

现代社会心理丛书



心理自测

PSYCHOLOGICAL SELF-TESTING

陈雪枫
莫雷 / 著

暨南大学出版社

G449.1

453789

C60

·现代社会心理丛书·

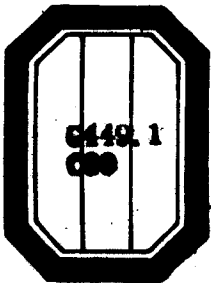
心理自测

PSYCHOLOGICAL SELF - TESTING

陈雪枫 莫雷 编著



00453789



暨南大学出版社

图书在版编目 (CIP) 数据

心理自测/陈雪枫, 莫雪编著. —广州: 暨南大学出版社, 1999.1

(现代社会心理丛书)

ISBN 7-81029-545-4

- I . 心…
- II . ①陈…②莫…
- III . 心理测验
- IV . G449 . 1

DV79/21

出版: 暨南大学出版社出版(广州·石牌)

印刷: 广东省汕头长途电信印刷厂印刷

发行: 新华书店经销

开本: 850×1168 1/32

印张: 11

字数: 27万字

印数: 10001-15000册

版次: 1996年11月第1版

印次: 1999年1月第3次印刷

定价: 16.50元

目 录

第一章	心理测验概述	(1)
第二章	气质自测	(19)
第三章	性格自测	(52)
第四章	能力自测	(88)
第五章	心理健康自测	(146)
第六章	职业素质自测	(187)
第七章	社交心理自测	(233)
第八章	恋爱心理自测	(265)
第九章	家庭心理自测	(305)
附 录	参考资料	(346)

第一章 心理测验概述

孟子曰：“权，然后知轻重；度，然后知短长。物皆然，心为甚。”可见，早在我国战国时期，已经出现了关于心理测量的基本思想。而三国时刘劭在《人物志》一书中提出的“观其感变，以审长度”，则更进一步地阐述了心理测量的一条基本假设，即通过一个人的行为变化来推测其一般的心理特点。后隋炀帝创行开科取仕，科举制度在我国通行了一千三百多年，这是测量方面无可辩驳的实践成果，然而它的科学性和有效性并不理想。

“世界上最浩瀚的是海洋，比海洋更浩瀚的是天空，比天空还要浩瀚的是人的心灵。”法国作家雨果的这句话使我们领会到人类心灵的领域是多么宽阔，而如此宽阔的心灵可想而知又会有多么复杂。正如世界上没有两张一模一样的面孔，世界上也不会有两颗一模一样的心。在有心理测量的基本思想以来的两千多年里，人们执着地探究人类心理的个体差异。但是，真正的科学测量却始于19世纪的西方国家。

当时有两个方面的因素推动了人类心理科学测量的产生和发展：一方面是社会的急需。19世纪以前，智力缺陷者和精神病人被认为是恶魔缠身，并因此受到了许多非人的虐待。后来欧洲和美洲国家对此的认识和态度逐渐发生了变化，开始对智力落后者和精神失常者实行人道主义，并开设了许多收容所和特别医院。西方工业革命成功后，对劳动力的需求急剧增加，工厂大量雇用童工，许多地方官与工厂主签约，每雇20个童工，必须带雇一个低能者。法国医生沈干于1837年创办了第一所专门教育

智力落后儿童的学校。此后，欧美其他一些国家也陆续设立了低能学校，或在普通学校设立特别班。于是，确立一种鉴别各种精神疾病，以及智力缺陷的统一标准已是势在必行之事。

而另一方面的推动因素则是科学心理学的发展。1879年，法国心理学家冯特正式创立世界上第一个心理实验室，这标志着心理科学的诞生。冯特在他的实验室里确定了一批典型的心理学实验，致力于对感知觉、反应时间、记忆和注意等低级心理过程的量的研究。他的研究方法和思路为科学地测量个体差异奠定了基础。

