



“十二五”普通高等教育本科国家级规划教材

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

PSYCHOLOGICAL & EDUCATIONAL (第四版)
MEASUREMENT

戴海崎 张 锋 主编



暨南大学出版社
JINAN UNIVERSITY PRESS



“十二五”普通高等教育本科国家级规划教材

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

PSYCHOLOGICAL & EDUCATIONAL
MEASUREMENT

(第四版)

戴海崎 张 锋 主编



暨南大学出版社
JINAN UNIVERSITY PRESS

中国·广州

图书在版编目 (CIP) 数据

心理与教育测量/戴海崎, 张锋主编. —4 版. —广州: 暨南大学出版社, 2018. 5
ISBN 978-7-5668-2370-0

I. ①心… II. ①戴…②张… III. ①心理测量学—高等学校—教材②教育测验—高等学校—教材 IV. ①B841.7②G449

中国版本图书馆 CIP 数据核字 (2018) 第 075786 号

心理与教育测量 (第四版)

XINLI YU JIAOYU CELIANG (DISIBAN)

主 编: 戴海崎 张 锋

出 版 人: 徐义雄

责任编辑: 暨 南 黄圣英 黄 球 黄 斯

责任校对: 刘舜怡

责任印制: 汤慧君 周一丹

出版发行: 暨南大学出版社 (510630)

电 话: 总编室 (8620) 85221601

营销部 (8620) 85225284 85228291 85228292 (邮购)

传 真: (8620) 85221583 (办公室) 85223774 (营销部)

网 址: <http://www.jnupress.com>

排 版: 广州市天河星辰文化发展部照排中心

印 刷: 广州市穗彩印务有限公司

开 本: 787mm × 1092mm 1/16

印 张: 22.75

字 数: 554 千

版 次: 1999 年 2 月第 1 版 2018 年 5 月第 4 版

印 次: 2018 年 5 月第 41 次

印 数: 292001—302000 册

定 价: 69.80 元

(暨大版图书如有印装质量问题, 请与出版社总编室联系调换)

编委会名单

主 任：许尚侠

副主任：莫 雷 刘华山 白先同

编 委：（按姓氏笔画排列）

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

第四版前言

在心理与教育测量应用社会需求量大而情切的召唤下，心理与教育测量学理论和技术的发展也不断与时俱进。在着手本教材第四版修订时，我们面临着很多的选择。选择的新内容如何既体现时代的创新发展，又不失教材循序渐进的逻辑性；既展现学科发展的前沿性，又保持本科教学水平的基础性？在出版社的倾力支持下，本书的原作者们坐在一起，对新内容作了认真的剪裁。

本次修订全书结构和主要内容改进如下：

1. 调换了一部分章节的前后次序，使全书更符合学科内在逻辑。
2. 概论章加强了对《心理测验管理条例》和《心理测验工作者职业道德规范》两文件意义的阐述。
3. 测量信度章增加了对分层 α 信度系数和成套测验合成总分测评信度估计方法的介绍。
4. 测量效度章增加了效度验证的举证模式。
5. 项目分析章增加了第五节“项目功能差异分析”。
6. 能力测验章推出了对知名测验新版本的介绍。
7. 人格测量章增加了第四节“内隐联想测验”。
8. 第十六章改名为“测量应用实务”，加强了测量在各领域实际应用的介绍，并在每节后增加“测量误用与滥用的表现及危害”内容。
9. 第十七章改名为“测量理论与应用的新发展”，并增加了第三节“新一代测验理论与认知诊断简介”。
10. 各章的练习与思考题均增至10题及以上。

希望此次修订后更加有利于学习者的学习，也更加有利于应用者的应用。

修订后，各章节负责人如下：张锋，第一、七、十四章；罗黎辉、张锋，第二章；杨志明，第三、四、五章；龙文祥，第六、八章；陈雪枫，第九、十二、十三章；戴海崎，第十、十一、十七章；龙立荣，第十五、十六章。另有戴步云、倪雨菡等青年学者参与了本书此次的修订工作，特此致谢。

此次修订，我们又参考和摘引了国内外一些新资料，在此再次向本书所有参考资料的作者们表示感谢。

编者

2018年1月5日

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八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、知名测验性能三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：

