

心理测验的编制技术与使用方法

陕西旅游出版社

王晚钧 朱鹏 编著

内容提要

本书是我国第一本系统论述心理测验编制技术和使用方法的专著。全书共分十一章，分别阐述了优秀心理测验的特点，智力测验、人格测验、客观测验、投射测验、标准组测验、因素分析测验、计算机适应性测验的编制技术，以及信度、效度、项目分析、分数转换、常模等检验心理测验科学性的方法，最后讨编了心理测量科学的发展趋势。

本书适用于心理学、教育学工作者，临床医生，人事部门和厂矿企业管理部门的有关人员，社会科学工作者以及专门从事测量学工作的研究者。也可做为文科大学生及研究生的辅助教材。

目 录

第一章 优秀心理测验的特点

- 一 量表的类型..... (1)
- 二 信度..... (4)
- 三 效度..... (9)
- 四 区分度..... (14)
- 五 标准化与常模..... (14)

第二章 智力与能力测验项目的编写

- 一 智力与能力测验项目编写的一般要求..... (17)
- 二 智力与能力测验的理论基础..... (21)
- 三 智力测验项目编写..... (25)
- 四 能力和成就测验项目编写..... (35)

第三章 人格测验项目的编写

- 一 人格的性质..... (64)
- 二 人格特质..... (66)
- 三 优秀人格测验的一般特点..... (68)
- 四 人格测验项目编写中常遇到的问题..... (70)
- 五 人格测验项目编写..... (73)
- 六 反应定势的消除..... (82)

第四章 客观测验和投射验测的设计

- 一 客观测验

(一)	客观测验的一般概念.....	(96)
(二)	客观测验设计的一般要求.....	(99)
(三)	客观测验的得分设计.....	(101)
(四)	设计人格及动机客观测验暗示的消除.....	(104)
(五)	客观测验有待解决的一些问题.....	(111)
二	投射测验	
(一)	投射测验的一般概念.....	(114)
(二)	投射测验设计中的一些问题.....	(117)
第五章	其它类型测验的编制	
一	标准组测验	
(一)	准标组测验的一般概念.....	(121)
(二)	编制标准组测验应注意的问题.....	(125)
(三)	标准组测验编制的步骤.....	(127)
二	因素分析测验	
(一)	因素分析的原理、逻辑和解释.....	(129)
(二)	因素分析测验的编制.....	(133)
三	计算机适应性测验.....	(136)
第六章	测验信度的的计算方法	
一	测验分数变化的原因.....	(140)
二	经典信度理论.....	(144)
三	估计信度的方法.....	(146)
四	信度估计与误差.....	(154)
五	信度系数的应用.....	(155)
六	影响信度的因素.....	(158)
七	信度的特殊问题.....	(163)
八	信对度指标的要求.....	(168)

第七章 测验效度的估计方法

一	相关法	(172)
二	因素分析法	(176)
三	专家判断法	(182)
四	多特质—多方法矩阵	(183)
五	效度范围	(186)

第八章 项目析分

一	项目分析的目的和特点	(188)
二	离散度分析	(190)
三	难度分析	(192)
四	项目的区分度分析	(197)
五	项目特性的相互关系	(201)
六	项目特征曲线和项目反应理论	(205)

第九章 测验分数的转换

一	线性转换	(215)
二	面积转换	(219)
三	非常态化曲线的转换	(221)
四	分数的等值转换	(224)

第十章 建立和使用常模的方法

一	建立常模的步骤	(231)
二	常模类型	(234)
三	解释常模应注意的问题	(244)

第十一章 心理测量的发展趋势

一	测验使用者的角色	(248)
二	测验编制的技术方法	(251)

三 测验分数的解释.....苏式林苗的测验解释 (259) 章

(后记.....苏式林苗的测验解释 (263) 章)

