

黎光明
编 著

心理测量

清华大学出版社



心 理 测 量

黎光明 编著

贵州师范学院内部使用

清华大学出版社
北 京

内 容 简 介

心理测量是我国各大学心理学专业学生必修的重要专业课,它在心理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的基础,是从实际应用研究的工具课。

本书由经典测验理论、具体心理测验和现代心理测验理论三大部分构成,既注重基础理论、基本原理的准确性阐述,又强调测验编制与使用的基本技能训练。在内容上,本书力求反映国内外心理测量的最新发展,将理论与实践结合在一起,突出心理测量的科学性与实用性。

本书可供高等院校心理学专业学生使用,也可供相关专业本科生、研究生和社会各界从事心理测量工作的人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

心理测量/黎光明编著. —北京:清华大学出版社,2019

ISBN 978-7-302-46340-5

I. ①心… II. ①黎… III. ①心理测验 IV. ①B841.7

中国版本图书馆 CIP 数据核字(2017)第 021625 号

责任编辑:姜志敏

封面设计:杨学勇

责任校对:王荣静

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:15.5 字 数:376 千字

版 次:2019 年 8 月第 1 版 印 次:2019 年 8 月第 1 次印刷

定 价:45.00 元

产品编号:065623-01

前 言

心理测量是我国高等院校心理学、教育学等专业重要的必修课程。应清华大学出版社的邀请,我们编写出版了《心理测量》这本书,将其作为高等院校心理学、教育学等专业的教材。

在本书撰写过程中,一方面借鉴了国内外已有的成果;另一方面结合了作者多年从事心理测量教学与研究的经验。为了写好本书,我们力图有一些特色和新意,更加突出实践性,使其更能满足在社会科学领域的研究与应用中对测量知识的需求。

本书在内容体系的设计上不仅强调学生应该具备基本的测量思想、掌握基本的测量方法,还强调应该培养学生对心理测量方法的分析与解决问题的能力。本书在编写过程中,根据社会科学类专业的特点,努力贯彻“学以致用”的原则,尽可能做到结构合理、概念明确、条理分明、深入浅出。在内容上,本书包括三大块:

第一大块:经典测验理论。经典测验理论体系构建清晰、易理解,始终围绕“心理测验”这个中心讲解,再演绎这个中心概念的三大要素,即行为样本、客观化和标准化。行为样本包含在常模这一章(第二章);客观化包含在“四度”中,即信度(第三章)、效度(第四章)、难度(第五章)和区分度(第五章);标准化包含在“四化”中(第六章:测验编制标准化、测验实施标准化、测验评分标准化和测验解释标准化)。

第二大块:具体心理测验。本书仅选择常见的三大心理测验,即能(智)力测验(第七章)、人格测验(第八章)和心理评定量表(第九章),作为具体心理测验的代表,而且仅选择一些比较有代表性的智力测验、人格测验和心理评定量表作为示例。

第三大块:现代测验理论。现代测验理论包括项目反应理论(第十章)和概化理论(第十一章),这也是本科生或研究生等必须掌握的一部分内容。本书较为详尽地阐述了这部分内容,尽量做到与国际前沿接轨。

本书从立项到完成,尽管整个写作过程由作者一人执笔,但却是集体力量 and 智慧的结晶。首先,要感谢国内外心理测量同行专家及学者,书中引用的大量文献资料是他们多年潜心研究的成果;其次,要感谢我的研究生们,如王幸君、蒋欢、梁正妍、谢晋艳、侯桂云、张晓婷、甄锋泉等,在编写过程中她(他)们为本书提供了大量宝贵的资料,这些资料让本书增色不少;最后,还要感谢清华大学出版社为本书所付出的辛勤劳动。

尽管倾尽全力,有时为了某种理论、某种方法或某个公式的科学性和准确性不厌其烦地查阅各种文献;尽管有许多好的想法和构思,在内容编排和组织上精益求精,但由于作者水平有限,因此书中难免会有疏漏与不足。为此,恳请国内外的专家同行及广大读者朋友批评指正。

作 者

2019年2月于华南师范大学

目 录

第一章 绪 论	1
第一节 心理测量概述	1
第二节 心理测验的分类与功能	9
第三节 心理测验的历史沿革与发展	14
第四节 如何正确对待和使用心理测验	20
第二章 常 模	25
第一节 常模团体	25
第二节 常模类型	31
第三节 常模应用	39
第三章 信 度	46
第一节 信度定义	46
第二节 信度估计	49
第三节 信度影响因素	59
第四章 效 度	63
第一节 效度定义	63
第二节 效度估计	66
第三节 效度影响因素	77
第五章 难度与区分度	82
第一节 项目难度	82
第二节 项目区分度	86
第六章 测验标准化	98
第一节 测验编制标准化	98
第二节 测验实施标准化	103
第三节 测验评分标准化	107
第四节 测验解释标准化	111
第七章 智力测验	116
第一节 智力测验概述	116
第二节 个别智力测验	126
第三节 团体智力测验	134
第四节 智力测验的理论问题	136
第八章 人格测验	142
第一节 人格测验概述	142
第二节 自陈量表	144

第三节 投射测验	154
第四节 人格测验的理论问题	160
第九章 心理评定量表	164
第一节 心理卫生综合评定量表	164
第二节 情绪及相关问题评定量表	169
第三节 应激及相关问题评定量表	174
第十章 项目反应理论	185
第一节 项目反应理论概述	185
第二节 项目反应理论分析过程	189
第三节 项目反应理论应用	199
第十一章 概化理论	205
第一节 概化理论概述	205
第二节 概化理论分析过程	212
第三节 概化理论应用	219
附 录	223
附 表	228
参考答案	232
参考文献	239

第一章 绪 论

心理测量是我国高等院校心理学、教育学等专业重要的必修课程。该课程是从事基础理论研究的方法课,也是从事应用研究的工具课。心理测量广泛应用于教育、经济、管理、卫生医疗、司法、军队等领域,并产生了重要的作用。本章主要介绍心理测量概述、心理测验的分类与功能、心理测验的历史沿革与发展,以及如何正确对待和使用心理测验。

第一节 心理测量概述

一、什么是测量?