第一，辟专章介绍了目标参照测验的理论与技术。

第二，介绍了认知心理学的一些测量学新观点。

第三，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。

第四，加强了对我国学者在测量学领域研究活动与成果的介绍。

第五，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，也可作为从事心理咨询、考试评价、人员测评等工作人员的参考书。

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：张锋编写第一、八、十四章，罗黎辉编写第二章，杨志明编写第三、四、五章，龙文祥编写第六、七章，戴海崎编写第九、十一、十七章，陈雪枫编写第十、十二、十三章，龙立荣编写第十五、十六章。

本书初稿由戴海崎、张锋分工审阅并提出修改意见，最后由戴海崎统校定稿。

本书的编写得到了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、安徽师大等高校均对本书的编写出版给予了很大的帮助，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。在编写过程中，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

1997年7月1日

目 录

第四版前言	1
编写说明	1
第一章 心理与教育测量概论	1
第一节 一般测量概述	1
第二节 心理与教育测量的性质	6
第三节 心理与教育测验的种类及其功能	11
第四节 心理与教育测量工作者的素质要求及道德准则	15
第二章 心理与教育测量简史	20
第一节 中国古代的心理与教育测量	20
第二节 现代心理与教育测量在西方国家的产生和发展	24
第三节 现代心理与教育测量在中国的发展	30
第三章 经典测验理论的基本假设	36
第一节 心理特质及其可测性假设	36
第二节 测量误差及其来源	37
第三节 真分数及其有关假设	39
第四章 测量信度	43
第一节 信度概述	43
第二节 信度的估计方法	46
第三节 提高测量信度的方法	56
第五章 测量效度	61
第一节 效度概述	61
第二节 效度的评估	64

第三节 提高测量效度的方法	72
第六章 测验的项目分析	75
第一节 测验项目的难度	75
第二节 测验项目的区分度	80
第三节 猜测问题与猜测率	89
第四节 多重选择题的项目分析	92
第五节 项目功能差异分析	94
第七章 心理与教育测验的编制与实施	99
第一节 编制心理与教育测验的基本程序	99
第二节 测验的实施	106
第八章 常模参照测验	114
第一节 常模参照测验概述	114
第二节 常模的编制	117
第三节 常模参照测验分数的解释与应用	129
第九章 目标参照测验	140
第一节 目标参照测验概述	140
第二节 目标参照测验的信度与效度	143
第三节 目标参照测验的项目分析	146
第四节 常用标准设定方法	151
第十章 测验等值	157
第一节 测验等值概述	157
第二节 测验等值关系计算的基本方法	163
第三节 常用测验等值设计方案	166
第十一章 学绩测验	173
第一节 学绩测验概述	173
第二节 标准化学绩测验	177
第三节 教师自编课堂测验	184
第十二章 能力测验(上)	189
第一节 智力测验的一般问题	189
第二节 个体智力测验	198

第三节 团体智力测验	208
第十三章 能力测验(下)	211
第一节 能力倾向测验	211
第二节 特殊能力测验	220
第三节 创造力测验	224
第十四章 人格测量	228
第一节 人格测量概述	228
第二节 自陈量表	231
第三节 投射测验	240
第四节 内隐联想测验	246
第十五章 其他心理与教育测验	256
第一节 焦虑测验	256
第二节 兴趣测验	260
第三节 态度和品德测量	267
第十六章 测量应用实务	276
第一节 人事测评	276
第二节 心理诊断	291
第三节 教育评价	300
第十七章 测量理论与应用的新发展	307
第一节 项目反应理论简介	307
第二节 概化理论简介	319
第三节 新一代测验理论与认知诊断简介	326
附录一 心理测验管理条例	333
附录二 心理测验工作者职业道德规范	338
附表一 正态分布表	340
附表二 t 分布检验临界值表	348
附表三 积差相关系数显著性检验表	350
参考文献	351

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
- 心理与教育测量的理论基础
- 心理与教育测量的量表与测验
- 心理与教育测量在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学专业和教育专业学生必修的重要专业课，它在心理科学、教育科学的基础研究和应用研究之间起着中介桥梁作用。一方面，它是开展心理学和教育学基础研究的方法论课程；另一方面，它又是心理学和教育学应用研究和解决现实问题的工具性课程。因此，学习心理与教育测量，对于理论研究和实际应用均具有重要的意义。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原理问题，以便为学习以后各章的具体内容提供一个基本框架。