(171) 测验分数的解释.....苏式林苗 (171) 章

(182) 测验分数的解释.....苏式林苗 (182) 章

(183) 测验分数的解释.....苏式林苗 (183) 章

(184) 测验分数的解释.....苏式林苗 (184) 章

(185) 测验分数的解释.....苏式林苗 (185) 章

(186) 测验分数的解释.....苏式林苗 (186) 章

(187) 测验分数的解释.....苏式林苗 (187) 章

(188) 测验分数的解释.....苏式林苗 (188) 章

(189) 测验分数的解释.....苏式林苗 (189) 章

(190) 测验分数的解释.....苏式林苗 (190) 章

(191) 测验分数的解释.....苏式林苗 (191) 章

(192) 测验分数的解释.....苏式林苗 (192) 章

(193) 测验分数的解释.....苏式林苗 (193) 章

(194) 测验分数的解释.....苏式林苗 (194) 章

(195) 测验分数的解释.....苏式林苗 (195) 章

(196) 测验分数的解释.....苏式林苗 (196) 章

(197) 测验分数的解释.....苏式林苗 (197) 章

(198) 测验分数的解释.....苏式林苗 (198) 章

(199) 测验分数的解释.....苏式林苗 (199) 章

(200) 测验分数的解释.....苏式林苗 (200) 章

(201) 测验分数的解释.....苏式林苗 (201) 章

(202) 测验分数的解释.....苏式林苗 (202) 章

(203) 测验分数的解释.....苏式林苗 (203) 章

(204) 测验分数的解释.....苏式林苗 (204) 章

(205) 测验分数的解释.....苏式林苗 (205) 章

第一章 优秀心理测验的特点

一种心理测验要成为优秀的测验，应具备某些特点。它至少应该是一个等距量表；具有较高的信度，效度以及区分度；有优秀的，代表性很强的常模；并且具有实际应用价值。

本书作者力图集中探讨的重点问题是：如何将这些特点完美而富有创造性地融汇于编制测验的整个过程，在此之前，有必要对一些基本概念加以讨论和区分，这些概念不论在编制测验中，还是在使用测验时都非常重要。

这些概念之所以重要，是因为它们能提高心理测量的正确性和准确性，并且和心理测量的科学性密切相关。我们认为，自然科学的进步有赖于优秀测量方法和手段的发展，心理学也不例外。此间讨论的每一个特点都在于提高心理测验的科学性。

一 量表的类型

在测量学上，有各种不同类型，或不同水平的测量量表。斯蒂芬斯（stevens）曾多次指出，在测量学上，有四种测量量表对研究和应用十分重要，它们是，名称量表，等级量表，等距量表和比率量表。

名称量表（nominal scale）用于将人或物进行分类。

在名称量表测量中，被测的成员用名称或符号代替。但是，这种代表符号完全是任意的，是指派性质的，只要原则上能达到区分被试的目的，可用任何符号来代表。这是一种比较粗糙的测量，因为它只能做一般的区分，无法提供个人或事物具体的测量状况。

等级量表 (ordinal scale)。这类量表测量的是个体在某一团体中所处的位置。等级测量能反映出事物之间的区别性和次序性，但这种区别或次序之间的差距并不相等，所以等级测量只有序列区分，而不具等距性质。

等距量表 (interval scale)。这类量表测量的是人或物之间的差异，但它有一个特殊要求，即一定的差异量应在表上呈等距分布，而且所用的单位应相同。由于个体差异是心理测量学家最关心的问题，所以等距量表在教育 and 心理测量中非常有用。很多测量实际上都是以等距的形式编制的，然而这类量表编制起来难度较大，其中困难之一就是无法确定哪种差异具有等距性，这主要是参考系的问题。

比率量表 (ratio scale) 测量的是两个或两个以上被试之间特点的比率。这里的比率是用确切的，有一定的参考标准的单位加以测量的。因而和等距量表有区别。在等距量表中，测量单位虽是等距的，然而却有任意性，其零点是任意指派的，无绝对零点可言。在比率量表中，零点不是任意的，无论什么，只要做比率测量，其零点总有确切的意義。

四种测量量表有明显的等级差别。名称量表是最原始的和信息量最少的量表，它反映不出被试的特点；次序量表较名称量表高一等，它除了具备等级量表的性质外，还可以将被试按次序排队；等距量表的功能高于前两者，除具有名称

量表和等级量表的功能外，它还具有反映人或物之间相对差别大小的功能；比率量表最完美，除具有前三种量表的特点外，还具有决定和比较的特点。

在数学处理上，四类量表的等级差别也很明显。比率量表几乎可用所有统计方法对结果加以处理；等距量表稍次，多数只能用平均数，标准差，积距相关，t检验，F检验等，如绝对差异量，等比量就不适用；等级量表更次，只能用中位数，百分位，等级相关，秩次检验等；最次是名称量表，仅能用次数，百分比， χ^2 等方法。