科学心理测量的创始人首推英国生物学家和心理学家高尔顿。他于1884年在伦敦国际博览会上设立了“人类测量实验室”，在前后若干年时间里做了大量的感觉辨别能力的个体差异的研究与测量，积累了大量资料，并且发明了许多测量的工具。与此同时，高尔顿也开始采用等级评定量表、问卷法和自由联想法等手段，成为心理测验的先驱。不过，“心理测验”这一术语，却是美国心理学家卡特尔于1890年首创，并被人们一直沿用至今。

19世纪末的心理测量的兴趣集中在能力的个体差异的测量，然而当时关于能力的实质的心理学理论研究还很薄弱，人们对于能力的理解混乱而模糊，这导致实际测量出来的结果难以被正确、清晰地分析与解释。直到1904年，英国心理学家斯皮尔曼才首先对智力做出明确的理论阐述，认为智力结构中存在一种一般因素，并从多个方面对它定义：经验的领会、关系推断和相关推断。在他的智力理论的影响下，法国心理学家比奈和医生西蒙首次有能力测量中发现复杂心理过程的重要性，并于1905年制订了世界上第一个智力量表。

从此，心理测验运动开始蓬勃发展，各种各样的测验层出不穷，并且逐渐不再局限于能力的测量，而扩展到气质、性格、兴趣、价值观、人际关系、情感适应和动机等等大量的心理学领域。

(一)

心理测验一经产生，便被广泛用于社会各个领域。随着社会的发展，人们对于心理测验的需求有增无减。而今，心理测验已在社会很多领域中行使着各种各样的功能，扮演着一个相当重要的角色。让我们来看一看这些领域。

教育领域

教育的过程简言之就是教和学的过程，教师教的效果和学生学的效果便直接反映了教育的效果。如何去有效评估教与学的效果呢？大家已经有目共睹，采取各种测验或者考试已成为极普遍的、人们公认的一种重要途径，其结果既是对教师教的成绩的评估，也是对学生学习成绩的评估。

测验成绩为教和学的过程所带来的反馈信息，一方面促使教育的发展，使之不断改进教育体制、教育目标、教材和教法；另一方面成为因材施教的前提，使教师在更加准确客观地了解自己的学生的基础上而采取有的放矢的、更加有效的教学措施。

教育中的学生选拔同样离不开测验。最典型的一个事实是：从上小学到攻读博士学位，每上一个台阶，都有一次升学考试关口，考试成绩领先的学生才能冲破一个又一个关口，直升而上。且不论这种考试制度是否会为教育带来某些弊端，如盲目追求升学率等，但至少这样的考试制度，将有助于提高发现人才的可靠性和科学性。

医学领域

当初心理测验产生的原动力之一就是出于对智力缺陷或精神疾病的鉴别的需要，而今随着社会的发展，人与人之间的竞争日

益激烈，人们越来越多地生活在压力之中，心理咨询与治疗机构也就相应普及起来，在这些机构中，心理测验是应用最广泛的诊断工具。

工业领域

在工业领域，人事测评与管理日益科学化，凭借个人经验选拔人员已不能满足实际的需要。选拔最有可能成功的人，并把他安置在最适合他的岗位上，使得人尽其才，这是目前人事工作者们的核心任务。各种心理测验已成为有效的决策辅助工具。

随着社会消费水平的提高，人们对工业消费品不再仅仅有实用性的要求，渐渐也产生了心理上的需求，比如美观大方的包装，名牌的效果，温馨浪漫的心理感受等等。于是，有关消费者对商品的使用态度、喜欢和满意程度等方面的市场调查迅速普及。而市场调查的重要手段之一就是采用心理测量技术编制的问卷调查法。

军事领域

第一次世界大战前，美国为招募具有战斗力的新兵入伍，并在新兵中选拔军官，就采用了大规模的心理测验。由于要在短短几个月内对 170 多万士兵进行测试，团体智力测验便应运而生。自此，心理测验便在军队中的士兵选拔和安置工作方面发挥着极其重要的作用，军队的战斗力因此而大为提高。

职业咨询领域

随着社会分工的日益细化，新兴行业不断涌现，如今我们面临的是一个庞大的职业王国。如何根据自身的能力、兴趣、个性等各方面特点，准确有效地找到自己的就业方向？关键之一在于准确清晰地认识自身的特点。于是，无论是对于职业咨询机构而

