

教育和心理测量



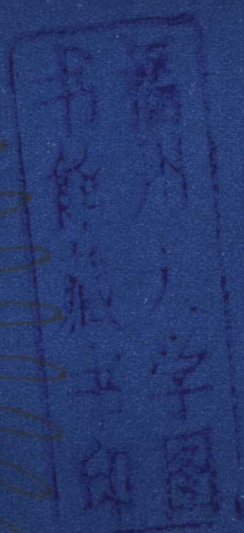
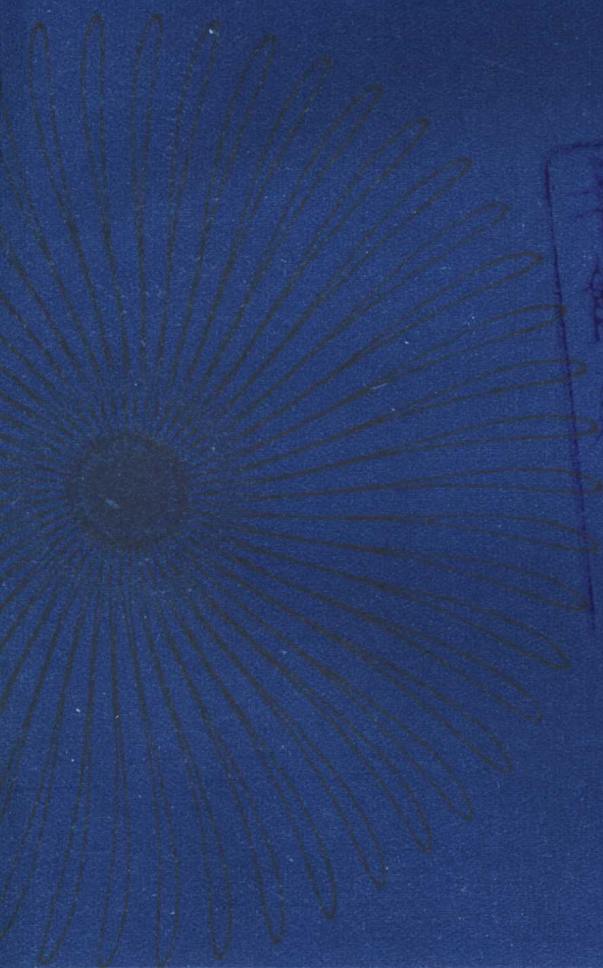
JIAOYU
HE
XINLI
CELIANG

余嘉元
编 著

• 江苏教育出版社

教育和心理测量

江苏教育出版社



教育和心理测量

余嘉元 编著

江苏教育出版社出版

江苏省新华书店发行 淮阴新华印刷厂印刷

开本 850×1168 毫米 1/32 印张 12.25 插页 2 字数 268,000

1987 年 1 月第 1 版 1987 年 1 月第 1 次印刷

印数 1-5,250 册

书号: 13351·028 定价: 2.60 元 (压膜)

责任编辑 孙峪峨

前 言

随着国家教育事业的发展，人们对于教育和心理测量的要求越来越高。作为合格的教师，必须能及时地了解自己的教学效果；作为教育行政部门的管理者必须对学校的教育质量有清晰的了解；作为教育科学工作者，必须能对教学过程中的各种变量进行准确的描写和控制；作为家长，也应该知道孩子测验分数所代表的真实意义。要解决这些问题，就必须掌握教育和心理测量的基本原理和方法。

本书是根据这种需要，在教育系学校教育专业、学校管理专业和心理学的研究生班试用的基础上，经修改后定稿的。可供各级各类师范院校和教育学院作为教材，每周四学时，一学期可教完。

本书有一定数量的习题，书末附有习题的答案，便于自学。广大中小学教师、教育行政管理工作者，只要掌握了本书的内容，就可以在自己的工作中加以运用。

教育和心理测量的基本原理，非但适用于教学工作，而且还可用于人员的功能测评、选拔和安置，故本书还可供各企事业单位的有关人员参考。

在本书的编写过程中，得到南京师范大学教育系领导、专家、教授和其他同志的关心，得到国内许多同行的鼓励，谨此表示衷心的感谢！由于编著者水平有限，缺点错误在所难免，希望广大读者给予批评指正。