目前心理测量科学的发展一般达到的水平可以完全达到等级量表的水平，勉强达到等距量表的水平。这里的勉强指的是心理学家常由于数学处理上的方便，将实质上属于等级量表水平的心理测验按等距量表处理。而距比率量表水平相差甚远。这主要是由于心理现象无法找到绝对零点这一困难造成的。

对以上四种量表类型的分析可以清楚地看出，从理想的角度来看，心理测验的编制者应以编制比率量表为目的，但就目前心理测量学的发展上看，这几乎不可能。退一步讲，如果要使测量结果适合于多种统计分析，我们更倾向于编制等距量表。由于测验效度研究不可避免地要使用这些统计分析，由于心理测验可以产生比从假设中得出定性结论更为优越的定量分数，因而结果很明显，没有等距量表是不行的。事实上，正如布罗恩（Brown）在1976年指出的那样，多数心理测验都近似于等距量表，并且将测验分数视为由等距量表得出的有用结果。

二 信度

信度的最一般含意是指测验的可靠性。估计信度的目的在于确定测验分数的一致性或者测量误差的大小。平行测验模式从概念上提出了估计信度的可能性。根据平行测验模式，可以编制同一种测验的两种形式，它们在使用上是等价的。这就是说，将被试从A测验中得到的真分数看做是B测验中的真分数，而将测验分数之间的差异看做是测量误差。如果两种测验分数差异很大，便可断定测量误差是测验分数变异的主要原因；如果两种测验分数很接近，可以认为测量误差很小，此测验的可靠性很好。

在实际估计信度中，有四种方法很常用，并分别具备平行测验的许多优点。它们是：

1. 重测信度

重测信度是估计测验分一致性的最早的一种方法，从表面上看，它也是最实际的方法之一。重测信度通过对同一组被试在不同时间，不同场合两次施测同一测验，直接估计测验分数之间的一致性。估计重测信度分下列三个步骤：

- (1) 对被试施测某种测验；
- (2) 在稍后时间内对同组被试施测同一测验；
- (3) 计算两次分数的相关，并以相关高低评价信度。

对重测信度的估计，一般采用皮尔逊积差相关公式计算。

很显然，重测信度很重要。如果一个测验不能在不同时间，不同场合情况下对同一组被试测得相同的分数，那么一切将无从谈起。重测信度的测量是实质性的，其相关系数能

够反映出测验在时间上的一致性。按照吉尔福特(Guilford)的观点,重测信度系数不应低于 70,低于此系数的测验,由于分数的标准误太大,其结果的可靠性无法得到令人满意的解释。关于标准误的意义和内涵在第六章讨论经典测量误差理论时再提及。

尽管重测信度的计算相对简单,但要注意的问题是,

(1) 不应在把两种测验分数合并时人为地提高信度系数;

(2) 样本必须能够代表所要测的群体。

2. 复本信度

从表面上看,用复本法估计信度和平行测验模式最接近。使用这种方法的关键在于编制出从内容,反应过程到统计分析等方面都等值的复本。例如,在卡特尔十六种人格特质问卷中就有 A、B、A + B 三种等值测验。用复本法估计信度有以下三个步骤:

(1) 对一组被试施测复本 A;

(2) 对同组被试施测复本 B (相隔一段时间或者紧接着);

(3) 计算二种复本测验分数之间的相关,并用相关系数估计复本信度。

复本信度可以解决重测信度中存在的一些问题。例如,由于两种测验形式不同,尽管两种测验可能出现类似的反应项目,然而相同项目出现的几率很低。这就有可能降低测验的回忆效应。虽然被试在第一种测验中的反应可能会对第二种测验的反应产生影响,但相对于重测信度而言,反应效应得到了适当的控制,可以认为反应性效应的影响在复本信度

中较弱。由于复本信度减弱了重测信度中的回忆效应，所以，在两种复本测验之间不必间隔很长时间，可以在一种复本测验后，立刻进行第二种测验。

虽然复本信度避免了重测信度中存在的一些问题，但从技术上仍有待于商榷。例如，尽管复本法和重测法都需要施测两次，但复本法由于要测试两种不同的复本，所以花费更大，实际困难大于重测法；编制一个测验的几个复本，如果不是不可能，其难度也将很大。如果等值测验编的不够理想，那么用其分数的相关估计出的信度，其误差来自于测验本身呢？还是来自于复本之间呢？有时很难说清楚；再有，仅仅为了估计测量信度而编制几套等值测验在应用上也不实际。

