东高等教育出版社

· 广州 ·

图书在版编目(CIP)数据

教育测量与评价/胡中锋,李方编著:—广州:广东高等教育出版社,1999.3

ISBN 7-5361-2315-9/G·732

I.教… II.①胡… ②李… III.①教育-测量 ②教育评价 IV.G449

中国版本图书馆CIP数据核字(1999)第02045号

广东高等教育出版社出版发行

地址:广州市广州大道北广州体育学院内20栋

邮编:510075 电话:87553782 83792953

广东工业大学印刷厂印刷

850毫米×1168毫米 32开本 15.5印张 388千字

1999年3月第1版 1999年3月第1次印刷

定价:24.00元

内容提要

本书内容包括教育测评的基本原理、方法与各类基础教育测评两大部分。作者尽量采用通俗易懂的语言，准确地阐述了教育测评的质量指标、编制方法与结果处理方法，详细介绍了几种主要的教育测评类型。本书能够理论联系实际，始终关注当前教育测评的重点、热点、难点问题，对很多问题都有自己独特的见解，对指导当前的教育测评工作有较为重要的意义。

目 录

第一篇 教育测评的基本原理与方法

第一章 教育测评概述	(3)
第一节 教育测量概述.....	(3)
第二节 教育评价概述.....	(9)
第三节 教育测评的发展阶段	(20)
第二章 教育测量的质量指标	(31)
第一节 信度	(31)
第二节 效度	(41)
第三节 难度	(50)
第四节 区分度	(55)
第三章 教育测验的编制与实施	(59)
第一节 测验目标的确定	(59)
第二节 测验题型的确定	(71)
第三节 命题设计	(77)
第四节 试题的编制	(92)
第五节 测验实施	(95)

第四章 教育评价方案的编制与实施	(99)
第一节 教育评价的基本原则	(99)
第二节 教育评价方案的制订	(103)
第三节 教育评价的实施	(152)
第五章 教育评价的方法	(158)
第一节 教育评价方法概观	(158)
第二节 收集教育评价信息的方法	(167)
第三节 处理教育评价信息的方法	(175)
第六章 教育测量结果的整理、转换与组合	(191)
第一节 教育测量数据的特点与种类	(191)
第二节 教育测量分数的初步整理	(196)
第三节 教育测量分数的转换	(199)
第四节 教育测量结果的组合	(204)
第七章 教育评价心理及其调控	(209)
第一节 教育评价心理概述	(209)
第二节 评价者心理与调控	(217)
第三节 被评价者心理与调控	(228)

第二篇 各类基础教育测评

第八章 语文成绩测验	(238)
第一节 语文成绩测验的历史与发展	(238)
第二节 语文成绩测验的种类	(240)
第三节 语文测验的编制	(257)

第四节	语文高考·····	(262)
第五节	作文评分的方法与误差控制·····	(265)
第九章	数学成绩测验·····	(272)
第一节	数学测验的历史与发展·····	(272)
第二节	数学教学目标·····	(275)
第三节	数学测验的编制·····	(287)
第四节	学生数学能力测验·····	(292)
第五节	中考数学命题·····	(297)
第六节	高考数学命题·····	(299)
第七节	数学竞赛试题的编制·····	(304)
第十章	标准化考试·····	(309)
第一节	主观经验性考试的弊端·····	(309)
第二节	标准化考试的特点·····	(313)
第三节	标准化考试的基本程序·····	(316)
第四节	标准化考试的优缺点·····	(321)
第五节	我国高考标准化考试改革的实践·····	(325)
第十一章	智力测验·····	(335)
第一节	智力测验的意义和种类·····	(335)
第二节	几种重要的智力量表·····	(342)
第三节	智力测验结果的解释·····	(346)
第四节	情绪智力及其内涵·····	(354)
第十二章	学生品德测评·····	(358)
第一节	品德测评概述·····	(358)
第二节	品德测评的指标体系·····	(364)

第三节	品德测评的方法·····	(378)
第十三章	教学工作评价·····	(393)
第一节	教学工作评价概述·····	(393)
第二节	教学工作评价的宏观依据·····	(399)
第三节	教学工作评价的内容·····	(404)
第四节	教学工作评价的指标体系和实施·····	(426)
第十四章	普通中小学办学评价·····	(435)
第一节	办学评价概述·····	(435)
第二节	办学评价的基本内容和指标体系·····	(437)
第三节	办学评价组织实施应注意的问题·····	(482)
主要参考资料 ·····		(485)
后记 ·····		(487)

第一篇

教育测评的基本 原理与方法



第一章 教育测评概述

第一节 教育测量概述

一、什么是测量

(一) 测量的定义

测量的定义有很多，但迄今多数人都能接受的是史蒂文斯(S. S. Stevens)给测量下的定义：“从广义而言，测量系根据法则给事物分派数字。”从字面上看，此定义包含了三方面的特性：法则、事物、数字。

