

八省师范大学合编心理学主干课程系列教材

心理

教育测量

戴海崎 张锋 陈雪枫 主编

暨南大学出版社

·八省师范大学合编心理学主干课程系列教材·

心理与教育测量

主编 戴海崎 张 锋 陈雪枫

华南师大心理系教材建设基金资助出版

暨南大学出版社

编写说明

《心理与教育测量》是华南师范大学心理学系组织南方 8 省师范大学编写的心理学主干课程系列教材之一。本书从测量学基本原理、测验编制技术、知名测验性能 3 个方面总结前人所编教材的经验，力求反映测量研究领域的当代特色。本书主要有如下特色：

1. 辟专章介绍目标参照测验的理论与技术。
2. 介绍了认知心理学的一些测量学新观点。
3. 增加了测验等值、题库建设、教师自编测验等实用技术的介绍。
4. 加强了对我国学者在测量学领域研究活动与成果的介绍。
5. 专章介绍了现代测量理论两个主要分支项目：反应理论与概化理论的新发展。

本书可作为心理学、教育学、社会学等专业的测量课教材，也作为从事心理咨询、考试评价、人员测评等工作的人员的参考书。

全书体系是在主编提供的框架基础上集体讨论而定的，编写分工如下：

张 锋：第一、八、十四章

罗黎辉：第二章

杨志明：第三、四、五章

龙文祥：第六、七章

戴海崎：第九、十一、十七章

陈雪枫：第十、十二、十三章

龙立荣：第十五、十六章

本书初稿由戴海崎、张锋分工审阅提出修改意见，最后由戴海崎统校定稿。

本书的编写得到心理学主干课程教材编委会的指导，特别是受到了编委会主任莫雷教授的格外关心；江西师大、云南师大、华中师大、湖南师大、安徽师大等校均对本书的编写出版给了很大帮助，在此一并深表谢意。我们还得感谢暨南大学出版社对本书的扶持。在编写中，我们参考了国内外大量资料，有些还作了摘引，在此也向这些作者表示感谢。

编者

心理学主干课程系列教材

编委会成员名单

主任：许尚侠

副主任：莫 雷 刘华山 白先同

编 委：（按姓氏笔划）

王守恒	史健生	白先同	许尚侠
沙毓英	刘英才	刘华山	
沈家鲜	李 铮	李山川	李巨才
陈沛霖	杨鑫辉	郑和钧	
罗黎辉	胡启先	胡晓莺	莫 雷
傅 荣	彭运石	漆书青	

目 录

第一章 心理与教育测量概论	(1)
第一节 一般测量概述.....	(2)
第二节 心理与教育测量的性质	(10)
第三节 心理与教育测验的种类及其功能	(18)
第四节 心理与教育测量工作者的素质要求 及道德准则	(26)
第二章 心理与教育测量的产生与发展	(33)
第一节 中国古代的心理与教育测量	(34)
第二节 现代心理与教育测量在西方国家的 产生和发展	(40)
第三节 现代心理与教育测量在中国的发展	(51)
第三章 经典测验理论的基本假设	(58)
第一节 心理特质及其可测性假设	(59)
第二节 测量误差及其来源	(61)
第三节 真分数及其有关的假设	(64)
第四章 测量信度	(68)
第一节 信度概述	(69)
第二节 信度的估计方法	(72)

第三节 提高测量信度的方法	(82)
第五章 测量效度	(88)
第一节 效度概述	(89)
第二节 效度的估计	(92)
第三节 提高测量效度的方法	(103)
第六章 测验的项目分析	(108)
第一节 测验的难度	(109)
第二节 测验的区分度	(116)
第三节 猜测问题与猜测率	(126)
第四节 多重选择题的项目分析	(131)
第七章 测验常模	(135)
第一节 分数转换	(136)
第二节 分数合成	(147)
第三节 常模编制	(153)
第八章 心理与教育测验的编制与实施	(168)
第一节 编制心理与教育测验的基本程序	(169)
第二节 测验的实施	(181)
第九章 测验等值	(194)
第一节 测验等值概述	(195)
第二节 测验等值计算的基本方法	(204)
第三节 常用测验等值设计介绍	(209)
第十章 目标参照测验	(218)
第一节 概述	(219)
第二节 目标参照测验的项目分析	(221)
第三节 目标参照测验的信度与效度	(231)
第四节 测验分数的解释——分数分界点 的确定	(236)