测量,简单地说,就是要把“量”给测出来。所谓“量”其实就是数字。如此看来,测量其实就是要将数字测出来。何谓数字?就是人们通常理解的阿拉伯数字或它们的排列以及组合等构成的符号。例如,人的身高多高,桌子多长等,这些测量最后都需要用数字来表达。

史蒂文斯(S. S. Stevens)(图 1-1)于 1951 年给测量下了定义:
“从广义而言,测量就是根据一定的法则用数字对事物加以确定。”

史蒂文斯的关于测量的这一定义概括了物理测量、社会测量、生理测量、心理测量等的共性。从史蒂文斯测量的定义中,不难发现,测量包含三个要素:测量对象(事物)、测量法则(法则)和测量结果(数字)。这三个要素的关系可形如: $y=f(x)$ 。其中, x 表示测量对象, f 表示测量法则, y 表示测量结果,即:



图 1-1 史蒂文斯

- ① 事物(x)——测量对象
- ② 法则(f)——测量法则
- ③ 数字(y)——测量结果

由此看来,测量需要包含以下三个要素。

第一个要素:测量对象。测量要得到测量出来的数字,应该先要有测量的对象,即要告知是测什么。“巧妇难为无米之炊”,即便测量工具如何好,测量者如何负责,没有测量对象也是无稽之谈。

第二个要素:测量法则。测量人的身高与测量人的体重,所使用的法则是不同的,而法则往往体现在工具中。测量人的身高用米尺,测量人的体重用磅秤,这是工具本身蕴含的法则有所不同。人们知道身高和体重具有不同的测量法则,不能相互混用。用磅秤来测量人的身高,或用米尺来测量人的体重,都是不妥当的。

第三个要素:测量结果。测量结果就是通常所指的“量”,即数字,这是人们真正想要的。这样看来,测量(liáng)其实就是测量(liàng),就是对测量对象通过一定的法则把量(liàng)给测出来,人们最终要的就是这个量。例如,桌子长 2 尺,温度 4 摄氏度等。

当然,因测量对象的不同,测量的种类也不同。如果测量对象是物理现象,那么测量就是物理测量,如桌子的长度。如果测量对象是社会现象,那么测量就是社会测量,如民意调查。如果测量对象是生理现象,那么测量就是生理测量,如血压多高。如果测量对象是心理现象,那么测量就是心理测量,如心理健康程度。

依据 $y=f(x)$, 我们知道, x 和 f 的目的是为了得到 y , 这表明 y 是测量结果, 即量或数字, 是非常重要的。但是, 得到了 y 是远远不够的(至多只能说完成了 $1/2$), 这是因为只有 y 也不一定知道 y 的含义和价值。这就如同一个学生通过测试考了 90 分, 这符合 $y=f(x)$, 其中 x 是学生能力, f 是测试题目, y 是考分 90 分。若不知道参与测试的学生的总体成绩分布, 则即便知道这个学生考了 90 分, 是高还是低呢? 人们也仍然无从知道。因此, 只有 $y=f(x)$, 是远远不够的, 还必须知道 y 的具体含义和价值。这就是说, 测量把量给测出来是主要目的, 但量要表达的含义和价值也是非常重要的。把量给测出来, 这就如同把孩子生出来固然是相当重要的, 这一步相当于生成量, 即 $y=f(x)$ 。但是, 把生出的孩子培养成人也是非常重要的, 只有通过这一步才知道孩子出生后的含义和价值, 这相当于 y 的含义和价值(不是生出 y)。一句话, 只生出孩子是不够的, 至多只完成了 $1/2$, 还必须继续把他(y)培养成人, 才能完成剩下的 $1/2$ 。

那么, 如何来表达 y 的含义和价值呢? 这需要将 y 表达在一个系统的量纲(scale)中, 从总体俯视 y 在整体中的地位, 才能有效获知 y 的含义和价值。在此处, scale 意指“量纲”或“量尺”。

依据上述例子, 如果一个学生测试考了 90 分, 而大多数同学只考了 60 分左右, 那么说明他比较厉害, 分数相对高些, 可形容为“鹤立鸡群”! 如果大多数同学都考了 95 分以上, 那么说明他就比较差了! 同样的一个分数, 在不同群体比较中, 含义和价值是不同的。因此, 我们需要将分数表达在整体系统的量纲中。

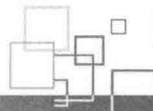
试想, 如果有人在你面前放置 1 元钱, 且无法知道 1 元钱的货币系统的量纲, 即不预先告知是哪一种货币, 那么你也无法知道 1 元钱的值或含金量。当然, 若在你面前放置 1 英镑、1 美元、1 元人民币、1 卢布、1 泰铢等, 则你会选择哪个呢? 答案很显然是 1 英镑。这表明, 只知道数字(量)是不够的, 还必须告知是哪一种货币系统的量纲, 因为不同的货币系统的量纲, 相同的数字(如 1 元), 其价值或含金量可能是不同的。同理, 心理的量也存在相同之处, 即仅测量出量是不够的, 还必须指定量是在哪一个系统的量纲上。