言，还是对每一个个人而言，针对人的各种心理特质而进行测量的心理测验无疑是一种很好的辅助工具。

社会科学研究领域

在教育、社会学、政治学、法学和经济学等许多社会科学领域的研究中，都会或多或少地涉及到人的各种心理及其个别差异。而心理测验的最基本功能就是测量个体差异，因而常被用作这些领域的研究工具，用来收集定量化的资料，从而帮助这些学科的研究建立和检验假设，最终推动社会科学的理论的发展。

在心理测验对社会各种领域做出重大贡献的同时，它对于个人的重要性也越来越不容忽视。人贵在自知之明。一个对自身有清晰而准确认识的人，往往会在人生的重要转折点上作出明智的选择，并以明朗乐观的心态去适应复杂的社会生活。不过，并不是每一个人都能凭借直觉和经验对自身产生清晰认识的。俗话说：“旁观者清，当局者迷。”甚至有人认为，一个人最不了解的就是他自己。在此，可靠而有效的心理测验的作用是已被实践证明了的，它可以帮助个人形成清晰的自我意识和准确的自我评价，树立恰当的人生目标，从而更好地适应社会。

可见，无论对于社会还是对于个人，有关人的心理特质的科学测量已是必不可少。然而，作为心理测量和心理测验来说，它们的科学概念与理论却远不如它们的实践那样在人群中得到普及。

从科学的定义来说，所谓测量，是指根据一定的法则用数字或符号对事物及其属性加以确定。这里的事物及属性，在物理上指物体的长短、轻重、快慢、强弱等，而在心理上则指人的智力、个性、兴趣、态度、学绩等；一定的法则在物理上指物理测量所依据的理论或规则，如杠杆原理、热胀冷缩原理，在心理上则指有关人的各种心理特质的理论，如气质理论、智力理论、动

机理论等。因此，所谓心理测量，是根据一定的心理学理论，使用一定的操作程序，给人的行为确定出一种数量化的价值。

物体的物理特征是看得见、摸得着的，比如人的高矮胖瘦等，因此物理测量可以直接测出人或物的物理特征的水平。然而，人的心理特征却是内在的，看不见也摸不着，不可能直接进行测量。因而心理测量往往是通过人们在面对问题情境时，所表现出来的外显行为来推论其心理特质，这一作法来源于一个基本的心理学理论假设：心理特质是个体特有的、稳定的、可辨别的特征，是在遗传与环境的双重影响下，个体对刺激作出反应的一种内在倾向，表现为一组内部相关或内在联系的行为，比如智力高的人在涉及智力的任务中往往表现得既快又准确；气质外向的人往往表现出活泼、热情、善于社交、合作性高等行为特点。因此，心理测量可以从个体的外在行为模式推知其内在的心理特质水平，它是一种间接的测量。

心理测量又是相对的。作为任何测量而言，都必须有参照点这一测量要素，即当把事物及其属性数量化时，必须有一个计算的起点。现有测量中存在两种参照点，一是绝对零点，如物理测量中的长短轻重等，此时若说某物长度为0，即意味着该物一点儿长度都没有；二是相对零点，如物理测量中的温度以冰点为零点，陆地高度以海平面为零点，此时若说某物温度为零，显然并不意味着一点儿温度也没有。在心理测量学中，没有理想的绝对零点，而只有相对零点，如智商为0，指的是0岁儿童的平均智力水平；学习成绩为0，表明在一次考试中全部没有做对，却并不表明在所测的能力上为0。因此，一般情况下，每个人的心理测量结果都在一个连续体上占据一个位置，我们只是从人与人之间的相对位置，对一个人的某种心理特质水平作出判断。