第十一章 学绩测验	(243)
第一节 学绩测验概述	(244)
第二节 标准化学绩测验	(250)
第三节 教师自编课堂测验	(263)
第十二章 能力测验(上)	(272)
第一节 智力测验的一般问题	(273)
第二节 个体智力测验	(288)
第三节 团体智力测验	(300)
第十三章 能力测验(下)	(305)
第一节 能力倾向测验	(306)
第二节 特殊能力测验	(318)
第三节 创造力测验	(325)
第十四章 人格测量	(332)
第一节 人格测量的一般问题	(333)
第二节 自陈量表	(337)
第三节 投射测验	(353)
第十五章 其他心理与教育测验	(362)
第一节 焦虑测验	(363)
第二节 兴趣测验	(371)
第三节 态度和品德测量	(381)
第十六章 测量的综合应用	(395)
第一节 测量在心理咨询中的应用	(396)
第二节 测量在人事测评中的应用	(403)
第三节 测量在教育评价中的应用	(412)
第十七章 心理与教育测量理论的新发展	(422)
第一节 项目反应理论简介	(423)
第二节 概化理论简介	(441)

附 录

附录一 心理测验管理条例（试行）	(455)
附录二 心理测验工作者的道德准则	(457)
参考文献	(459)

第一章 心理与教育测量 概论

本章提要：

- 测量的基本性质及其要素
- 测量量表的四种水平
- 心理与教育测量及其理论基础
- 心理与教育测量的量表和测验
- 心理与教育测量科学研究和实际工作中的意义
- 心理与教育测量工作者的素质要求与道德准则

心理与教育测量是我国各大学心理专业和教育专业学生必修的重要的专业课，它在心理科学、教育科学的基础学科和应用学科之间起着一种中介作用。一方面，它是基础心理学科和基础教育学科的深化，是从事基础理论研究的方法课；另一方面，它又是应用心理学科和应用教育学科的基础，是从事实际应用研究的工具课。在本章里，我们将讨论心理与教育测量的若干基本概念和基本问题，以便为学习以后各章的具体知识提供一个基本的框架。

第一节 一般测量概述

一、测量及其种类

测量（measurement）是人类生产和生活中普遍存在的现象。农业生产要丈量土地面积，工业生产要测定产品的技术指标，地质勘探要测定海拔高度和地质指标，医疗工作要测定人体的生理指标，教育工作要测定学生的学业成绩。至于科学研究中的测量活动就更加普遍，也更加严格了。那么，究竟什么是测量呢？

简单地说，测量就是依据一定的法则使用量具对事物的特征进行定量描述的过程。

所谓“一定的法则”，是指任何测量都要建立在科学规则和科学原理基础之上，并通过科学的方法和程序完成测量过程。例如，用杆秤测量物体的重量，依据的是物理学上的杠杆

原理；用温度计测量温度，依据的是热胀冷缩原理。有的测量依据的法则比较科学和完善，测量的结果比较准确、可靠，而有的测量依据的法则比较粗糙和欠成熟，测量结果的准确性和可靠性较差。有的测量依据的法则的操作比较直观和简单，一般的人不需要经过专业训练就很容易掌握，而有的测量所依据的法则的操作复杂，需要经过专门训练才能逐步掌握。

所谓“事物的特征”，是指所要测量的事物的特定属性。例如，物体的重量、长短、高矮；物体运动的速度；物体中某些特定成分的含量等等。这些不同的特征就是测量的特定对象。一种事物有各种各样的特征，对不同的特征要用不同的测量工具、依据不同的法则进行测量。有些事物的特征直观明显，具有外显性（如物体的重量、长度等），所以在测量中容易被确定，测量的结果具有无可争辩性，容易被所有的人认同和接受；而有些事物的特征不那么外露，具有内隐性（如人的智力水平、性格特点等），所以在测量中难以准确界定，测量的结果不容易获得清楚的解释，因而也不容易取得多数人的认同和接受。

所谓“量具”，是指测量中所使用的工具。例如，重量测量中的杠秤、电子秤，长度测量中的木尺、皮尺，体温测量中的体温计等等。不同的测量要用不同的量具，不同量具所使用的单位和参照点也不同。

