

跨世纪教育丛书

心理与教育测量

Xinli Yu Jiaoyu Celiang

戴艳 / 编著

天地出版社
Tian di chu ban she



四川省教委项目成果
· 跨世纪教育丛书 ·

心理与教育测量

戴 艳 编著

天 地 出 版 社

《跨世纪教育丛书》顾问及编委名单

(按姓氏笔画为序)

顾 问 杨宗义 张敷荣 黄希庭

主 编 吴定初

编 委 刘海燕 吴定初 曹正善

曾德琪 戴 艳

目 录

祝贺与希望	张敦荣 (1)
前 言	(1)

第一编 测量的基本问题

第一章 心理与教育测量的性质与功能	(1)
第一节 心理与教育测量的性质	(1)
第二节 心理与教育测量的种类和功用	(14)
第三节 正确对待测验, 防止测验滥用	(22)
第二章 心理与教育测量的产生和发展	(26)
第一节 心理测量的产生和发展	(26)
第二节 教育测量的产生和发展	(36)
第三节 心理与教育测量在中国的发展史	(42)
第三章 测量的一般步骤和方法	(50)
第一节 测验的编制过程	(50)
第二节 试题的编制技术	(62)

第二编 测量的基本原理

第四章 测量的误差	(84)
-----------------	--------

第一节	误差的含义	(84)
第二节	误差的来源	(88)
第五章	测量的信度	(97)
第一节	信度的意义	(97)
第二节	估计信度的方法	(102)
第三节	测量的标准误	(120)
第四节	影响信度的因素	(124)
第六章	测量的效度	(127)
第一节	效度的意义	(127)
第二节	内容效度	(131)
第三节	结构效度	(135)
第四节	效标关联效度	(141)
第五节	提高测量的效度的方法	(151)
第六节	效度的应用	(153)
第七章	项目分析	(166)
第一节	项目难度	(166)
第二节	项目区分度	(175)
第三节	常模参照性测验的项目分析	(184)
第四节	目标参照性测验的项目分析	(190)
第五节	项目分析的一个实例	(194)
第八章	测验分数的解释	(197)
第一节	常模参照的分数解释	(198)
第二节	效标参照的分数解释	(220)

第三编 各类心理与教育测验

第九章 成就测验	(227)
第一节 成就测验的性质和应用	(227)
第二节 成套成就测验	(234)
第三节 单科成就测验	(246)
第十章 能力测验	(254)
第一节 智力测验	(254)
第二节 团体智力测验	(268)
第三节 婴幼儿测验	(273)
第四节 多重能力倾向测验	(277)
第五节 创造能力测验	(279)
第六节 特殊才能测验	(294)
第七节 智力的自我测验	(301)
第十一章 人格测验	(323)
第一节 自陈量表	(324)
第二节 投射测验	(360)
附录 心理测验工作者的道德准则	(370)
附表 正态曲线的面积和纵轴	(372)
主要参考文献	(383)

祝贺与希望

张敦荣

再过一年多时间，人类就要跨入自己的第二个千年。在这激动人心的时刻即将来临之际，我们每一位教育工作者既应该回顾已逝的历程，认真总结历史的经验，更应该面向未来教育的发展，明确自己肩负的重任。

为了适应教育改革的需要，推动教育事业的发展并为新世纪培养更多德才兼备的优秀人才，四川师大教育系吴定初教授主持编写了《跨世纪教育丛书》。在《丛书》付梓之前，定初嘱我作一序言。定初是我指导的第一个硕士生，《丛书》的大多数作者也是西南师大教育系的毕业生，我为他们的成绩感到欣慰并且理应支持他们。但是，自己因目力不济未能通看全书，故只好简单写上几句，既祝贺《丛书》的出版，又表达一个老教育工作者的些许心愿。

首先，希望大家更加努力地学习，包括学习古今中外一切有利于我们教育事业发展的思想、理论和学说。因为只有不断吸取全人类在教育研究方面的优秀成果，教育科学才可能长足地发展。其次，希望大家更加重视理论联系实际，尤其是联系我国当前和今后的教育改革实践。因为实践的需要既是理论研究的原动力，又是理论研究的归宿。再次，希望大家更加勤奋地投身于教育科研，出版更多高质量的研究成果。因为教育科学的发展离不开一代又一代中青年学人的不懈奋斗，大家为国家的教育事业多