对于任何测量，客观性是最基本的要求，对心理测量也不例外。测量的客观性就要求测量工具的客观性。心理测量的工具主

要是心理测验，于是就有了心理测验的客观性的问题。

通过观察人的少数有代表性的行为，对于贯穿在人的全部行为活动中的心理特点，作出推论和数量化分析的一种科学手段，就是心理测验。

这一定义首先告诉我们，心理测验的直接测量对象是人的某一类全部行为活动，但其最终目标则是隐含在这类行为中的人的某种心理特质，因而反映出心理测量的间接性特点。

与此同时，定义中又指出测验是通过观察人的少数有代表性的行为来推论其内在在特质水平的。这是因为，在一次测验中不可能囊括与所测特质相对应的全部行为领域。比如测量人的智力：一种智力理论认为，智力这一特质包含了言语理解、推理、记忆、数字运算、知觉速度等多种因素。显然，对每一种因素的测量都可能包含有无数的题目。仅以数字运算中的自然数的加法运算为例，我们可以在这一范围内编制出无数的庞大的题目群。但是在实际测量中，这些题目不可能被一一用到。因此，一次测验所包含的题目只是全部可能题目中的一个样本，这个样本的结构必须和题目总体的结构类似，从而可以较好地代表总体，为进一步的推论提供依据。

定义中的“科学手段”针对于心理测量的客观性，指出心理测验是心理测量的科学工具。然而，测验的科学性与客观并不是凭空而来的，也不是轻易就可以得到保证的，而必须经过测验的编制者和使用者的努力才能达到。

(二)

保证心理测验的科学和客观，首先必须保证测验的质量。对一个测验的质量评估，一般从3个方面着手：项目分析、信度评估和效度评估。

项目分析

在测验编制过程中，首先必须确定测验的目的、对象和内容；然后有针对性地收集大量资料；第三步是选择题目形式，并编写出题目初稿，初稿应有计划中测验题目数的2~3倍，以备筛选；第四步将题目初稿编辑成测验，一般情况下题目由易到难排列；第五步便是预测，即将测验施测于某一团体，该团体应是将来正式测验中准备应用的群体中的一个代表性样本。通过预测，可以得到预测团体在每道题目上的反应结果，对这些结果作出的分析就是项目分析，目的在于筛选题目，保证题目的质量。

项目分析一般从两个方面进行：题目的难度和区分度。

题目的难度

测验题目的难度一般以全部被试团体中正确回答的人数比例来表示，比如难度为1，表明题目非常容易，所有的人均回答正确；难度为0，表明题目极难，没有人能够正确回答；而难度为0.5，说明题目难度中等，有一半的人通过。一般来说，一道题目的难度是否适宜，取决于测验的目的、内容和形式。如果测验目的在于考察被试者对某些方面的知识技能的掌握程度，比如教育测验，那么只要是教育中认为重要的内容就可以编入测验，对难度并无特别要求。如果测验的目的在于选拔和安置人员，那么显然题目过难或过易均不合适，一般情况下难度值接近0.5是比较理想的，但这时的0.5是指整个测验的平均难度，而并不要求每道题目的难度均为0.5。由于人的心理特质基本上是呈正态分布的，即处于中间水平的人多，而高水平或低水平的人少，因此，为了对人群做最大程度的区分，题目的难度最好也是呈正态分布，即难题和易题相对少，中等难度的题目多。如果题目形式是选择题，那么题目难度一般应高于随机答对的概率值。比如四选一的选择题，随机猜对的概率是0.25，那么如果实际得到的

难度值居然不到 0.25，则说明题目有问题或者题目太难，这样的题目一般不保留。

题目的区分度

区分度是指测验项目对所测量的心理特性的区分程度或鉴别能力。确定题目区分度的指标和估计方法，一般以被试对题目的反应与某种效标测量间的关系为基础。所谓效标，是指与测验所测特质有关的可以直接而独立地测量的行为。比如，智力测验常以年龄作为效标，察看通过某一项目的人数是否随年龄而增长，若是，则该项目具有一定的区分度，否则，项目没有区分度，应考虑删除它；又如，成就测验可以用教师评定的等级为效标，若通过某一题目的人数随着学生的不同评定等级而发生相应的变化，那么该题目便有区分度，否则不予保留。当测验的外在效标难以得到时，通常以测验总分来代替，这种作法的基本假定是：测验作为一个总体是对欲测之心理特质的一个适当的测量，因为个别题目易受随机误差的影响，但当题目足够多时，全部题目上的误差作用方向不同，互相抵消，所以可假定全部题目组成的测验考察结果是有效的。这看起来是一个无力的假设，但经验证明，这个方法在大多数情况下是可行的。