所谓“定量描述”，是指任何测量的结果总是对事物特征的量的确定。虽然有时人们把诸如“1”代表男，“0”代表女这样的做法也叫做测量，但这里的数字仅仅是一种符号，并不是数量。所谓“数量”不仅指事物特征的符号，而且指一种有序的量。数量具有4个特征：一是它的区分性，即一个数（如“1”）不同于另一个数（如“2”）；二是它的序列性，即 $1 < 2 < 3 < 4 \cdots$ ；三是它的等距性，即 $2 - 1 = 1$ ， $3 - 2 = 1$ ，所以，

$2 - 1 = 3 - 2$ ；四是它的可加性，即一个数加另一个数产生第三个数。数的这些特点是一切数学运算的基础，同样，也正是数的这些特点使得对事物特征的差异的测量成为可能。有的测量对事物的特征定量描述的精确度高些，而另一些测量对事物的特征定量描述的精确度差些。测量的精确度既与测量对象的性质有关，也同测量时所用的工具有关。

首先，测量的精确度决定于测量对象本身的性质。我们可根据测量对象的性质把它分为3种类型：①确定型，即在一定条件下，事物的量保持恒定不变。例如，物体的长度和重量，只要物质的温度不变，受力状况不变，其长度也就不会改变；只要物体在地球表面的水平位置和垂直高度不变，其重量也不会改变。②随机型，即事物的量随机改变。例如，人的短时记忆的容量，尽管实验者在实验过程中每次向被试呈现刺激的条件保持恒定，但每次测量的结果总是存在差异，不过，这种差异又总是保持在一定的范围内，量的改变趋势也呈现出一定的规律。③模糊型，即事物的量本身就是模糊不定的，难以获得确定的量。例如，对人的性格特征，尽管人们习惯于用热情奔放或冷若冰霜等词汇来描绘，而且也能够区分出两个同是热情奔放的人在程度上的差别，但这种差别的量却是很模糊的。显然，对确定型的事物进行定量描述要比对随机型和模糊型的事物进行定量描述要容易得多，因此，测量的精确度也要高得多。但是，即使对确定型的事物也不能做出绝对精确的描述，在任何测量过程中都会有误差存在，所不同的是误差的大小而已。

其次，测量的精确度决定于测量工具（量尺）的精密性。不言而喻，使用技术上完善的测量工具比使用技术上粗糙的测量工具，其测量结果要精确得多。对于长度的测量，用皮尺测量比之用脚步测量，其结果要精确得多；而用激光测量比之用

皮尺测量，其结果又要精确得多。同样，对于重量的测量，用杆秤测量比个人的主观估计，其结果要精确得多；而用电子秤测量比用杆秤测量，其结果又要精确得多。因此，尽可能使用技术精密的测量工具，是保证测量精确度的重要条件。但是，不论使用何等精密的测量工具，实际测量中仍然会有误差存在，所不同的也仅是误差的大小而已。测量学的目标之一是设法尽可能把误差减少到最低程度，而不可能完全消灭误差。

测量技术被广泛用于工农业生产、商业活动、科学研究和人们的日常生活领域。根据测量对象的性质和特点，可以将各种不同形式的测量大致分为4种类型：①物理测量：即对事物的物理特征的测量。如长度测量、重量测量、面积测量、速度测量等等均属物理测量。②生理测量：即对机体生理特征的测量。如对动植物各种化学成分含量的测量，对人体各种生理机能的测量等等均属生理测量。③社会测量：即对社会现象的测量。如在人口普查、经济统计、民意调查中所使用的测量技术均属社会测量。④心理测量：即对人的心理特征的测量。如智力测量、人格测量、职业兴趣测量、态度测量等等均属心理测量。狭义的教育测量主要指对学生学业成绩和知识水平的测量，此时，教育测量可以被包括在心理测量的范畴之内。但是，广义上的教育测量不仅包括对学生学业成绩和知识水平的测量，而且包括对教育领域中其他教育现象的测量。如对教师教学水平的测量、对整个学校办学质量的测量、对学校管理水平的测量等等。此时，教育测量当属社会测量的范畴。本书所使用的教育测量是指狭义的教育测量，但为了与其他心理测量有所区别，将教育测量与心理测量这两个术语并列使用。