1. 法则

这是测量的标准，也就是测量的依据和准则，即我们根据什么来进行测量。例如，用天平衡量物体的重量，依据的是杠杆原理；用温度计量体温，依据的是物体热胀冷缩的原理。“法则”是一个测量所具备的最重要的特性，对于同一个测量对象按照不同的法则实施测量，其结果是不一样的。比如，测量一个物体的长度，用不同的尺子去量度，其结果肯定是不一样的。

法则也有好坏之分。使用好的法则可以得到比较理想的测量结果，而较差的法则只能导致不准确的测量结果。要寻找和建立较好的测量法则是一件极为困难的事情。有些事物的属性相对简单一些，建立法则可能容易一些；而有些事物的属性可能较为复杂和多变，这就给我们的测量带来了很大困难。但不管是什么事

物的测量，其法则都有一个渐进的过程，即使是对物体长度的测量，也是一个逐步精确的过程。可以这样说，人类的测量史是一个不断建立和完善测量法则的过程。

2. 事物

这是测量的对象，也就是对什么进行测量，更明确地说，就是引起我们兴趣的事物的属性或特征。不同的事物其属性也是不同的。所测量的事物是否具有—致性，这是影响测量结果是否精确的一个重要原因。

3. 数字

这是测量结果的表现形式。测量的结果必须以数字的形式表现出来，这是区别测量与定性评价的一个标志。

数字大多数时候具有量的意义，这样它就是代表一定的数值。这些数值具有自然数的特点，如区分性、序列性或等级性、等距性和可加性。数的这些性质，是我们进行运算的基础。

必须指出，在测量中，我们是根据事物的属性和属性的大小来分派数字的，因此，必须考虑数字是否具有如上自然数系统的功能，还必须考虑到事物的性质和指派数字的原则。

(二) 测量的三个要素

不管是物理属性的测量，还是精神属性的测量，都必须具备以下三个要素：

1. 单位

单位是我们用于计算的标准，如测量物体的重量可用吨、千克、克等为—单位，测量物体的长度可用千米、米、分米、厘米、毫米等为—单位，测量学业成绩可用分数的“分”作单位。单位是测量的基本要求，没有单位就无法测量。

理想的单位必须具备两个条件：一是要有确定的意义，即对同一单位，所有人的理解意义要相同，不能出现不同的解释；二是要有相等的价值，即相邻两个单位点之间的差别总是相等的。

物理测量的单位是比较明确的。但教育测量的单位一般来说是相对的：首先，它的意义比较模糊。例如，对原始分数而言，每一个“1分”在人们心目中的含义是不同的；其次，相同单位不等值，如数学的“1分”不一定等于语文的“1分”（因为很难保证两科的难度完全相等），即使同一次考试的“1分”也不一定完全等值，如选择题的“1分”就不等于填空题的“1分”。因此，一般而言，教育测量的单位是不能直接加减乘除的。

2. 参照点

参照点是计算的起点，参照点不统一，量数所代表的意义就不同，测量的结果就无法进行比较。参照点一般有两种：绝对零点和相对零点。绝对零点里面，“0”表示“没有”的意思，如“0米”就表示没有长度；相对零点里面，“0”不一定表示没有，如温度的“0℃”就不是表示没有温度，而是表示零上与零下的交界点。教育测量的参照点一般也是相对的，“0分”并不表示你对这门学科“一无所知”（可能知道的没有考！），而另一方面，得分即使不是“0分”，也可能表示就是真正意义上的“一无所知”（至少你可以猜测得分！）。

3. 量表

量表一般称为测量的工具，它是具有一定单位和参照点的连续体。如尺子是度量长短的量表，天平是权衡重量的量表。教育测量中所使用的量表多以文字试题的形式出现，也有以图形、符号、操作要求的形式出现的。

由于制订量表的单位和参照点不同，量表的种类也不同。不同量表的精确度是不一样的。将量表从低级到高级排列，则可分为四种量表水平，即类别量表水平、顺序量表水平、等距量表水平、比率量表水平。

类别量表是最低水平的一种测量量表，它只是用数字代表事物和事物的归类，没有任何数量的意义。如对性别进行分类，可

以用 1 表示男性，0 表示女性。类别量表不能作数量化分析，更不能进行加减乘除。

顺序量表比类别量表要精确，其中的数字不仅指明类别，同时指明不同类别的大小等级或具有某种属性的程度。例如，品德评定的等级：优、良、中、差；考试的排名次；甚至考试的原始分数等，都是顺序量表。

等距量表不仅有大小关系，而且有相等的单位和相对的零点。例如，测量温度就是一个等距量表， 60°C 与 50°C 之间的差别等于 40°C 与 30°C 之间的差别；智力测验的分数、标准分数都是等距量表。等距量表可以进行加减运算，但由于没有绝对零点，故不能进行乘除运算。

比率量表是最高级和体现最精确的测量水平的一种测量量表。它既有等距的单位，又有绝对零点，因此可以加减乘除。大多数物理测量量表是比率量表，而教育测量的量表却很难达到这一水平。

二、什么是教育测量

（一）教育测量的涵义

教育测量是根据测量学的原理和方法对教育现象及其属性进行数量化研究的过程。它主要包括对学生内在精神属性的测量，如测量学生的学习成绩、智力水平、品德状况、人格特征等方面。因此，教育测量主要是一种间接的测量。