那么, 如何来定义测量的量系统的量纲呢? 为方便表述, 这里, 我们把“测量的量系统的量纲”缩称为“测量量纲”。依据一定标准, 1946 年史蒂文斯将测量量纲从低到高分成 4 种水平, 如表 1-1 所示。

表 1-1 4 种测量量纲在要素上的属性

要素 量纲	是否有大小之分 (大小)	是否有相等单位 (单位)	是否有绝对零点 (零点)	是否能进行代数运算 (运算)
称名量纲	×	×	×	×
顺序量纲	√	×	×	×
等距量纲	√	√	×	√/×
等比量纲	√	√	√	√

注: “×”表示不具备某一类要素属性; “√”表示具备某一类要素属性。



在表 1-1 中,测量量纲的要素包括 4 种,即大小、单位、零点和运算。“大小”表示某一类量纲是否有大小之分、强弱之分、等级之分。“单位”表示某一类量纲是否有相等单位,即不同量级之间是否能等量齐观。“零点”表示某一类量纲是否有绝对零点,而不是人为定义的零点,这里的零点又称为参照点、起点、原点、出发点。“运算”表示某一类量纲是否能够进行代数上的数学运算,如是否能够进行加、减、乘、除等运算。

根据测量量纲的要素,表 1-1 明确规定 y 的测量量纲可分为 4 种,即称名量纲(nominal scale)、顺序量纲(ordinal scale)、等距量纲(interval scale)和等比量纲(ratio scale)。 y 有 4 种测量量纲,这就表明即使有相同的 y 值,在不同系统的量纲上其含义也是不同的。例如,对于数字(量)1 和 2,这两个人们最常见的数字(量),在不同系统的量纲上其含义也是不同的。为了说明这个问题,现举例如下。

如果 1 和 2 在称名量纲上,表示数字命名,1 代表男,2 代表女。我们不能说 2 大于 1,因为性别没有大小之分,仅仅是用 1 和 2 来表示它们的分类。这一类数字既没有相等单位,也没有绝对零点,更不能作代数运算。例如,用 0 代表藏族,1 代表傣族,2 代表汉族,我们不能说 $2-1=1-0$,因为没有相等单位。也不能说这里的 0 就是没有,这仅仅代表藏族,所以这里的 0 不是通常人们所说的绝对零点或起点。也不能作运算,如用 1 代表藏族、2 代表傣族、3 代表汉族,不能说 $1+2=3$,即藏族+傣族=汉族,这里的数字仅起分类或称名的作用。电话号码、手机号码、学生证号码、车牌号码等都是称名量纲。称名量纲在大小、单位、零点及运算四个量纲要素上都是“ $\times$ ”。称名的意思是指这类数据(或量)仅仅是为了分类或命名。

如果 1 和 2 在顺序量纲上,表示考试成绩排名,第 1 名和第 2 名。我们可以说第 1 名比第 2 名好,有顺序(大小)之分。但这一类量纲没有相等单位,因为我们不能说第 2 名-第 1 名=第 3 名-第 2 名,第 2 名与第 1 名的差距不一定等于第 3 名与第 2 名的差距。如果第 1 名 80 分,第 2 名 70 分,第 3 名 50 分,那么就显而易见了, $3-2 \neq 2-1$ 。顺序量纲也没有绝对零点,因为第 0 名并不是第 1 名的起点,更何况通常没有第 0 名的说法。顺序量纲也不能作代数运算,这是因为我们不能说第 1 名+第 2 名=第 3 名。顺序量纲在大小量纲要素上是“ $\sqrt{}$ ”,但在单位、零点及运算三个量纲要素上都是“ $\times$ ”。顺序的意思是指这类数据(或量)能够区分出名次(顺序)。

如果 1 和 2 在等距量纲上,表示温度 1 摄氏度和 2 摄氏度。我们可以说温度是有大小之分的,2 摄氏度要比 1 摄氏度热一些,3 摄氏度与 2 摄氏度之间的差值等于 2 摄氏度与 1 摄氏度之间的差值($3-2=2-1$)。温度的单位是常温常压下人们规定水结成冰定义为 0 摄氏度,水变成水蒸气定义为 100 摄氏度,从 0 摄氏度到 100 摄氏度切 99 下,细分成 100 段,取其中的 1 段即为温度的单位,每 1 段都是相等的。但是,温度的零点(或称参照点、起点、原点、出发点)是相对的,因为温度的零点是人们将常温常压下水结成冰时的温度定义为 0 摄氏度,当然人们也可以不这么规定零点的,所以这是相对的。就相当于测量珠穆朗玛峰的高度,为 8844.43 米,是人们通常以海平面为起点的。当然了,从珠穆朗玛峰山脚到山顶是没有那么高的,所以起点也可以选在山脚或测量者现在自己的脚下。很显然,类似于温度或山的高度,零点是相对的,不是绝对的。正因为温度有相等单位所以它能作加减运算,正因为温度没有绝对零点所以它不能作乘除运算,于是我们在表 1-1 中表达成“ $\sqrt{}/\times$ ”。等距量纲在大小、单位这两个要素上都是“ $\sqrt{}$ ”,但在零点要素上是“ $\times$ ”,在运算要素上是“ $\sqrt{}/\times$ ”。等距的意思是指这类数据(或量)单位是等距的。