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？

斯蒂文斯（Stevens, 1946）提出了一个简洁的测量定义，认为测量是指“根据法则给客体或事件指派数字”。后来，有研究者（Lord, Novick, 1968; Torgerson, 1958）认为，测量的对象不是客体（如水果或空气）本身，而是客体的特定属性（如水果的重量或空

气中有毒气体的含量)。因此,测量的定义被修改为“根据法则给客体或事件的属性指派数字”,或者通俗地说,测量是指依据一定的法则使用量尺对事物的属性进行定量描述的过程。这一定义包含四个关键词,简单说明如下:

所谓“一定的法则(rule)”,是指任何测量都要建立在某种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,所依据的是物理学上的杠杆原理;用温度计测量温度,所依据的是热胀冷缩原理;用尺子测量物体的长短,所依据的是把尺子零点对准物体的一端,指认出沿直线到物体另一端的距离。有的测量所依据的法则比较稳定和完善,所以测量的结果比较准确、可靠;而有的测量所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便较差。有的测量所依据的法则操作比较直观和简单,一般不需要经过专门训练就能很容易掌握(如称重量或量长短);而有的测量所依据的法则操作程序复杂(如科技领域的专业测量活动),需要经过专门训练才能逐步掌握。

所谓“事物的属性(attribute)”,是指所要测量的客体或事件的特定特征(characteristic)。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属性,对不同的属性要用不同的测量工具依据不同的法则进行测量。有些事物的属性直观明显,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,并容易被多数人认同和接受,测量的结果具有无可争辩性;而有些事物的属性不那么外显,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,也不容易取得多数人的认同和接受,测量的结果也就不容易获得清楚的解释。

所谓“量尺(scale)”,是指测量中所使用的度量工具。例如,重量测量中的杆秤、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尺,不同量尺所使用的单位和参照点也不同。

所谓“定量描述”(quantitative description),是指测量的结果总是对事物属性的量的确定。虽然有时人们把诸如“1”代表男、“0”代表女这样的做法也叫作测量,但这里的数字只是一种分类符号,并不是有意义的数量。所谓“数量”不仅指描述事物特征的符号,而且指一种有序的量。数量具有四个特性:一是区分性,即一个数(如“1”)不同于另一个数(如“2”);二是序列性,即 $1 < 2 < 3 < 4 \cdots$;三是等距性,即 $2 - 1 = 1$, $3 - 2 = 1$,所以, $2 - 1 = 3 - 2$;四是可加性,即一个数加另一个数产生第三个数。数量的这些特点是一切数学运算的基础,同样,也正是这些特点使得对事物特征差异的测量成为可能。有些测量对事物特征定量描述的精确度高些,而另一些测量对事物特征定量描述的精确度低些。测量的精确度既与测量对象的性质有关,也与测量时所用的工具有关。

首先,测量的精确度取决于测量对象本身的性质。我们可以根据测量对象的性质把它分为三种类型。

(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,只要物体的温度不变、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,只是这种差异总是保持在一定范围内,量的改变趋势也呈现出一定的规律。

(3) 模糊型,即事物的量本身是模糊的,难以获得确定的量。例如,对于人的性格特征,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,只是误差的大小不同而已。

其次,测量的精确度取决于测量工具的精密性。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,用皮尺测量比用脚步测量,其结果要精确得多;而用激光测量比用皮尺测量,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,不同的仅是误差的大小而已。测量学的目标之一是设法把误差降到最低限度,但不可能完全消灭误差。

测量技术被广泛用于工农业生产、商业流通、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四种类型。

(1) 物理测量:指对事物物理特征的测量。如长度测量、重量测量、面积测量、速度测量等均属于物理测量。

(2) 生理测量:指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均属于生理测量。

(3) 社会测量:指对社会现象的测量。如在人口普查、经济统计、民意调查中所进行的测量等均属于社会测量。

(4) 心理测量:指对人心理特征的测量。如智力测量、人格测量、职业兴趣测量、态度测量等均属于心理测量。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,教育测量可以纳入心理测量的范畴。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、对学校办学质量的测量、对学校管理水平的测量等。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

在数学上,测量的数量是由零(0)和任何正负数组成的实数系,该实数系通常可用实数轴直观地表示出来(见图1-1)。实数系是可被无限分解为不同单位的数字连续体,其中每个数值都处于该连续体上的一个特定位置。从图1-1可以观察到,一个完善的测量数量必须具备两个基本要素,即测量的参照点和测量的单位。