由于一般情况下题目的区分度是以题目得分和效标分数之间的相关系数来表示的，因此，计算得到的区分度会在 $-1 \sim +1$ 之间变化。当区分度为 $+1$ 时，说明题目具有相当理想的区分度，或者说被试在测验题目上的反应与在效标上的反应完全一致，在题目反应上高水平的被试在效标上也一定是高水平的。若区分度为 0 ，说明题目对于水平高低的被试不具有一点儿区分度。而当区分度小于零时，则表明被试在题目上的反应与在效标上的反应一致性很低，出现在题目反应上高水平的被试在效标上水平反而低，或者在题目反应上低水平的被试，在效标上反而成为高水平的这样的反常现象，这时意味着题目出现问题，应予以

检查，或者修改后保留，或者删除。正常情况下对于区分度的分析，一般也应结合测验目的考虑：就教育测验而言，若目的在于考察被试者对某部分教学内容的掌握水平，那么以内容的重要性为主要考虑因素，题目的区分度并不重要，只要不出现负值即可。而对于一般的测验来说，区分度达到 0.20 以上基本上就属于可接受的范围。

在项目分析过程中应注意，题目的难度和区分度均是针对某一被试团体而言的，比方说同一道题目对于高水平的被试的题目难度小些，而对于低水平的被试则显然难度要大一些；难度越大的题目对于高水平的被试区分度好一些，而难度越小的题目则对低水平的被试者的区分度好一些。因此，做项目分析时必须保证施测的被试团体对将来正式施测团体的代表性。此外，项目分析过程对每道题目的难度和区分度不是孤立看待的，而是二者结合起来考虑题目的删或留的问题。例如一个难度适中的题目，若其区分度很低或出现负值，则应考虑删除；而若一道区分度可以接受的题目，其难度过低或过高，那么也应考虑修改或删除。总之具体情况具体分析。

经过项目分析之后，对测验中的题目或删除，或修改，或保留，根据正式测验所需的题目，教师选出最终的题目，组成正式测验。

信度评估

对于一个测验的质量评估，仅仅做项目分析是远远不够的。考察测验的质量好坏，一个必须考虑的因素是测验的信度。

所谓信度，是指测验结果的可靠性和稳定性，或者说是同一测验对同一被试者前后几次施测的结果之间的一致性程度。事实上，无论是物理测量还是心理测量，若对同一人或物施行若干次测量，其测得的每一次结果都不可能完全一致。这是由于测量误差的影响。正如物理测量中的概念一样，心理测量的误差亦可分

为两种：一为随机误差，是指由偶然因素引起的无规律的误差，导致测量结果围绕某一个值产生不一致、不稳定的变化；二是系统误差，是指由某种恒定无关因素引起的有规律性变化的误差，导致测量结果偏离真值，但每次偏离的方向和大小是稳定的，不会影响测验结果的一致性。既然信度是考察测验结果的一致性，那么显然只有随机误差对信度会有影响：随机误差越大，测验结果一致性越低，于是测验的信度便越低。

对于测验的信度的评估，理论上认为平行测验模型是一条理想的途径。该模型的主导思想是：将同一测验以两种形式表示，形式 A 与形式 B 须等值，即针对同一属性而测量，且其结果对于任一相当大的被试样本都能得到同样的分数分布。若不存在随机误差，那么同一个人在 A 和 B 上的得分应该是一致的。而若两个得分不一致，且差异很大，说明随机误差对测验分数的影响是相当大的，也就表明测验结果是不可靠的，换言之，测验信度很低。但若两个得分间差异很小，则说明尽管随机误差对测量结果产生了影响，这种影响却并不显著，也就是说，测验基本上还是可靠的，或者说，测验还是具有一定的信度的。