编著者

1986年1月

目 录

第一章 测量的基本问题	1
第一节 测量的定义和要素	3
第二节 测量的方法和量表	5
第三节 教育和心理测量的性质	20
第四节 评价	25
习题	34
第二章 测验的基本问题	35
第一节 测验的定义	35
第二节 测验的种类	37
第三节 良好测验的条件	43
第四节 测验的功能	47
第五节 编制测验的基本知识	53
第六节 测量误差、信度和效度	62
习题	72
第三章 测验分数的处理和解释	73
第一节 次数分布表和次数分布图	74
第二节 集中量数和差异量数	78
第三节 相关系数	88
第四节 正态分布	97
第五节 测验分数的解释	107
习题	144

第四章 项目分析	148
第一节 常模参照测验的项目分析	149
第二节 标准参照测验的项目分析	174
第三节 速度测验的项目分析	176
第四节 项目分析的作用和试题的改进	178
习题	184
第五章 信 度	185
第一节 信度的理论	185
第二节 信度系数	187
第三节 测量标准误	212
第四节 标准参照测验和速度参照测验的信度	220
第五节 影响信度的因素	223
习题	231
第六章 效 度	234
第一节 效度的定义、种类和性质	234
第二节 内容效度	236
第三节 效标关联效度	242
第四节 结构效度	259
第五节 标准参照测验的效度	265
第六节 影响效度的因素	267
第七节 效度的应用	274
习题	283
第七章 测验的编制	285
第一节 编制测验的基本原则和步骤	285
第二节 客观测验的编制	292

第三节	论文式试题的编制·····	309
第四节	操作测验的编制·····	316
第五节	观察评分法及其工具的编制·····	323
第六节	测验的编辑和实施·····	335
习题	·····	339
习题答案	·····	340
参考文献	·····	342
附表一	正态分布表·····	343
附表二	积矩相关系数检验表·····	348
附表三	等级相关系数检验表·····	349
附表四	χ^2 检验表·····	350
附表五	T分数的计算辅助表·····	352
附表六	弗拉南根项目分析表·····	353
附表七	肯德尔和谐系数显著性考验时S的临界值·····	381
附表八	t检验表·····	382

第一章 测量的基本问题

世界上的万事万物都是有差异的，无生命界是如此，有生命界也是如此。如何精确地对这些差异进行描述，千百年来一直是人类孜孜不断地进行探索和研究的课题。经过一代又一代科学家的努力，我们现在已经可以相当精确地对长度、重量、温度、电压、电流、亮度和时间等物理量进行测量了。物理学家制定了一整套的测量标准、单位制和测量方法，并创造了许多测量的工具。所有这一切，对于科学和技术的发展，都起了极为重要的作用。同样，生物学工作者和医务工作者在对生物体的许多生理量的测量方面，例如，对血压、心跳、呼吸、血小板、红血球等等的测量，也取得了相当令人满意的成就。好奇和探索是人类的本性，科学工作者在物理量和生理量的测量方面所取得的成就，激励着人们去探索对心理量的测量。我们常常说某些人聪明过人，而另外一些人则呆如白痴，这就是对人们智力的初浅描写。那如何对智力的高低作进一步的定量分析呢？这是许多人所关心的问题。同样，我们还常常会说某些人的知识渊博，而另外一些人则知识很贫乏；某些人的数学能力很强，而另外一些人则毫无数学头脑；某些人喜欢合群，善于交际，而另外一些人则喜欢独自一人，性格孤僻；某些人在团体中喜欢担任首领的角色，发号施令，支配别人，而另外一些人则习惯于服从权威，很少表现出独立性。自然界和社会造就了形形色色，千差万别的各种各样的人，那怎么对这些人们的心理特点进行测量呢？这是摆在心理学工作者和教育工作者面前的一个重要课题。

今日的地球是一个充满决策的世界，每个人都经常面临着需要作出决策的境地。决策是有各种层次的，有的是关于国家方针、政策方面的，有的是关于企业经营方向、管理方法方面的，有的则是关于人的方面的。我们所要研究的是关于人的决策方面的问题，特别是在教育和心理领域中的问题。

作为一个学校领导，往往要对课程的设置和安排作出决策。例如，要开设些什么课程？除了上级规定的课程外，为了开阔学生的知识面，培养学生的创造力，还要增加些什么课程或课外活动？为了适应社会的需要，是否要开设电子计算机和 BASIC 语言课？各门课程的时间怎么安排？这些都是经常要考虑到的。

作为一个对工作负责任的教师，往往对教学方面的许多问题，例如，这学期是用原来的教学方法还是进行一些改变？教学的进度是否要作些调整？对于某些后进生如何进行帮助？经常面临着决策。