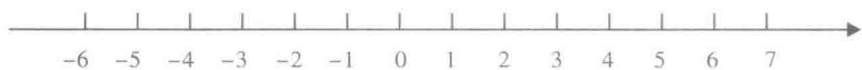


图 1-1 测量中的数量连续体

（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，必须有一个测量的原始起点，也就是测量前测量对象的数量固定原点，这个固定原点就叫作测量的参照点（anchor point）。在测量的数量连续体中，固定原点的数字被定为“0”。显然，要使两个测量数量能够相互比较，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。

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。如重量测量和长度测量就是建立在以绝对零点为参照点的基础上的测量。这个绝对零点的意义就是“无有”，即没有重量或没有长度，以此为测量的起点去确定某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作为测量的起点。如对海拔高度的测量，就是以海平面为测量的起点。人们假定海平面的高度为“零”，然后确定陆地高出海平面的垂直距离。再如对气温的测量，是以水的冰点为测量的起点。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”的度数。

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，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，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。在现实生活中，不同测量所用的单位有很大差异，如长度测量的单位是毫米、厘米、分米、米等，而重量测量的单位是毫克、克、千克、吨等。但抽象到实数轴的数字连续序列中，我们可以观察到测量数量的基本单位是“1”，即数量从原点“0”开始移动，数轴上每个连续整数都与前一个整数相距 1 个单位，数轴上相邻整数间的距离都是相等的。所以，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，所有人的理解都是相同的，不允许做出不同的解释。例如，所有人对重量单位“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 千克与 20 千克之差等于 40 千克与 30 千克之差。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。这一点我们在后文还会谈到。

三、测量的量表

如前所述,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度量值。这种能够使事物的特征数量化的数字连续体就是量表(scale),建立系统的法则,选择有意义的参照点及单位来量化事物属性的活动便称为度量(scaling)。制定量表的参照点和单位不同,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。斯蒂文斯(Stevens, 1946)根据测量中使用的不同参照点和单位,区分出四种不同水平的测量量表。

(一) 称名量表

称名量表(nominal scale)只是用数字代表事物的成分或用数字对事物进行分类,其中的数字只是事物属性的符号,并不具备有意义的固定原点、单位的等距性和数字的顺序性,因而该类数字没有数量的意义。称名量表又可细分为两种形式:一是命名量表,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类型。

在教育或心理学研究中,研究者通常对有关的调查材料用数字编码。例如,将男性编码为“0”,而将女性编码为“1”,但其中的“0”和“1”并不意味着两种性别存在数量的差异。因此,运用称名量表时,对数字不能作常用的数量分析,如不能说6号学生>5号学生>4号学生,当然也不能进行代数运算。适合对称名量表进行统计分析的统计方法有百分比(%)、次数(f)、众数(M_o)和 χ^2 检验。

(二) 顺序量表

顺序量表(ordinal scale)上的数字不仅能够指代类别,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,我们通常取前三名,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、程度来排列数字的量表就叫作顺序量表。

顺序量表中的数字与实数轴中的数值具有相同的顺序性,即不同的数字可表示测量对象的等级、大小和程度的差异,但它既没有相等距离的单位,也没有固定的测量原点。换言之,它既不表示事物特征的真正数量,也不表示绝对的数值,因此不能进行代数运算。适合于对顺序量表进行统计分析的统计方法有中位数(M_d)、百分位数(P_p)、等级相关系数(r_p)和肯德尔和谐系数(W)等。

(三) 等距量表

等距量表(equal interval scale)不仅能够指代事物的类别和等级,而且具有相等距离的测量单位。等距量表的数字是一个真正的数量,这个数量中各个部分的单位是相等的,

因此可以对其进行加减运算。例如，在测定气温时， 10°C 和 15°C 的差别与 15°C 和 20°C 的差别是相等的。

等距量表没有绝对的零点，它的零点是人为假定的相对零点。因此，对等距量表中的两个数量不能进行乘除运算，它们之间不存在倍数关系。例如，我们不能说 20°C 是 10°C 的两倍。适合对等距量表进行统计分析的统计方法有平均数 ($\bar{M}$)、标准差 (SD)、积差相关系数 (r)、等级相关系数 (r_p) 以及 t 检验和 F 检验。