如果1和2在等比量纲上,表示人的身高1米和2米。我们可以说2米的身高要比1米高,有大小之分。对于人的身高,我们可以说 $2-1=3-2$,有相等单位。等比量纲也有绝对零点,0表示没有高度。类似的还有重量,如1斤、2斤,如果是没有重量就是0。事实上,试想如果你去买肉,盘子上没有放肉,那么数字当然应该显示为0了。正是因为身高、重量等这类量有相等单位和绝对零点,所以它既能作加减运算,又能作乘除运算。我们既可以说1斤肉加3斤肉等于4斤肉,也可以说4斤肉比3斤肉多1斤肉,还可以说4斤肉是2斤肉的2倍等。等比量纲在大小、单位、零点和运算四种量纲要素上都是“√”。等比的意思是指这类数据(或量)有绝对零点,是可以相除或等比的。

在表1-1中,“×”表示某种测量量纲不具备某一类要素属性,“√”表示某种测量量纲具备某一类要素属性。四种测量量纲在要素上的属性存在不同。根据表1-1,由上分析可知,从出现“√”的频数看,称名量纲为0,顺序量纲为1,等距量纲为3,等比量纲为4,这表明量纲水平依次升高,量纲水平最高的是等比量纲,其次是等距量纲,再次是顺序量纲,最后是称名量纲。

综上所述,理解测量的概念需要注意以下几点。

第一,测量的量是根据 $y=f(x)$ 得到的。测量需要具备三个要素,即测量对象 x ,测量法则 f 和测量结果 y , y 相对更为重要,因为 x 和 f 都是为了“生出” y 。

第二,测量的量需要指定测量量纲。“生出”或测出 y 是不够的,还必须赋予 y 含义和价值,对于相同的 y ,不同的测量量纲其含义和价值是不一样的。测量量纲共有4种,包括称名量纲、顺序量纲、等距量纲和等比量纲。

上述第一点是把 y “生(或测)出来”,第二点是把 y “培养成人”,这两点功劳各半。只有基于以上两点,测量的量才是有价值的,仅测出量来也至多只完成了1/2,仍需知道测出来的量的含义和价值如何,这就需要通过判断它属于哪种测量量纲来表达。

二、什么是心理测量?

通过以上分析可知,测量因测量对象的不同可分为物理测量、生理测量、社会测量和心理测量等。心理测量(Psychological Measurement)仅仅是测量的一种。心理测量隶属于测量,其内涵是相同的,但外延变小了。与测量一样,心理测量仍然必须具备三个要素,即测量对象 x 、测量法则 f 和测量结果 y ,但它的三个要素的外延都要变小,“狭义化”或“特指化”。三个要素的关系也可形如为 $y=f(x)$:

- ① 事物(x)——测量对象——心理现象
- ② 法则(f)——测量法则——心理法则
- ③ 数字(y)——测量结果——心理的量

从心理测量与测量的形式看,两者其实没有多大差别,仅仅是心理测量对象的范围变小了,狭义化了,特指“心理”,所以它叫心理测量,加了“心理”这个特定称谓。

心理能测量吗?人们听起来有点儿觉得不可思议,因为心理测量不像物理测量那么直观。物理测量对象大多看得见摸得着,而心理测量对象既看不见,也摸不着。

然而,我国思想家孟子(图1-2)说过:“权,然后知轻重;



图1-2 孟子

度,然后知长短。物皆然,心为甚。”

孟子这句话的意思是说:“称一称我们就知道物体的重量,量一量我们就知道物体的长短,物理测量是这样,那么心理测量就更加是这样了。”人的心理之所以可以测量是源于人的心理特性是存在差异的。例如,孔子认为,人可分为上人、中人和下人。有了心理特性差异就为心理测量打下了量分的基础,有了心理特性差异人们就可以用数字来表达它们之间的相对或绝对差异。1918年,桑代克(E. L. Thorndike)提出:凡客观存在的事物都有其数量。1939年,麦柯尔(W. A. McCall)进一步提出:凡有其数量的事物都可以测量。上述两句话整合起来,意思就是:凡客观存在的事物都可以测量。心理现象是客观存在的事物。所以,心理是可以测量的。

对于心理测量,仍然可以表达成: $y=f(x)$ 。

对于 x (心理现象),因为存在不同种类的心理现象,所以测量工具也存在不同种类。若测量智力、能力倾向、学业成就等,则称为能力测验。若测量兴趣、态度、性格等,则称为人格测验。若测量心理状态(焦虑、压力、恐惧、成瘾等),则称为心理评定量表。

对于 f (心理法则),因为心理现象存在不同,所以法则也是千差万别的;更因为心理法则基于同一种心理现象存在多种理论,所以导致心理法则具有很大的“派别性”。例如,智力理论包括二因素论、群因素论、多元论、PASS模型等;注意理论包括过滤器理论、资源限制理论等;人格理论包括三个人格特质理论(EPQ)、五个人格特质理论(NEO-PI)、十六个人格特质理论(16PF)等。心理法则呈现出多样性,因此人们在使用心理学工具时一定要知道它是出自于哪一个理论,不同的心理学理论对于同一种心理现象的解释是不同的,即所谓“理论在先,工具在后”。