3. 分半信度

分半信度，故名思义，就是将测验项目分成两半来估计信度。其推理过程直接而明瞭，最简单的分半形式就是将测验项目分成两半，并且把分成的两半看成是复本。用分半法估计信度包括以下三个步骤：

- (1) 对一组被试施测某种测验；
- (2) 将测验项目分成两半；
- (3) 求两半测验分之间的相关，并以相关系数做为估计此测验的分半信度。

分半信度从理论上和实际应用上避免了重测信度和复本信度存在的一些问题。首先，分半法不用对被试测验两次，就可达到估计信度的目的。其次，由于分半信度的估计只需施测一次就可完成，所以，回忆效应，反应性效应以及随时间变化的特殊效应都降低到最小程度。这样，测验分半之间分

数的不一致性只可能反映出测验本身反应的不一致性，而不涉及两个复本或两次施测之间的误差。

对测验项目分半的方法很多，其基本原则是在进行项目分半时，所分的两部分应尽量做到相似，也就是说，它们包括的反应群的几率相等。为了减少系统误差，最简单，最常用的分半方法是奇一偶分半。即：奇数项目做为一个分测验，偶数项目做为另一个分测验。这样分的好处在于，两个分测验都包括了测验开始，中间和最后的项目，并且在项目数量上相等。

4. 内部一致性信度

内部一致性信度考察的是测验项目之间的自相一致。也叫同质信度。用内部一致性方法估计信度的步骤为：

(1) 对一组被试施测某种测验；

(2) 计算项目间的相关并计算这些内相关的平均值；

(3) 求内部一致性系数，并以此做为估计内部一致性信度的指标。

经典信度理论求内部一致性信度通常使用库德 (G · F · Kuder) 和理查逊 (M · W · Richardson) 提出的两个公式 $K-R_{20}$ 和 $K-R_{21}$ 。这两个公式对强迫性选择测验较为合适。在多重选择测验中，常用克隆巴 (L · J · Cronbach) 的 α 系数法。

内部一致性信度和以上讨论的几种信度有着密切的关系。从数学上和概念上都能证明这些关系的存在。内部一致性信度和分半信度有着数学上的关系。例如用 α 系数法得出的内部一致性信度，既代表这种信度最常用的和最一般的形式，又代表所有可能分半方式的平均信度系数。也就是说，

如果有一个30个项目的测验，将它所有分半形式下的相关系数都计算出来，那么这些信度系数的平均数将等于 α 系数。分半信度和内部一致性信度也有区别。这种区别主要是分析单位不同，分半信度将测验分成两半加以估计，内部一致性信度将各个项目和其它所有项目加以比较，进而做出估计。

内部一致性信度和一般信度在概念上也有密切关系。其中很重要的一点就是内部一致性信度在于回答“测验为什么可靠”的问题。内部一致性系数是测验项目和项目平均内相关的函数。如果把每个测验项目看做是一种可观察的行为，那么，内部一致性系数所代表的信度就是可观察行为的数目的函数，每个项目所代表的可观察行为也是在同等观察条件下其它项目的函数。如果每个项目都有效，并且如果项目数量很大，内部一致性系数将很高，这时我们可断定此测验可靠。

心理测量工作者都希望编制出具有高度内部一致性信度的测验，其原因显而易见，如果测量同一特点的测验有一部分测量某个变量，而另一部分与此变量不一致，那么无法充分测出此变量。这表明，一个测验要有效（即测量出所要测的特点）那么它必须一致，所以心理测量学强调内部一致性信度。高信度是高效度的先决条件，这是心理测量学界持有的普遍观点。唯有卡特尔（Cattell）等研究者提出了不同看法。卡特尔认为，高内部一致性信度实际上和效度相互矛盾。其理由是，任何一个测验项目的覆盖面都小于要测的变量标准。因此，如果全部测验项目高度一致，它们也应呈高相关。这样，一个有信度的测验仅能测量出一个变量的很少一部分。这里，有两点应引起注意：（1）克隆巴的 α 系数是