(四) 比率量表

比率量表 (ratio scale) 是最完善的测量量表，因为它除了具有类别、等级和等距的特征外，还具有绝对的零点或固定的原点。许多物理属性的测量都是在比率量表上进行的 (如以厘米测量长度，以千克测量重量，或以天数、月数、年数测量年龄)。在该类测量中，一旦确定绝对零点的位置，那么实数轴上的任何一个非零点的数值均可表示为与其他数值的比值。这意味着，使用比率量表不仅可以知道测量对象之间相差的程度，而且可以知道它们之间的比例。例如，在长度测量中，测得甲的长度为 9 米，乙的长度为 3 米，此时，我们不仅了解到甲比乙长 6 米，也了解到甲的长度是乙的长度的 3 倍。在重量测量中，测得甲的重量为 40 千克，乙的重量为 20 千克，那么，我们既可知道甲比乙重 20 千克，又可知道甲的重量是乙的重量的 2 倍。适合对比率量表进行统计分析的统计方法除了与等距量表相同的之外，还有几何平均数 (M_g)、变异系数 ($C.V$) 等。

第二节 心理与教育测量的性质

一、心理与教育测量的定义

根据一般测量的定义，我们可以将心理与教育测量定义为，根据心理学和教育学法则给人的心理特质和教育成就指派数字，或者依据一定的心理学和教育学理论在测验上对人的心理特质和教育成就进行定量描述的过程。与一般测量的定义相比较，心理与教育测量的定义既具有一般测量的共同属性，又具有其独有的特征。

首先，心理与教育测量依据的法则在相当程度上是一种理论，很难达到如同物理测量依据的法则那样普遍被研究者共同接受的科学水平。心理与教育测量学家凭借这些理论来编制测量的工具并完成测量工作。例如，测量学家在编制智力测验时对智力本身的性质存有非常不同的看法，对智力的结构成分也有非常不同的理解。因此，从一种智力测验上得到的测量分数与从另一种智力测验上得到的测量分数可能具有不同的意义。由于所依据的法则不够成熟，即使使用同一种测验测量，所得结果也不像物理测量那样准确和可靠。这种情况在人格测量领域表现得更加明显。

其次，心理与教育测量的对象是人的心理特质和教育成就。教育成就的含义比较明显，是指一个人通过接受教育而获得的知识水平。但心理特质的含义则比较含混，不大能

够给予清楚的界定。心理学上通常将“特质”(trait)理解为相对稳定的、对个人的行为具有持久调节作用的心理特征,如智力、兴趣、态度、人格等均可视为特质。但这些特质本身就是很抽象的概念,在测量工作中,将其具体化为可操作的测量对象本身就是一项非常复杂的工作。心理特质显然具有内隐性,我们不可能像测量重量或长度那样直接测量人的心理特质的量,而只能通过测量个人在特定情境中的外显行为来推断他的心理特质。这就决定了心理与教育测量只能是一种间接测量。

再次,心理与教育测量的量尺是由有关领域的专家经过长期的编制、试用、修订、完善而逐渐形成的标准化测验(standardized test),它的编制是一项高度专门化的系统工作,要达到科学所要求的水平绝非易事。

最后,心理与教育测量的目标虽然是对人的心理特质和教育成就进行定量分析,但这种定量分析的精确度远不及物理测量的精确度。这首先是由人的心理特质的高度复杂性所决定的,同时也与目前的测验缺乏令人十分满意的信度和效度有关。

二、心理与教育测量的理论基础

在阐述心理与教育测量的定义的过程中,我们在说明心理与教育测量同一般测量的共同特征的同时,着重强调了心理与教育测量的复杂性和难度。也许正因为如此,人们对心理与教育测量是否必要和是否可能持有怀疑态度。归纳起来,怀疑心理与教育测量的必要性和可能性的理由不外乎两个方面:一是人的心理现象和知识水平是一种主观存在,它的复杂性、流动性和内隐性的特征使人们不可能对其进行直接测量;二是目前的心理与教育测量的技术手段远未达到如物理测量那样准确和可靠的水平。那么,心理与教育测量究竟是否必要和可能呢?对此,我国古代学者孟子早在两千多年前就给予了明确的回答(参见第二章)。但直到20世纪初期,随着心理与教育测量运动的发展,这个问题才真正摆在了测量学家的面前。1918年,桑代克(E. L. Thorndike)曾提出,“凡客观存在的事物都有其数量”。1939年,麦柯尔(W. A. McCall)进一步指出,“凡有其数量的事物都可以测量”。这两个命题被公认为是心理与教育测量的理论基础。