二、测量的基本要素

任何测量都必须具备两个基本要素，即测量的参照点和测量的单位。

(一) 测量的参照点

从根本上说，测量是确定特定事物的特定特征的数量。因此，在测量工作中，必须有一个量的起点。这个起点就叫作测量的参照点。要使两个测量结果能够相互比较，必须使这两个测量建立在同一个参照点上。因为参照点不同的两个测量，其结果的意义完全不同，没有可比较的共同基础。

参照点有两种：一种是绝对参照点，即以绝对的零点作为测量的起点。如长度测量和重量测量就是建立在以绝对的零点为参照点的基础上的测量。这个绝对的零点的意义就是“无”，即没有重量或没有长度。以此为测量的起点，去确定某种事物有多重或有多长。另一种是相对参照点，即以人为确定的零点为测量的起点。如对地势高度的测量，就是以海平面为测量的起点。此时，人们假定海平面的高度为“零”，然后去确定陆地高出海平面多少，再如对气温的测量，是以水的冰点为测量的起点。此时，人们假定水刚刚能够结为冰的温度为“零”，然后确定气温高于或低于“零”多少度。

最为理想的测量参照点当然是绝对参照点，因为它的意义最为明确。但在许多情况下，人们难以找到绝对参照点，所以必须改用相对参照点。采用相对参照点为测量起点的测量结果只能进行加减运算，而不能进行乘除运算。因此，它的两个值

之间没有倍数关系。例如，在智力测量中，假定甲的智商为 100，而乙的智商为 50，我们不能说甲的智商是乙的智商的 2 倍，而只能说甲的智商高出乙的智商 50。

（二）测量的单位

测量的第二个基本要素是它的单位。不同测量所用的单位是不同的。长度测量的单位是毫米、厘米、分米、米等等，而重量测量的单位是毫克、克、千克、吨等等。理想的测量单位应当具备两个条件：一是要有确定的意义，即对同一单位，所有的人的理解都是相同的，不允许作出不同的解释。例如，所有的人对重量单位“千克”的解释都是一样的，没有歧义。二是要有相等的价值，即第一个单位与第二个单位之间的距离等于第二个单位与第三个单位之间的距离。例如，30 公斤与 20 公斤之差等于 40 公斤与 30 公斤之差。但是，在某些情况下，要具备这两个条件是相当困难的。例如，教育与心理测量中的单位就往往难以达到这个要求，它远没有其他测量中使用的单位成熟和完善。这一点我们在后面还会谈到。

三、测量的量表

要测量某一特定事物的特定特征的数量，必须首先选择一个具有确定单位和测量参照点的数字连续体，将欲测量的特征与这个连续体相比照，确定它的位置，看它距参照点的远近，就会得到该特征的一个测量值。这种能够使事物的特征数量化的数字的连续体就是量表（scale），制定量表的单位和参照点不同，就会编制出不同的量表；不同的量表具有不同的测量水

平，相应地测量的精确度也不同。根据测量的不同水平以及测量中使用的不同单位和参照点，我们把测量量表分为4种。

（一）命名量表

命名量表（nominal scale）是最低水平的测量量表，它只是用数字代表事物或用数字对事物进行分类。在这种情况下，数字只是事物的符号，而没有任何数量的意义。因此，运用命名量表时不能作常用的数量化分析。如我们不能说6号学生 > 5号学生 > 4号学生，当然也不能进行代数运算。正因为如此，有人认为运用命名量表进行的测量不能算作是真正意义上的测量。

命名量表又可细分为两种形式：一是名称量表，即用数字指代个别事物。如用数字给学生或运动员编号。二是类别量表，即用数字指代事物的种类。如用1、2、3、4、5……分别代表不同的职业。

适合于对命名量表进行统计分析的统计方法有百分比、次数、众数和 χ^2 检验。

（二）顺序量表

顺序量表（ordinal scale）是次低水平的测量量表，它不仅能够指代事物类别，而且能够表明不同类别的大小、等级或事物具有某种特征的程度。各种比赛、评估中的名次排列就是一种典型的运用顺序量表进行的测量。例如，在各种体育比赛中，我们通常取前3名，分别用1、2、3代表，那么，我们就可以说， $1 > 2 > 3$ 。这表示，第1名的水平高于第2名的水平，第2名的水平又高于第3名的水平。这种按照事物的大小、等级、程度而排列数字的量表就叫顺序量表。

在顺序量表中，数字只表示等级、大小和程度的顺序，它