作贡献就是我的最大心愿。

总之，祝愿大家“百尺竿头，大进几步”。

1998年7月

前 言

心理与教育测量学，是心理学与教育学的一个分支学科，是随着心理学、教育学和统计学理论的进步，在测验运动的实践中发展起来的。

目前，我国已有多种心理与教育测量的专著问世，但这些专著一般都较注重心理与教育测量的最新理论探索，较为详实地阐述编制测验的原理、方法和标准化，涉及较多的心理与教育测量的基础数理统计学知识，即使对具体的心理与教育测验方法的介绍，也较为专门化地论述各种测验的实施方法、计分方法以及临床应用中问题的研究等等。

本书的编写宗旨在于：试图在完整介绍心理与教育测量的最新理论的前提下，深入浅出，注重介绍各种常见的心理与教育测验量表知识。

本书的最大特点是：它不仅能使读者掌握编制和使用心理与教育测验的详尽的专门方法，而且能让读者通过学习本书，了解心理与教育测量及其量表的性质、种类、使用范围及测验结果的含义和用途，以便在使用和接受心理与教育测验时，使心理与教育测验这一科学方法能更好地发挥它的甄别、安置、诊断、咨询和评价等职能，避免心理与教育测验产生消极的影响。

本书分三编十一章，第一编是介绍测量的基本问题，第二编论述测量的基本原理，第三编分述各类测验。其目的是使读者比较系统和全面地掌握心理与教育测量的基本理论和方法。在编写过程中，编者根据多年教授《心理与教育测量》这门课程的体

会，力求理论联系实际，做到简明、准确、系统、完整。本书既可作为学校教学之教材，也适合广大喜好心理与教育测量的读者阅读。

在编写本书的过程中，得到校、系领导的大力支持和帮助；同时参阅了国内外许多测验专家、学者的论著，并引用了其中的一些资料。在此深表谢意。

我要感谢天地出版社和四川锦祝印务所的同志，是他们使我有机会将我的所学、所知奉献给我们的读者。

由于编者学识水平有限，书中疏漏与不当之处在所难免，敬请前辈专家和各方同道不吝赐教，欢迎广大读者批评指正。

戴 艳

1997 年冬于四川师大教育系

第一编 测量的基本问题

第一章 心理与教育测量的性质与功能

世界上无论是有生命还是无生命的物质都是有差异的，正如我们所说“人心不同，尤如其面”、“一棵树上没有两片完全相同的叶子”。如何精确描述或度量物质之间的差异，是千百年来科学工作者不断探索和研究的问题。对一些物理现象如长度、重量、温度、时间等的差异，已形成了一整套的测量工具、方式、标准和单位制，能够精确的度量。而对人的心理现象或教育现象能否像物理现象一样度量呢？心理与教育测量有何特性？有哪些种类？又有什么功能呢？在这一章里我们将解答这些问题。

第一节 心理与教育测量的性质

一、心理与教育测量的可能性

对于物理现象的测量由来已久。物理测量是依据物理学的原则对物理现象的测量，如用秤、尺子和秒表所进行的重量、长度

和时间的度量。物理测量属于客观性测量，基本上不受观察者的主观判断的影响，这已为世人所公认。

人的知识、技能、智力、兴趣、气质、性格、品德等精神特质是否能测量呢？其实，心理与教育测量的思想基础和程序原则与物理测量是一致的。美国心理学家桑代克（E·L·Thorndike）在1918年曾说：“任何现象，只要是存在的，总有一种数量”。美国测量学家麦柯尔（W·A·Mccall）于1922年又提出：“凡是有数量的事物一定可以测量。”也就是说，世界上的事物现象凡是存在，都有程度上的不同，程度上的不同也就是数量上的差异。既然有数量上的差异，就可以进行数量化的测定。学生的学业成就、人的心理品质存在差异已是不可争辩的事实，这种差异就意味着数量的存在，而有数量就可以根据某种工具进行测量。血压计被用来测量血压的差别，视力表则是在发现视力的差别之后发明的；当人的心理现象、学生的学业成就有差别的时候，人们就发明了心理与教育测验和其他测量方法，来度量这些差别。

二、测量的一般概述

（一）测量的定义

心理物理学家史蒂文斯（S·S·Stevens）说：“就其广义来讲，测量是按照法则给事物指派数字。”这就是说，测量是根据一定的法则用数字对事物加以确定。这是迄今大家最能接受的测量定义。