从哲学的观点看,任何事物都是质和量的统一,事物的质的差异是分类的前提,而事物的量的差异则是测量的前提。这里的“事物”不仅指外在的客观现象,而且指人的内在的心理特质。根据这一前提,我们可以认为:第一,人的心理特质和知识水平如同其他一切物理现象一样是有差异的,这种差异不仅包含质的方面,而且包含量的方面。因为有差异,所以有必要测定差异的数量,描述差异的程度。第二,心理特质和知识水平虽然不是物理实体,不能直接测量,但是必须表现于人的外部行为之中,并调节着人的外部行为。因此,通过观测人的外部行为的差异,就有可能测量出人的心理特质和知识水平的差异。第三,心理与教育测量的准确性、可靠性和精确度如同其他一切测量技术一样是相对的,也同其他一切测量技术一样必然随着科学技术的进步和发展而逐步提高。尽管目前心理与教育测量的科学性还达不到人们所期望的高度,但测量不准不等于不能测量,目前测量不准也不意味着将来永远测量不准。一百多年来,心理与教育测量学家正是抱着这种信念进行了大量的研究,取得了明显的成效,已经初步形成一套比较科学的测量原理与技术。可

以设想,随着研究工作的拓展和深化,心理与教育测量的科学水平将会进一步提高。

三、心理与教育测量的量表

在本章第一节里,我们已经谈到,测量中所使用的单位和参照点不同,就会有不同水平的测量量表。那么,心理与教育测量的量表属于哪一水平呢?从本质上讲,心理与教育测量的量表属于顺序量表。其原因如下:

(1) 从所使用的参照点来说,教育测量和心理测量领域的参照点均为相对零点,而非绝对零点。例如,在学期末的学科考试中,通常的做法是把学生的成绩确定在 0 ~ 100 分之间。显然,这个“0”是命题人人为假定的测量起点。因为即使某生在该试卷上得了 0 分,也不能说该生在本学期内没有学到任何知识,或者说该生的知识水平为“零”。在智力测量中,假定某一儿童不会做任何一道题目,那么,他的成绩也为“0”,但这个“0”也并不表示他的智力水平为“零”。这就决定了心理与教育测量的量表不可能达到比率量表的水平。

(2) 从所使用的单位来说,心理与教育测量的单位远没有其他测量的单位成熟和完善。一是心理与教育测量所使用的单位意义不太明确。例如,在各种形式的考试中,虽然使用的单位都是“分”,但实际上,数学考试中的“分”和语文考试中的“分”意义是不相同的。学生在不同学科上的考试成绩所反映的不是同一个领域的知识水平。二是在心理与教育测量中的单位常常不等值。例如,同一次数学考试,学生做对一道较简单的题目,得到 1 分,同样做对一道较复杂的题目,也得到 1 分。从表面上看,前者的 1 分和后者的 1 分是等值的,但实际上,它们所反映的学生的知识水平是不相等的。由于单位的意义不同,单位的价值不相等,因此各科的考试成绩不能直接相加而求出总分,也不能根据总分求各科平均分。这就决定了心理与教育测量量表上的度量值不是等距量表上的度量值。然而,无论是考试还是能力测验,不同个体获得的分数具有顺序性,可以比较其大小。例如,虽然同一数学试卷上获得的成绩 80、85 和 90 分之间没有绝对零点,也不能认为彼此间的差异量相等,但我们可以对三个分数做出顺序排列: $90 > 85 > 80$ 。所以,它是基于顺序量表获得的度量值。

由于顺序量表的参照点没有绝对零点,而且它的单位不等值,大量的统计方法不能直接应用到顺序量表的分数上,因此在理论研究和实际应用工作中受到极大的限制。为了克服这些缺陷,心理与教育测量学家希望将从顺序量表上得到的分数转化到等距量表上去解释。也就是说,可以采用统计方法把顺序量表的分数转换到具有相等单位的另一个量表即等距量表上(见第八章和第十章)。目前,对大多数心理与教育测量的分数解释工作是在等距量表上实现的。

四、心理与教育测量中的测验

如前所述,心理与教育测量工作是在测验的基础上完成的,而测验是由有关领域的专家经过长期的编制、试用、修订、完善而逐渐形成的标准化测量工具。对于什么是测验的