对于 y (心理的量),是通过 $f(x)$ 转换过来的。如上所述,存在4种不同的测量量纲。但是,一般说来,心理的量相对不够完善,既无相等单位,也无绝对零点。心理的量无相等单位,例如,有两位同学甲和乙,甲第一次考了40分,第二次考了45分,提高了5分;乙第一次考了90分,第二次考了95分,也提高了5分。哪位同学的5分更容易提高呀?答案是甲,因为一般地从40分提高到45分的难度远远小于从90分提高到95分的难度,这表明 $5分 \neq 5分$ 。同理,如果都只提高了1分,那么 $1分 \neq 1分$,这说明心理的量1个单位不等于1个单位。另外,心理的量也很难找到一个统一的起点(绝对零点),一个学生考了0分,并不代表他一塌糊涂,也并不意味着他一点儿智商都没有,考了0分也只能说明对于这次考试相对于某些方面他是不足的,并不代表他全然不懂,甚至于完全没有智商了。我们只能说,该学生相对于其他同龄同学他的成绩在某些方面是不足的。然而,心理的量却有大小之分,所以学生的成绩仍可排名。

综上所述,心理的量既无相等单位,也无绝对零点,但有大小之分。根据表1-1,心理的量只能在顺序量纲上表达了。但是,人们往往会通过某些统计技术将心理的量从顺序量纲调整为等距量纲或等比量纲。

三、什么是心理测验?

安娜斯塔西(A. Anastasi)于1963年给心理测验下了一个定义:心理测验是对行为样本客观而标准化的测量。在这个定义中,包括了三个要素:行为样本、客观化和标准化。这表明,心理测验是心理测量的一种,所测的是一般心理素质与状态,工作追求客观化和标准化。

测量、心理测量、心理测验三者的关系如图 1-3 所示。

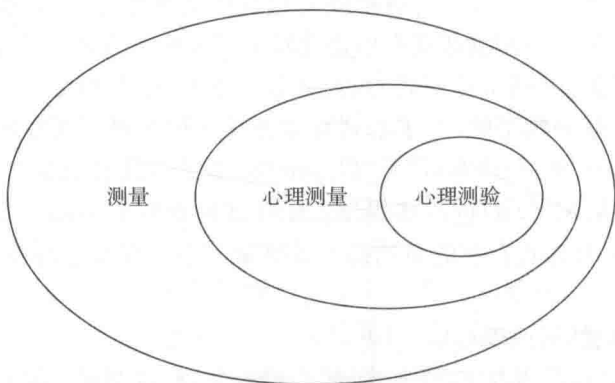


图 1-3 测量、心理测量和心理测验的关系

从图 1-3 可以看出，测量的概念最大，其次是心理测量，最后是心理测验。这表明，心理测量只是测量的一种，心理测验又只是心理测量的一种。

为了说明心理测验和心理测量之间的关系，举一例子来说明。有一位同学在网络上做了一个小小的测试，是关于心理年龄方面的，结果测出其心理年龄 80 多岁，非常郁闷！跑去问测量学老师该怎么办？测量学老师建议她按照某种规则过一周再测一次。这位学生按照老师的吩咐，真的过一周去网上再次测试了，心理年龄的结果变为 60 多岁，她很高兴地告诉了测量学老师。测量学老师建议她过一周再测一次，心理年龄的结果变为 40 多岁，她更高兴了。测量学老师依然建议她又再测一次，结果她的心理年龄回归正常——20 多岁了！最后，测量学老师还是建议她再测，结果，她不敢再测了！因为她害怕结果。

对于这个例子，讨论：

(1) 这个测试是心理测量吗？

(2) 这个测试是心理测验吗？

答案：(1) 这个测试是心理测量。这个测试符合 $y=f(x)$ ，其中 x 是心理年龄，属于心理现象； f 是这个测试制定者规定的测试法则； y 是 80 岁、60 岁、40 岁、20 岁等心理年龄（量）。符合心理测量定义及三个要素的要求，因此这个测试是心理测量。

(2) 这个测试不是心理测验。心理测验必须具备三个要素，即行为样本、客观化和标准化。行为样本来自实践需要，并根据科学理论取样获得，又叫常模样本、代表性样本、标准化样本。这个测试制定者没有制定有效常模，谈不上具有常模样本，其标准设定仅仅是根据自己的标准来给定测试者分数，依自我规则规定测试者做哪些题得多少分。客观化指标包括难度、区分度、信度和效度，这个测试每次测试分数摆动太大，没有什么稳定性（信度）可言，更谈不上什么有效性（效度）。另外，这个测试没有经过严格意义上的标准化，即没有经过严格意义上的测验编制标准化、测验实施标准化、测验评分标准化和测验解释标准化。因此，不难看出，这个所谓的心理年龄小测试不具备心理测验三个要素，不是心理测验。

根据上面这个例子，我们可以发现，是心理测量不一定是心理测验，但是心理测验就一定是心理测量。这表明，心理测验不是一般的心理测量，其追求客观化和标准化。心理测量要成为心理测验，必须具备以下三个要素。

第一个要素：行为样本。心理测量是间接测量，具有“间接性”，针对的是人的行为，因

为人的心理看不见也摸不着,不能直接加以测量,只能通过外显行为导引出来,即心理学常见公式:

$$S-O-R$$

(1-1)