这种平行测验模型只能作为估计信度的一种理论途径，因为我们很难编制出同一测验的完全等值的两种不同形式。因此，在实际中我们一般采用几种近似的方法评估信度：

再测信度

这是一种最古老而又直接的信度评估方法，它直接评估了对同一组被试两次施测同一测验所得结果的一致性程度，且以两次结果间的相关系数作为信度指标。

应该注意的是，两次施测之间的时间间距应该较好地把握，间隔太短，由于第一次测验产生的练习效应和记忆遗留效应会影响第二次的测验结果；而间隔太长，个体被测的特质可能已经由于学习、教育，以及发育的影响而发生了变化。

复本信度

此法是接近平行测验模型：编制与原测验完全等值的平行测验，即复本，然后将测验及其复本施测于同一组被试，计算两组得分之间的相关系数，即为复本信度。

复本信度避免了再测信度的时间间隔的难题，但编制平行测验谈何容易，因而只能做到使二者尽可能接近等值。

分半信度

将测验施测于一组被试，然后将测验分半，计算被试者在每一半上的分数的一致性程度，即为测验分半信度的初步指标。此法既不需专门建构复本，又不必两次施测，因而简单易行且能克服许多不良影响，一般来说是最常用并且行之有效的信度评估方法。

分半法的核心是将测验分为两半，使之相当于两个平行测验，因而原则上要求两半测验等值。常用的分半方法有完全随机分半或奇偶分半（奇数题为一半，偶数题为另一半）。

评分者信度

这种信度针对开放式题目的评分一致性而设。开放式题目一般没有唯一正确的答案，评分时必然会受到评分者的主观判断的影响，那么不同的评分者之间必然会有误差存在。为考察评分者之间的一致性程度，一般采取让多个评分者评多份试卷，然后计算这多个评分者之间的一致性程度，作为评分者信度指标。

由于信度的评估指标一般以相关系数来表示，因此信度值一般在0~1之间波动。在所有其他条件都等同的情况下，信度系数越高，测验质量越好。不过，不同类型的测验可接受的信度水平不同。当测验用于做最终决策时，或需要依据测验结果的细致的个体差异将被试分类时，对测验的信度要求较高；而当测验用于初期筛选，或需依据测验结果的显著个体差异对被试进行粗线条的划分时，对测验的信度水平要求较低。这样，我们就可根据

测验目的和测验可行性来确定测验标准。一般来说，大多数学绩测验所报告的信度在 0.92 以上，智力测验在 0.90 以上，客观人格测验在 0.70 左右，创造力测验几乎不可能超过 0.50，投射测验最低，只有 0.20 或更小。

效度评估

保证了测验的信度，并不意味着保证了测验的质量。好比打靶：射手每发子弹均打中五环的某一点上，弹孔密集，说明他的稳定性很高，只可惜他没打中靶心，说明他没有准确性，我们因而断定他还不算是一个优秀的射手。同样的道理，当我们选用某测验作为评价或决策的依据时，不仅要把握该测验的稳定性，还必须保证它的准确性或有效性，即测验的另一个更为重要的指标：效度。

效度代表测验的准确性，即测验能够测量出其所欲测量的被试的心理特质的程度。比如我们意欲评定被试的几何能力，但选用的测验却全由代数题目组成，那么显然该测验结果并不能真实反映被试的几何能力，于是我们就说该测验缺乏效度。这也说明效度是一种相对的概念，它是相对于特定的测量目的而言的，上例的测验虽对几何能力的测量缺乏效度，但对代数能力的测量却很可能是有效的。因此，对于测验效度的评估，应结合测验目的而进行。

在测量误差中，随机误差既影响分数的一致性，又影响分数的准确性，而系统误差只影响分数的准确性而不会影响分数的稳定性。因此，测验的信度只受到随机误差的影响，而效度则同时受到随机误差和系统误差的影响。当随机误差减小时，测验信度一定会增加，但此时由于系统误差的影响的缘故，不能保证测验效度一定会增加，可见测验信度高只为效度高提供了必要但非充分的条件；而若测验效度高，则说明测验的误差在减小，测验的信度必然也会提高。因此，对于一个测验的质量评估，重点在于