另外，作为一个单位，招收哪些人作为新学生或新工人？新招收进来的人员如何进行分班或分组？这些也都是决策问题。

最后，作为个人，在一生中会遇到许多生活的十字路口。例如，当中学毕业时，往往要作出决策，究竟是升学还是工作。如果是升学，就要进一步决定上哪所大学，学习什么专业；如果是工作，就要进一步决定是在目前居住的城市找工作呢，还是到边疆农村去，以及准备搞哪一项工作。这些问题，是每一个人都可能遇到的。

那么，如何作出正确的决策呢？一般说，应该在决策前尽可能多地获得有关的信息，以及掌握科学的决策方法。获得可靠而有效信息的方法有好多种，测量是其中最主要的方法之一，因此，对于决策过程来说，测量是极为重要的一个环节。对于教育

和心理学工作者来说，教育和心理测量是搞好日常工作和开展科学研究的重要手段。

经过许多年的努力，人们已初步建立了教育和心理测量的理论和方法。

第一节 测量的定义和要素

一、测量的定义

我们要研究教育和心理领域中的测量，就必须首先了解一般的测量的定义。谈到测量，人们总会联想到各种各样的仪器，例如测量长度的米尺，测量重量的天平，测量时间的钟表以及测量温度的热电偶，等等。但是，这些变量的测量并不一定要用仪器，例如有经验的钢铁工人可以根据炉火的颜色估计炼钢炉中的温度，音乐家可以用耳朵分辨出不同频率的声音，优秀的售货员可以根据手感确定货物的重量。这些都是仅仅依靠人的感官和大脑，而不利用仪器进行测量的例子。从以上可以看出，仪器的使用并不是测量的最基本特征。那么，测量的最基本特征究竟是什么呢？是将事物进行区分。因此，人们把对事物进行区分的过程称为测量。严格地讲，这种对事物进行区分的过程必须是按照一定法则的，区分的结果必须是能够用数学的方式进行描写的。所以，我们把测量定义为：按照一定的法则，用数学方法对事物的属性进行描写的过程。

二、测量的要素

根据测量的定义，我们可以看到，任何一个测量都包括：测量的对象——事物的属性；测量的工具——某种法则；测量的结果——某种数学表达形式（在很多情况下是用实数域中的数字来表

示)等三个要素。

下面，我们举几个例子来说明测量中的这三个要素。

例1 测量人的体重。

测量的对象（事物的属性）：人的体重。

测量的工具（某种法则）：要求被测量者穿戴尽可能少的衣着，静止地站立在磅秤上，读数者的视线要和磅秤的刻度盘相垂直。

测量的结果（数字）： $\times \times$ 公斤。

例2 测量学生的英语水平。

测量的对象（事物的属性）：学生的英语水平。

测量的工具（某种法则）：用预先编制好的英语试卷，按照测验的要求进行，包括规定测验的时间，不准作弊，等等。

测量的结果（数字）：测验的分数。

三、测量的精确程度

在测量的研究中，人们非常关心的一个问题就是某种测量的精确程度究竟如何。因为在一般情况下，测量的结果越是精确，它的价值就越高。

测量的精确程度首先是和测量对象本身的属性有关。我们可以把测量对象的属性分成三种类型：第一种是确定型，即在一定的条件下，它是保持恒定不变的。例如，物体的长度和重量，只要物体的温度不变，受力状况不变，长度也就不变；只要物体在地球表面的水平位置和垂直高度不变，重量也就不变，这些都属于确定型的问题。第二种是随机型，即事物的属性是随机变化的。例如，人的记忆广度，尽管各种条件都保持恒定，但每次测量的结果还是会有所差异。第三种是模糊型，即事物的属性本身是模糊不清的。例如，我们认为某人是热情奔放，而另一个人是

冷若冰霜，那么，什么叫热情奔放，什么叫冷若冰霜呢？这些都是模糊的概念。

显然，对于随机型和模糊型的事物属性，我们是不可能用数字作精确描写的。那么，对于确定型的事物属性是否能作出绝对精确的测量呢？也是不能的，因为在每次测量中都会受到一些无关变量的影响。例如，用千分卡测量一个工件的直径，如果测量一百次，那么这一百次的结果并不会完全相同，也就是说，总是有误差存在。

测量的精确程度还和测量所用的工具有关，假如我们要测量两个建筑物之间的距离，用米尺测量就要比用步长测量精确得多，而用激光测量又比用米尺测量要精确得多，就是激光测量也还是有误差。随着科学技术的发展，今后还会有更先进的测量工具诞生。但是，不论测量工具多么先进，总还是不能对事物的属性进行绝对精确的测量。