在公式(1-1)中,S表示刺激;O表示中介变量;R表示行为反应。

正是因为我们测量的人的心理只能通过测量行为来反映,所以这种策略是迂回策略,是间接的。当然,行为的产生需要通过一定刺激触发(导引出来),这便是S。可供导引行为产生刺激的触发方式比较多,如提问题、设情景、布现场等。所导引的行为可以根据自我观察、自陈报告评定、由他人观察报告评定、动作反应、口头与文字作答等来呈现。所导引的行为表现方式呈现出多样化。心理是产生在S—R中的中介变量O,只能间接推测出来。这样看来,行为才是测量所要研究的重点,因为心理依赖它间接推测出来。行为最好是总体的行为,与刺激联结起来推测时说服力最强,但是总体有时难以得到(如限于人力、物力、财力或信息不全面),或即使有时能得到,但也不划算。例如,我们要验血来做生理检查,最好是将血全部抽光放到一个大缸里,这是总体,但有谁愿意去抽血呢?抽完了还没检查完人就已经死了,不划算。那怎么办呢?医生通常只要求验血者抽取部分血液来代表总体血液就可以了,结果一般都比较准确可靠,这是因为只要抽取的血液样本有一定的代表性,就无须抽取总体血液,验血仍然有保障。同理,对于行为也是一样的,不一定需要行为总体,行为样本也是可以的,只要它具有一定的代表性,是能反映被试行为特征的一组行为,就可以了。但是,行为样本仍然存在某种程度的偏差,不能反映被试全部的心理功能。然而,只要行为样本具有足够的代表性,就可以推测行为总体。

有了行为样本,就可以用来制定常模。常模是用来提供比较的标准,是标准化样本分数的分布。在许多情况下,常模是一组有代表性被试群体的平均测验分数。这个平均测验分数表示的是普通人的一般状况。解释个人的测量分数就是将这一分数与常模分数相比较,看该分数高于或低于常模分数多少。例如,在能力测量领域,如果某一个人的测量分数高于常模分数,那么此人的能力水平高于普通人的平均水平;相反,如果某一个人的测量分数低于常模分数,那么此人的能力水平低于普通人的平均水平。常模既然是一组有代表性的被试群体的平均测验分数,那么编制测验常模的关键是要抽取有代表性的被试样本,它要求按照科学的抽样原则抽取样本中的每一个个体。这里需要特别说明的是,不要把常模的概念与我们通常理解的标准的概念混淆起来。标准指的是理想上期望达到的程度,而常模指的是被试群体实际达到的程度。以常模为参照编制的测验叫常模参照测验,以标准为参照编制的测验叫标准参照测验。

第二个要素:客观化。客观化是指测验项目的功能与强度、整个测验能测什么与测量精度的大小等,要经由“事实”来检验和判定。施测要结构化、规范化,要严格控制误差。评价心理测验的客观化指标包括“四度”,即信度、效度、难度和区分度。总之,心理测验要采取“实证模式”。

评价一个测验是否科学的重要指标是信度和效度。信度是指一个测验的可靠性,即用同一测验多次测量同一团体所得结果之间的一致性程度。我们用钢片卷尺去测量一根木杆的长度,所得结果是可靠的,因为无论是由一个人数次测量,还是分别由数个人去测量,所测得的结果都将是高度一致的。如果改用橡皮软尺去测量一根木杆的长度,那么一个人数次测量或分别由数个人去测量的结果就难以高度一致。这就是说,橡皮软尺这种

测量工具的信度不高,弹性太大,不稳定。由此可见,信度是衡量测验科学性的最基本的指标。效度指的是一个测验的有效性,即一个测验在多大程度上能够测到它所要测量的心理特质。若一个测验所测得的不是它所要测得的特质,则这个测验就是无效的。例如,智力测验所要测得的特质应该是智力,如果一个智力测验测到的不是智力,而是知识,那么无论它的信度有多高,这个智力测验对于测量智力都是无效的。由此可见,效度是衡量测验科学性的最重要的指标。

评价一道题目是否合理的重要指标是难度和区分度。难度和区分度是题目质量重要的参考指标。难度反映的是题目的难易程度,而区分度反映的是题目的鉴别能力。

诚然,一份测验或一道题目好还是不好,都要经由“事实”说了算,不是某个人说了算的(这是主观的),这才是客观化。

第三个要素:标准化。标准化是编制测验的一个重要步骤,也是使用测验的一个重要条件。为了使接受测量的不同个体所获得的分数具有可比性,测验的条件必须对所有的个体都是相同的。在相同的测验情境中,唯一的自变量是正在接受测量的个体的心理特质,这样的测量结果才具有可比性。所谓标准化,是指测验的编制、实施、评分以及分数解释的程序的一致性。测验标准化需要具备下列条件。

(1) 测验编制标准化。测验编制要求编制的程序规范,题目合理,所有接受测量的个体实施相同或等值的测验内容。如果测验内容不同,那么所测得的结果便没有可比较的基础。

(2) 测验实施标准化。这是测验标准化的第二个条件,即所有接受测量的个体必须在相同的施测条件下接受测验,其中包括:①相同的测验情境;②相同的指导语;③相同的测验时限。

(3) 测验评分标准化。测验评分标准化要求评分结果具有客观性,也只有当评分的结果具备了客观性,才能将测量分数的差异归之于个体心理特质的差异。为此,测验中所制定的评分规则,要足以使不同的评分人的评分结果保持最大限度的一致。

(4) 测验解释标准化。编制测验的一个重要步骤是编制测验常模。在心理测量领域,由于测量分数没有绝对零点作为参照点,所以,孤立地看待一个测量分数是没有什么意义的,只有将该测量分数与他人的测量分数相比较,才能显示出它的意义。常模的功能就是给解释测量分数提供一个可比较的参照点。对测验分数进行解释时,既要参照测验所提供的常模,也要参考其他一些重要的指标,如信度、效度等。

至此,我们可以建立起心理测验三个要素的框架,如图 1-4 所示。

这样看来,根据图 1-4 的心理测验三个要素的框架,本书第二章~第六章都统一到心理测验的三个要素之下,心理测验就成为经典测验理论(Classical Test Theory, CTT)的核心概念、中心思想或灵魂!