随项目内部相关的提高而增加的；（2）在任何的多变量预测研究中，当变量互不相关时，测验和效标之间的最大多重相关可以求得。很明显，如果两个变量高相关，那么其中一个必定无法提供更多的新信息。按照卡特尔的观点，最大的效度来自于测验项目不完全相关，但是每个项目却都和效标具有正相关。我们认为，卡特尔的理解是正确的，然而，就目前掌握的资料和心理测量理论发展现状来看，尚没有测验编制者着手去编制既和效标相关，而相互之间又不相关的测验项目。巴特利等人曾于1982年考察过卡特尔自己编制的人格问卷（卡特尔16种人格因素问卷，简称16PF），卡特尔自己也曾努力这样做过，然而结果并不理想。现在心理测量学界普遍坚持的观点仍然是：精确的效度测验是高度一致的。

三、效度

这里首先讨论效度的性质，因为它是优秀测验的第二个特点。正象对信度的讨论一样，对效度讨论的目的在于使读者充分掌握效度，了解在测验编制过程中，应把效度看做另一个目标。

说一个测验有效度是指该测验测到了它所要测的特点或心理变量。但仅凭这一点，尚不能完全表达效度的全部含义，相反，它带来了这样一个问题，即怎样才能知晓一个测验测到了它所要测的特点。实际上，有许多考察测验效度的方法，每一种都代表着效度的不同方面。一般认可的效度有以下几种：

1. 内容效度

内容效度是通过分析测验内容和测验项目，来考察测验对目标行为测量的有效性。

考察内容效度的第一步是对测验项目进行分析，如果所有项目均和测量目的一致，并且如果测验项目真正代表了所要测的目标行为，那么测验就有效。这里，对内容范围的分析是建立内容效度的基础。所谓内容范围，指的是要测特点所包含的一组整体行为。通过对这组整体行为的分析，可以明瞭他们和要测特点的关系，在此基础上编制的测验项目将真实地代表了这组行为，因而能够对要测特点进行准确的测量。在物理等典型自然科学的测量中，内容范围相对确定，然而在心理测量中，很多变量，如智力，人格等很难准确的确定出清晰的内容范围，有时连内容分界也比较模糊。所以心理测量内容效度的确定相对困难。但我们可以大致上按照以下基本过程评估测验的内容效度：

(1) 分析内容范围；

(2) 确立内容范围所包括的各个部分，并编制或修改出可操作的测验项目，把它们归类到各个部分中去；

(3) 将测验目的和内容范围结构加以比较。建立内容效度的关键是考察测验在一个特殊内容范围内的取样是否充分。这一步工作以合理的判断为主，而确立内容范围又是系统而可靠判断的关键。对内容范围的分界和结构定义得越细，内容效度的判断也越可靠。

2. 建构效度

人的许多心理品质都具有抽象的性质，如幸福感、智力、社会化、人格等。这些品质很难定义出字面上，或物理意义上的概念。然而为了应用，为了发展和完善心理学理论，我们必须测量它们。

测量抽象特点的问题并非心理学所独有，物理学家常常

测量隐藏在事物内部的，看不见的特点。例如，物体的质量本身既看不见，也摸不着，然而，这类抽象的物体特性很重要，而且可以被测量。我们把质量，幸福感、智力、人格等抽象的特点称之为“建构”（Construction），建构的概念代表着科学上的抽象，它能帮助科学家从有关的现象和观察中总结出规律。

建构有两个特点：

1. 建构在性质上是规律的抽象概括；
2. 建构和具体的，可观察的事件或实体相联系。

心理测量是以具体的，可观察的行为为基础的操作过程。因此，在心理测量中，为了获得行为样本，为了确定测验能否测出心理学建构，必须将抽象的特点转化为具体的，行为化了的，可操作的项目，这实际上是特定行为和抽象建构的关系问题，我们称之为“建构的解释”。它是确定测验是否有建构效度的关键。

建构的解释过程有以下三个步骤：

1. 确定和目标建构有关的行为；
2. 确定其他建构，考察这些建构是否和目标建构有关；
3. 确定其他建构的有关行为，在分清各建构之间关系的基础上，确定每一种和目标建构有关的行为。

通过以上步骤，可以揭示建构群之间以及建构与行为之间的关系，这种关系系统，通常称为“法则网络”。法则网络分析的越细，对建构行为分析的就越全面，对建构的解释也越精确。

确定建构效度的方法，并无一定的模式可依。因为建构效度的建立从测验最初编制就开始了，测验编制的整个过