既然测量总是有误差存在，那么，测量究竟要精确到多大的程度才能令人满意呢？这就要根据测量的目的来决定。例如，在工程技术中，对工件的测量如果有 0.1 厘米的误差，就显然是不允许的了。但是在教育和心理测量中，如果有千分之一的误差，人们并不在乎。例如，测量学生掌握英语词汇的水平，甲生能默写 90.1% 的词汇，乙生能默写 90.2% 的词汇，对于他们的差异，我们并不很重视。一般来说，在教育和心理测量中对精确程度的要求比工程技术中的要求低。

第二节 测量的方法和量表

一、测量的方法

测量的方法有很多种，如实验法、观察法、测验法。在教育

和心理测量中对测验法用得较多。

根据测量能否直接测到所需的事物属性，可以将测量分为直接测量和间接测量两类。

直接测量可以用工具直接测得事物的属性，例如，用天平称物体的重量，用尺子量物体的长度。

间接测量是根据测量的结果去推测事物的属性。例如，水银温度计是根据水银柱的长短来推测温度的高低；在英语测验中是根据考生所得到的分数来推测英语水平如何。在教育 and 心理测量中，基本上都是间接测量。

二、量表

测量的本质是根据某一规则将事物数量化。数量化的价值是在一个定有参照点和单位的连续体上表现出来的，这个连续体称为量表。下面介绍一下关于量表的两个基本概念。

1. 单位

单位是计量事物的标准量的名称，如果没有单位，则数量的多少和大小就无法表示。好的单位必须满足两个条件：

(1) 具有确定的意义 同一单位在所有人的心目中都有同一意义，不允许有不同的解释。这个问题在物理学中很好解决，如：1 公斤、1 米、1 小时，所有人的理解都是一致的。但在教育和心理的测量中，这个问题就有点困难，特别是问答题、论文式题目，什么样的回答算满分，什么样的回答应该扣 1 分或扣 2 分，就不太好十分精确的决定。

(2) 单位的距离等值 在物理学里这个问题比较好解决，例如，有 A、B、C、D 四个物体，它们的重量分别为 30 公斤、40 公斤、90 公斤、100 公斤，那么可以肯定地讲 A、B 之间的重量差和 C、D 之间的重量差是相等的。但是，在教育 and 心理测量

中，问题就不那么简单了。例如，有 A、B、C、D 四个学生，在一次测验中，他们的成绩分别为 30 分、40 分、90 分、100 分，那么，我们能否说 A、B 之间的学业水平之差和 C、D 之间的学业水平之差相等呢？显然，我们是无法肯定这一点的。因为作为分数单位的题目，其难度单位是否等值，事先并没有作过检查。因此，首先要有个标准来确定分数的标准，也就是说要有科学的依据来确定考试中每个题目应该得到的分数。在目前所用的考试方法中，非但分数单位不等值，而且没有确定的值，这就很成问题，因为等值单位是作为量具的最基本要求。

2. 参照点

所谓参照点，就是计算的起点，也称为零点。若参照点不同就无法进行直接比较，例如，摄氏温度计将冰点定为零度，而华氏温度计将冰点定为 32 度，因此 10°C 和 10°F 就不能直接进行比较，而必须经过换算才能进行。

参照点可以分为两种：

(1) 绝对零点 例如：长度、重量都是有绝对零点的。

(2) 相对零点 例如：山脉的高度以海拔 $\times \times \times \times$ 米表示，就是以海平面作为相对零点的。摄氏温度计是以冰点作为相对零点的。这种相对零点都是人为确定的。

教育和心理测量中的参照点一般都是相对零点，例如，在某一次测验中得零分的学生未必对这方面的知识一点也没有。

在相对零点量表上的得分，只能表示差异的大小，而不能表示倍数。例如，有甲、乙两个学生，在一次测验中，甲考了 80 分，乙考了 40 分，我们只能说甲比乙的成绩高 40 分，而不能说甲的成绩是乙的成绩的两倍。

三、四种不同的测量水平和量表

根据对事物属性的数量化描写的不同水平，可以将量表分成四类，若将它们从低级到高级依次排列，就是类别量表、顺序量表、等距量表和等比量表。各种量表的性质并不完全相同，每一种较高级的量表都具有较低级量表的全部特性，并且还有自己独特的性质。