根据前述,心理测量就是要获得有关人们心理特性的数量化认识。随着测量实践的发展,对测验过程与结果进行量化分析的技术,即心理计量学的理论与技术,就不断地发展起来。其中,最先发展起来的心理计量学理论是经典测验理论。它认为测验中的观察分数等于真分数与随机误差分数之和。到 20 世纪 30 年代末,这种理论已经提出了一整套分析测验量化的技术,如常模(第二章)、信度(第三章)、效度(第四章)、难度与区分度(第五章),还发展了比较完善的标准化技术(第六章),包括测验编制标准化、测验实施标准化、测验评分标准化和测验解释标准化。通过经典测验理论,也编制了一些具体的实用的心理测验,如智

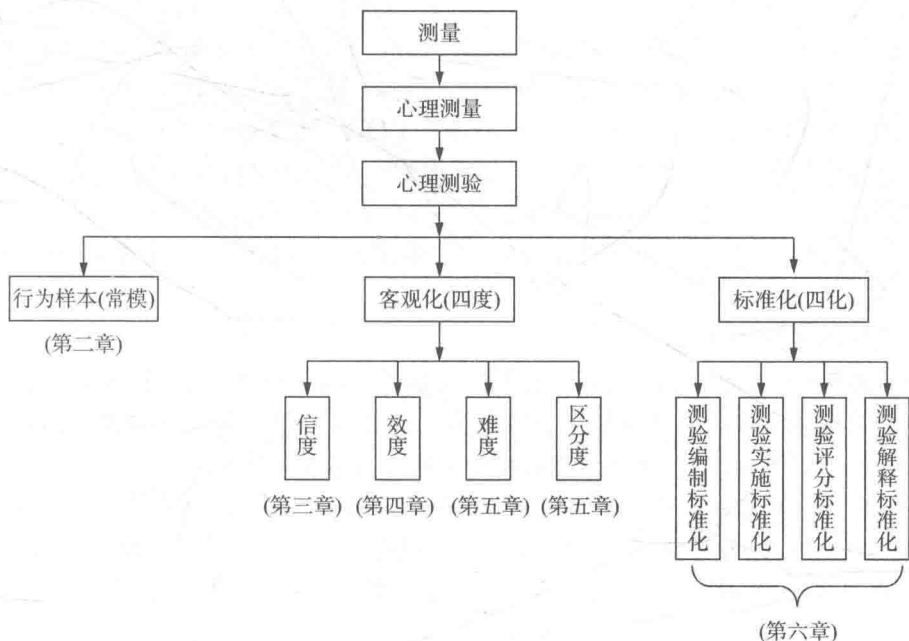


图 1-4 心理测验三个要素的框架

力测验(第七章)、人格测验(第八章)和心理评定量表(第九章)。

自 20 世纪五六十年代起,在经典测验理论基础上,一些心理测量学专家又陆续提出了两个现代测验理论,包括项目反应理论(Item Response Theory,IRT)(第十章)和概化理论(Generalizability Theory,GT)(第十一章)。这两个现代测验理论从不少方面深化和拓展了经典测验理论,但必须使用更为高深的现代数理统计学方法,一般都要借助计算机来完成分析与计算。

第二节 心理测验的分类与功能

心理测验的根本作用在于判别人的心理的个别差异。但是,人的心理是复杂的。因为心理测验需要在不同条件下为不同的目的服务,所以心理测验的分类与功能也就有多种。

一、心理测验的分类

分类要遵照一定的标准。分类的标准不同,分类后所得的结果也就不同。心理测验按不同的标准,可以得到下述的不同分类。

(一) 按测验所测的心理特性分类

(1) 能力测验。能力测验测量的心理特性是能力,包括一般能力测验、能力倾向测验、学业成绩测验、学业成就测验等。一般能力测验又称智力测验。能力还可分为实际能力和潜在能力两种。对潜在能力的测量,通常又叫对能力倾向或“性向”的测量。学业成绩测验和学业成就测验也可视为能力测验。学业成绩测验是指对个体在一个阶段的学习或训练之后所掌握的知识和技能的发展水平进行学习成绩测定的测验,如语、数、英等学科测验。学

业成就测验是指在特定领域为检测应试者对有关知识和技能的掌握程度而设计的测验,是测量人在接受教育后所取得的学业成就的测验,如大学入学时学业水平测验、驾照考试等。

(2) 人格测验。人格测验测量的心理特性是人格,是个性中除能力以外的部分,亦可看作是能力测验,主要测量性格、气质等方面的个性心理特征及其相关行为。

(3) 心理评定量表。心理评定量表的主要用途是对个体或群体的心理状况作出评定。使用心理评定量表,不但要对个体或群体的心理与行为作出评定,而且还要对与心理与行为相联系的身体、环境和社会等因素作出评定。心理评定量表是用来量化观察中所得印象的一种心理测量工具。

(二) 按测验的功能分类

(1) 智力测验。如比内(Binet)智力量表、韦氏(Wechsler)智力量表、瑞文(Raven)推理测验等。智力测验测量的心理特性是一般能力。随着智力理论研究的深入,人们不但重视对逻辑分析智力的测量,而且更加重视对社会实践智力的测量。

(2) 特殊能力测验。如音乐、绘画、机械、文书、飞行员测验,公务员考试,专业技能考试等。特殊能力是指音乐、美术、体育、机械、飞行、创造力等方面的特殊才能。随着社会发展的加速,人们不但重视对音乐、美术等特殊能力的测量,而且更加重视对创造性能力的测量。