（二）教育测量的特点

由于教育测量主要是测量学生的内在心理特性，因此，它具有与物理测量不同的特点。主要有以下四个方面：

1. 测量结果的间接性

教育测量一般是间接测量。因为我们无法直接测量学生的内在心理特性，而只能通过其外显的行为，来间接测量其心理活动

的特点与水平。也就是说，我们只能通过学生对测验题目的反应和一些行为表现，运用推理、判断的方法，来间接的测量出他们的知识水平、智力高低和品德好坏。在教育测量中，除了对学生身体素质 and 发育的测量可以直接测量外，其余绝大多数是间接测量。

2. 度量单位的相对性

教育测量的另一个特点是它的度量单位一般是相对的。这在前面已有说明。因此，必须对教育测量的数据进行转换，否则，不能进行代数运算。这将在第六章详细说明。

3. 测量对象的复杂性

教育测量的对象主要是学生的精神属性。首先，它是内在的，不能直接测量。其次是它的多变性。因为，学生的智力、学业成绩、品德等各方面情况是在不断变化、发展的，随着年龄、年级的升高，这些特性将会发生变化。另外，有些主、客观因素，也会影响测量的结果。如对学生的品德进行测量，有些学生可能故意掩饰其本来的品德情况。

4. 测量目的的针对性

任何一个测量都具有明确的目的，教育测量必须为实现教育目的服务，即教育测量是为了掌握学生在学业、智力和思想品德等方面的情况，了解教育和教学的效果，以便更合理地组织教育活动，安排教学内容，选择教学方法，贯彻因材施教，促进学生在德、智、体、美、劳等方面全面发展。我们不能脱离教育目的和教材的要求，随意地制订量表，任意地进行测量。

三、教育测量的误差问题

无论是在天文领域中的测量，或是在物理、生理、心理、教育等领域中的测量，都存在一定的误差。有三种情况，会导致测量的误差：

(1) 测量的仪器和设备。测量工具的精密程度会影响测量

的精确程度，测量时使用的仪器和设备越好，误差就会越小。

(2) 所测量的事物的一致性程度。所测量的事物是否具有-致性，这是影响测量结果是否精确的一个重要原因。

(3) 测量者的因素。比如，由于测量者的技术不熟练，或观察能力不强，一时注意力不够集中等，都会导致测量的不精确。

由于教育测量的上述特点，导致教育测量的结果存在较大的误差。教育测量中所使用的量表(工具)很难把学生的知识、才能、思想一点不漏地全部测量出来，这就给教育测量带来了很大的困难。但是，这并不等于说教育测量是不可能的。因为世界上的事物、现象都有程度上的不同，亦即数量上的差异，既然有数量上的差异，也就可以进行数量化的测定。虽然教育测量不如物理特性的测量那样精确、易行，但绝非不可能。而且，教育测量的长期实践，也证明了教育测量的可能性。目前，我们对有些方面的测量已经达到了较为科学、客观的程度，如智力测验、学业成绩测验等。

四、教育测量学的作用

学习教育测量对于提高教育、教学质量，提高教育管理水平，进行教育研究等都具有十分重要的作用。

(一) 有利于提高教育、教学的水平

教育和教学的最后一个环节就是对成果的检验，同时在教育和教学的过程中，也需要随时了解教育和教学的情况，诊断学生学习中存在的问题，实行因材施教，所有这些方面都需要用到教育测量的理论与方法。例如，如何编制测验、组织考试、评分以及解释分数等，都属于教育测量的内容。因此，正确地利用教育测量，有利于提高教育、教学的水平。

(二) 有利于提高教育管理的水平

加强教育测量，能提高教育管理的科学性，能有效地改变长

期以来形成的凭主观经验衡量教育效果的做法。教育测量可以根据一定的教育目的，编制客观的标准化的量表，进行广泛的测量，作出科学的分析和评价，它对于合理地使用教育经费，科学地选拔人才，决定教育政策，进行教育规划等，都起着重要的作用，是教育管理不可缺少的手段。

（三）有利于发展教育研究的方法

对于教育测量的研究，本身就是教育研究的重要组成部分，教育测量的理论与方法的丰富和发展必然会充实和发展教育研究的方法。很多教育问题的研究要直接运用教育测量的方法，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，既是教育研究的重要课题，也是教育测量的重要研究问题。此外，教育测量的方法还是进行教育研究的不可缺少的重要手段。

第二节 教育评价概述

一、什么是评价

“评价”一词早在 900 多年前我国的北宋时期就已出现。《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”是讨价还价、评论货物的价格。我国 1979 年出版的《辞海》，对“评价”一词的解释是：“评价，评论货物的价格……今亦泛指衡量人物或事物的价值。”评价就是根据某种价值观对事物及其属性进行判断、衡量，亦即对人或物做出好与坏、真与假、善与恶、美与丑、优与劣等等的判断。评价意味着对某一事物的价值给予一般的衡量。

价值观在评价中起着十分重要的作用。人们的价值观不