1. 类别量表

它也称为名称量表或称名量表。

所谓类别量表，就是根据某一特点，对两个或两个以上的对象进行分类，并用一定的数字（或符号）加以表示。

例如，我们用计算机进行人口统计时，可以根据性别，将人口分成两类，并用“1”表示男性，用“0”表示女性。

交通管理部门规定江苏省的汽车用“09—”表示，上海市的汽车用“08—”表示。

在电子线路中，用“1”表示高电位，用“0”表示低电位。

在上面各个例子中，数字只起到某种名称符号的作用，它只是对事物的属性进行归类区别，而没有序列的意义。因此，我们不能认为，江苏省的汽车用“09—”表示，上海市的汽车用“08—”表示，9大于8，江苏省的汽车就比上海市的汽车大一些或好一些。另外，类别量表中的数字不具有可加性，因此，不能进行任何运算。例如，在用计算机统计人口时，规定男性为“1”，女性为“0”，并不意味着两个男性相减，可以得到一个女性。所以，类别量表是所有量表中最低级、最粗糙的一种量表。

运用类别量表可以对每个类别中所含事物的个数（或次数）进行累计。例如，可以计算某城市中男性公民和女性公民的人数，某停车场上江苏省汽车和上海市汽车的数目。

对于类别量表所用的统计方法，都属于对次数的统计处理，

如：百分比、 χ^2 检验、列联相关等等。

2. 顺序量表

所谓顺序量表，也称为等级量表，就是根据某一特点，将事物分成等级，并且用数字表示。

例如，在百米赛跑中，根据运动员到达终点的先后，将他们分为第一名、第二名、第三名等等。

再如，学校中对学生的思想品德进行评定时，可以根据学生各方面的表现，用 5、4、3、2、1 这五个分数，表示优、良、中、及格和不及格这五个等级。

在对事物进行等级划分的同时，当然也包括了事物进行类别划分。例如，上述的思想品德的五个等级，也就是五个不同的类别，所以，顺序量表与类别量表都是对事物进行了分类。顺序量表具有优于类别量表的特点，那就是它标志着各类别之间的序列关系。例如，得 4 分者不仅反映了他是和其他得分者属于不同类别，而且还表示比得 5 分者低一等，比得 3 分者高一等，即 $5 > 4 > 3$ ，同样，还有 $3 > 2 > 1$ 的序列关系。

顺序量表比类别量表要高级一些，但它还是有缺点的，即它只能反映事物的类别和顺序关系，而不能进一步告诉我们每一个等级差是否具有相等的距离。例如，百米赛跑中，第一名、第二名之间和第二名、第三名之间虽然都相差一个等级，但他们的成绩之差却不一定相等。也就是说，顺序量表不具有等距性，因此也就没有可加性。由此可见，它也是一种比较低级的量表。

对于顺序量表所能运用的统计方法，仅仅限于中位数、百分位数、等级相关系数、肯德尔和谐系数以及秩次的变差分析等。

3. 等距量表

也称为间距量表，是较类别量表和顺序量表高级的量表。因

为，它除了具有上述两种量表的性质外，还具有等距性，即它的单位是等值的，它对事物分析所作的数量划分是等距离的。例如，摄氏温度计就是一个典型的等距量表，我们可以认为 25°C 和 20°C 之差，与 20°C 和 15°C 之差是相等的。正因为具有等距性，所以能够对等距量表上的数字进行加减运算。

等距量表的缺点是没有绝对零点，即它的起点是人为确定的，而不是事物本身属性所客观存在的。如摄氏温度计上的 0°C ，就是人为规定的。因此，我们对等距量表上的数值不能进行乘除运算。

虽然等距量表有上述缺点，但教育学家和心理学家对它还是特别感兴趣，其原因是：

①教育和心理测量中的许多结果，都可转化为等距量表，而且，教育和心理上所要测量的人的成就和能力等客观上并没有绝对零点，这和等距量表的特性是相符合的。

②等距量表具有这样一个良好特性，即如果我们对等距量表上的分数加上或减去一个常数，甚至乘上或除以一个常数，将不改变原来各分数之间的关系。这样，几个不同单位的分数可以转换到同一个量表上，使我们可以对不同测量方式得到的结果进行比较。

③对于等距量表的数值，大多数统计方法都可以运用。如，计算平均数、变异数、皮尔逊积矩相关系数等统计量，以及应用 t 检验、 F 检验等。

4. 等比量表

也称为比率量表。它是这四种量表中最高水平的测量量表，除了含有类别量表、顺序量表和等距量表的性能外，还有绝对零点。因此，非但可以进行加减运算，而且还可以进行乘除运算。