从字面上看，该定义包含了3个要素：事物、数字和法则。

所谓“事物”，代表的是我们所要测量的对象，说得更确切些，就是引起我们兴趣的事物的属性或特征。测量就是确定这些属性或特征的差异。在心理与教育测量中，我们所感兴趣的、所要测量的则是心理能力、人格特点和学业成就等。由于这些特征不可能直接测量，因此，我们所测量的其实是这些特征的外显行

为或者说表现特征。

所谓“数字”，是代表某一事物或事物某一属性的量。数字代表某一事物时，仅具有符号的功能，如邮政编码、学号、运动员的编号等等，它们本身不具有任何意义，只有我们赋予它们意义的时候，它们才具有意义。数字大多数时候具有量的意义，这时它就代表一定数值，这些数值具有数系统的特点，即：①区分性，1就是1，2就是2，没有任何数与1或2这个数相同，被1（或2）指派的事物的每一个体在所测量的方面是相同的。②序列性或等级性，如 $1 < 2 < 3$ 。③等距性，1和2，2和3之间的绝对值总是相等的， $2 - 1 = 1$ ， $3 - 2 = 1$ 。④可加性，某数与他数之和，必得另一个独特的数。由于数的这四个特性，因此，就能通过数的运算所获得的结果来描述事物。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，数字是否具有如上数系统的功能，还必须考虑到事物的性质和指派数字的原则。

所谓“法则”，代表的是测量所依据的规则和方法。如用秤称物体的重量，依据的是杠杆原理；温度计测物体的温度，依据的是热胀冷缩规律；测量人的智力，依据的是有关智力的理论。法则有好有坏，使用好的法则可以得到较为理想准确的测量效果，使用坏的法则，则会得到不准确的测量效果。许多事物易于测量，是因为它们所使用的法则易于制定和便于操作。人类的有些特征如发色、肤色、身高、体重等很容易测量，可人类大部分的特征却是难以测量的，主要是因为设计一个好而清晰的法则非常困难。心理测量是标准化的测量，它的测量法则包括测量的内容、步骤和评分的标准以及测验分数数量化的方法，测验编造得好坏，取决于法则的好坏，其结果决定了心理测量的有效性和可靠性程度。同样，在教育测量上，人们对一些考试方法的意见，实际上就是说这些考试方法作为体现测量学生知识或能力的法则是够理想的，有相当多的缺点。

(二) 测量的要素

任何测量都应具备两个要素，即参照点和单位。

1. 参照点

所谓参照点，就是计算的起点，也称为零点。参照点不同，量数所代表的意义就不同，测量结果就无法进行比较。参照点有两种：绝对零点和相对零点。

绝对零点是以“恰恰没有”作为计算起点，表示什么都测不到的一点，长度、重量都是有绝对零点的。相对零点是人为规定的计算起点，并不表示在这一点什么都没有，海拔的高度，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。理想的参照点是绝对零点，如用相对零点作为参照点，最好尽可能使它接近零点。

心理和教育测量中的参照点一般都是相对零点，因为人的智力、人格、学生的学业成就、思想品德等的绝对零点都很难确定。在相对零点量表上的得分，只能表示差异的大小，不能以倍数来解释得分的多少。在一次智力测验中，甲的智商是 140，乙的智商是 70，我们只能说甲的智商比乙的高 70，而不能说甲的智力是乙智力的两倍，因为没有零智力。

2. 单位

所谓单位，是计量事物的标准量的名称。单位是测量的基本要求，没有单位，数量的多少和大小就无法表示。单位的种类、名称繁多，即使是测量同一事物，也可以用许多单位，如时间的单位有分、秒、时、日、月、年等。理想的单位必须具备以下两个条件：

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。如 1 米、1 000 千克，其定义十分确定，所有人的理解都是一致的，不会引起不同的解释。但在心理与教育测量中解决这个问题就有一定的困难。如以百分度为单

位来测量学生的学业成绩时，这个单位就不如物理单位的意义明确。

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。如甲、乙、丙、丁4个学生，他们的体重分别为40千克、50千克、80千克、90千克，那么我们可以肯定甲和乙的体重与丙和丁的体重差是相等的。而在心理与教育测量中，经常使用的单位都不是绝对等价值的。如甲、乙、丙、丁4个学生，他们在一次测验中的成绩分别为40分、50分、80分、90分，那我们是否能说甲和乙的学业水平之差与丙和丁的学业水平之差相等呢？显然我们不能确定。因为每一个分的价值并非完全相等。