(3) 人格测验。如 MMPI、16PF、EPQ、罗夏墨迹测验、主题统觉测验(TAT)等。人格测验主要测量性格、气质、情绪状态、人际关系、动机、兴趣、态度等,一般不把能力包括在内。

(4) 心理评定量表。如 SCL-90、SDS、SAS、SES、LES、SSRS、CSQ 等。心理评定量表主要包括心理卫生综合评定量表、情绪及相关问题评定量表和应激及相关问题评定量表等。

(三) 按测验材料的性质分类

(1) 文字测验。或称纸笔测验、言语测验。如 MMPI、16PF、EPQ、韦氏智力量表中的言语测验等。所用测验材料为文字,作答也用文字。此类测验易受文化背景的影响,文化程度太低者更难于接受。

(2) 非文字测验。或称操作测验、非言语测验,如罗夏墨迹测验、TAT、Raven、韦氏智力量表中的操作测验等。大多要求对图形、实物、工具、模型进行辨识和操作,无须使用文字作答,故不易受文化因素的影响。

(四) 按测验材料的严谨程度分类

(1) 客观测验。或称自陈量表,自陈就是自我陈述、自我报告。其特点:刺激材料的意义明确、作答有明确规定,如 MMPI、16PF、EPQ 等。客观测验一般为调查表。所谓调查表,就是了解被试情况的细目表,相当于一个标准化的访谈提纲。一个客观测验往往包含几个分量表,可以同时测量几个特质,通常采用纸笔形式,但容易产生反应心向和反应方式。所谓反应心向是指作出不符合实际情况的回答,以便使别人对自己形成某种特殊印象。所谓反应方式是指独立于题目内容以外的反应倾向,如猜测、默认、折中等。

(2) 投射测验。其特点:刺激材料的意义不明确、作答没有明确规定。测验刺激无组织(如墨迹图),没有明确意义,问题模糊,对被试的作答内容要求并无明确规定。部分人格测验属此类型,如罗夏墨迹测验(联想法)、TAT(构造法)、词语联想测验(联想法)、句子完成测验(完成法)、绘画测验(表露法)等。图 1-5 是罗夏墨迹测验其中一张图;图 1-6 是 TAT 其中一张图。

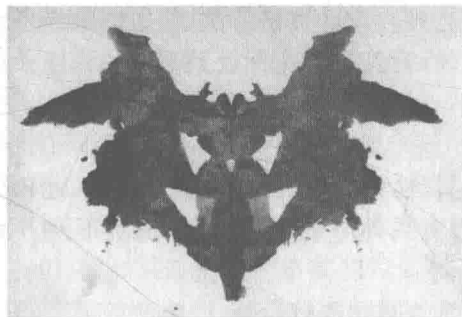


图 1-5 罗夏墨迹测验中的图片示例



图 1-6 TAT 中的图片示例

客观测验与投射测验在以下三个方面存在不同：一是人格的假设，客观测验认为人格是意识层面的，而投射测验认为人格是潜意识层面的；二是测验的材料，客观测验采用的是意义明确的调查表，而投射测验采用的是意义不明确的材料；三是被试的反应，客观测验通常是封闭的，而投射测验通常是开放的。客观测验的优点：容易实现标准化，效率高；缺点：容易出现反应偏差。投射测验的优点：可以对人格做综合分析，被试不容易有虚假反应；缺点：信效度较低，不容易标准化，对主试的要求较高。

（五）按接受测验对象的规模分类

（1）个别测验。每次只对一名被试施测，即以一对一的形式进行，如比内智力量表、韦氏智力量表等。其优点是主试对被试的作答反应有较多观察与调控机会，如对一些幼儿和特殊群体（如弱智者等）进行测验，非得面对面施测，则采用个别测验较好。但是，这类测验费时，短期内不易收集到大量人群的资料，施测手续复杂，主试应专门训练，因而使用中 also 显露出一定的局限性。

（2）团体测验。可以同时多名被试施测，时间经济，但对被试观察和控制差，易产生测量误差，如陆军甲种和乙种测验、MMPI、16PF、EPQ 等。世界上第一个团体智力测验是陆军甲种和乙种测验，产生于第一次世界大战期间，其目的是分门别类甄选近 200 万新兵，陆军甲种测验适合有一定文化水平、母语为英语的被试，共有 8 个分测验，为文字测验；而陆军乙种测验适合文化程度较低或母语为非英语的被试，共有 7 个分测验，为非文字测验。

（六）按测验对作答行为的要求分类

（1）最高成就测验。或称最佳行为测验、最高作为测验、最高行为测验。此类测验要求被试的作答行为尽可能地做好，其内容与认知过程有关，答案有正误之分，能力测验属于此种类型。

（2）典型行为测验。此类测验要求被试按通常习惯方式作答，其内容与典型行为有关，答案无正误之分，人格测验属于此种类型。

（七）按测验分数解释的参照体系分类

（1）常模参照测验。参照体系是常模，即所测被试团体在所测特性上实有水平的分布状态。大多数能力测验和人格测验属于常模参照测验，如我国的中考与高考、EPQ、MMPI 等。这类测验上的观察分数要转换成导出分数，然后根据导出分数并利用常模资料来说明被试在所属团体中的相对地位，从而解释清楚测验结果的应有意义。

（2）标准参照测验。参照体系是社会的应有要求，即所设置的标准。一些能力测验属