（三）测量的量表

量表是测量的工具，是表示量数的方法，它是具有单位和参照点的可以使事物数量化的连续体。如尺子是度量长短的量表，中国比内测验是度量儿童智力水平的量表。

由于制定量表的单位和参照点不同，量表的种类也不同。根据测量的精确度，史蒂文斯（S·S·Stevens）将测量从低级到高级分成4种量表水平，高级量表除包括低级量表的条件假设和功能外，还有本身的特点。

1. 命名量表

命名量表（nominal scale）也称为称名量表或类别量表。

这是最低水平的一种测量量表，只是根据法则用数字或其他标志来代表事物或把事物归类，这里的数字仅是符号或称呼，没有任何数量化的意义。命名量表又可细分为作为标记的和作为分类的两种，前者仅用数字代表个别事物，如学生的学号、运动员的编号等；后者则是用数字来代表具有某一属性的事物的全体，如用1代表男性，2代表女性等。

在命名量表中，数字只用来作标记和分类，而不能作数量化分析，既不能说 $A > B > C$ ，也不能做加、减、乘、除的运算。

所适用的统计有次数、众数、百分比、偶发事物相关（如四分相关、中相关）以及 χ^2 考验。

2. 顺序量表

顺序量表（ordinal scale）也称等级量表，就是根据某一特点，将事物分成等级，并用数字表示，也就是以事物某种属性程度的大小依次排列的等级顺序。如运动会上各种比赛项目中的名次、学生的考试名次、工资级别、能力等级、对某事物的喜爱程度等，都是顺序量表。

顺序量表比命名量表水平高、也更精确，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。这里的数字包含有大小关系，代表符号“>”。如将能力分为上、中、下3个等级，相应地指派数字3、2、1，在这种情形下，3、2、1构成 $3>2>1$ 的顺序关系。但顺序量表不能说明每一等级之差是否具有相等的距离，如在测验成绩中，第1名和第2名之间、第2名和第3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。

这说明在顺序量表中，既无相等的单位，又无绝对零点，数字仅表示等级，而不表示某种属性的真正量或绝对值，因此不能作加、减、乘、除的运算。它所适用的统计有中位数、百分位数、斯皮尔曼等级相关系数和肯德尔和谐系数以及秩次的变差分析等。

3. 等距量表

等距量表（Interval Scale）也称间距量表，这种量表不仅具有区分性、等级性，而且具有等距性，也就是说一定数量的差异在整个量表的所有部分是相等的。因此，它是比命名量表和顺序量表较高一级的量表。

等距量表上所有的单位都是相等的，所以能对等距量表上的数字进行加减运算。但等距量表没有绝对零点，它的起点是人为

确定的，因而在等距量表上的数值不能进行乘除运算。温度计就是等距量表的典型例子。我们可以认为 20°C 和 15°C 之差，同 15°C 和 10°C 之差是一样的，但是我们不能说 20°C 是 10°C 的两倍。

等距量表上的数值加或减一个常数，或者同一个常数乘或除，不会破坏原来数据之间的关系，这样，一个量表上的数值可以转换为另一个具有不同单位的量表上的数值，而且几个不同单位的测值可以转换到一个通用量表上，使我们可以对不同测量方法得到的结果进行比较。对于等距量表的数值，可采用的统计方法有：计算平均数、标准差、积差相关、阶层相关，应用 T 检验和 F 检验。

4. 比率量表

比率量表 (ratio scale) 是测量的最高水平，它具有相等的单位，又具有绝对零点，此量上所得的数值既可确定一个事物比另一个事物大多少，又可确定大多少倍。因此，量表值可以进行加、减、乘、除四则运算。

比率量表在物理测量中容易见到，长度、重量、时间等都是比率量表。例如，甲的体重 100 千克，乙的体重 50 千克，我们既可以说甲的体重比乙多 50 千克，也可以说甲的体重是乙的两倍。

比率量表除了适用上述几种量表所适用的统计方法外，还可以计算几何平均数和变异系数。

5. 四种测量量表的比较

命名量表、顺序量表、等距量表、比率量表是依照从低到高的次序排列的。量表的次数越高，对于描述事物的数所能允许的算术运算也就越多。高一级的每一种量表，除了包括低一级各种量表的性质之外，还具有其特殊的性质。

人们常常认为运用命名量表、顺序量表进行的测量是定性测