

北京市高等教育自学考试委员会组编



心理测验

——原理与实践

彭凯平 编著

陈仲庚 审校



华夏出版社

心理测验

——原理与实践

北京市高等教育自学考试委员会组编

彭凯平 编著

陈仲庚 审校

华夏出版社

1989年·北京

北京市高等教育自学考试委员会组编

心理测验

——原理与实践

彭凯平 编著

陈仲庚 审校

*

华夏出版社出版发行

(北京东直门外香河园柳芳南里)

新华书店总经销

北京通县燕山印刷厂印刷

*

850×1168毫米 32开本 14.125印张 360千字

1989年5月第1版 1989年5月北京第1次印刷

印数1—5600册

ISBN 7—80053—375—1/B·022

定价:6.60元

序 言

心理测验一直被看作是心理学专业的一门基础课程，这是因为心理测验是研究行为和心理现象的方法之一。任何介绍方法的课程，都不外两种讲法：一种是从方法论的角度介绍原则、应用评价等理论；一种是从具体技术着手，介绍各种方法及其操作和注意事项。心理测验发展到今天已经形成一套有确切概念和完整结构的学科系统，理论上已有很大发展，具体测验种类和数目也大为增多。一本书如果试图介绍具体测验，很难避免挂一漏万的缺点；如果缺乏对具体测验的解释，也会陷入心理测量学的理论争论之中。本书试将两种讲法结合起来，即从理论着手，辅以有代表性的典型测验。本书还对心理测验的研究历史做了回顾，可从中了解测验方法的演变过程；对于编制测验的要求、标准化的过程及检验心理测验的多种方法等，本书尤其做了较为详细的介绍。

虽然心理测验的历史较长，但它仍是发展中的学科，而且其发展和应用颇有戏剧性。在国外三、四十年代蓬勃兴盛甚至狂热之后，现已逐渐平稳发展；在国内消声匿迹几十年后又起死回生，并且极为时兴了。现在有人探讨一个问题就想编一个测验，岂知编制一个合格的测验并非一件容易的事。我希望心理测验的书多出几本，使大家对测验的性质有更多了解。使得专业人员在编制和修订测验时有更严格的要求；有关的工作人员在应用测验时，更能充分掌握它们的特点、功能和限度，以便对结果作出恰如其分的解释。这样我国心理测验的研究和应用就能正常顺利地发展。

本书是彭凯平同志在讲义的基础上编写而成的。从讲义到教

材，作者耕耘细作，力图做到深入浅出，适合于初学者之用。但是读者的情况各有所异，书中某些内容不见得都能读懂。因此提醒读者，在学习心理测验之前，应学习普通心理学和心理统计学。

本书内容充实，材料丰富，层次分明，文风严谨、活泼。在送交出版社之前，书稿修改了多次，并由我审校。我非常愿意向读者推荐这本书。

陈仲庚

1988年元月于北京大学心理学系

0 N 11/35

目 录

第一章 绪论	(1)
第一节 心理测验的历史.....	(1)
第二节 心理测验的性质.....	(19)
第二章 测验的编制	(33)
第一节 测验目标分析.....	(33)
第二节 测题的编写技术.....	(37)
第三节 测验的编排和组织.....	(52)
第三章 测验的实施和计分	(58)
第一节 测验的实施.....	(58)
第二节 测验的计分.....	(77)
第四章 常模和测验分数的解释	(86)
第一节 参照常模的分数解释.....	(86)
第二节 参照效标的分数解释.....	(113)
第五章 信度	(118)
第一节 信度的性质.....	(118)
第二节 信度的类型及估计方法.....	(124)
第三节 影响信度的因素.....	(138)
第四节 信度的特殊问题.....	(142)
第六章 效度	(146)
第一节 效度的性质.....	(146)
第二节 效度的类型.....	(149)
第三节 效度的表示方法及影响因素.....	(168)
第四节 效度的应用.....	(179)
第七章 项目分析	(199)

第一节	项目难度	(199)
第二节	项目鉴别度	(205)
第三节	项目分析的特殊问题	(211)
第四节	项目反应理论	(214)
第八章	智力测验	(221)
第一节	个别实施的智力测验	(221)
第二节	团体实施的智力测验	(242)
第三节	智力测验中的心理学问题	(247)
第四节	智力的本质	(261)
第九章	成就测验	(277)
第一节	成就测验的性质和应用	(277)
第二节	常见标准化成就测验	(284)
第三节	教师自编成就测验	(296)
第十章	能力倾向测验	(303)
第一节	能力倾向测验的性质	(303)
第二节	特殊能力倾向测验	(307)
第三节	多重能力倾向成套测验	(320)
第四节	创造力测验	(327)
第十一章	兴趣、态度和价值观测量	(334)
第一节	兴趣问卷	(334)
第二节	态度量表	(349)
第三节	价值观及有关变量的测量量表	(359)
第十二章	人格测验	(365)
第一节	人格测量的理论基础	(365)
第二节	人格问卷	(374)
第三节	投射技术	(396)
第四节	人格评估的其它方法	(408)
参考文献		(429)
后记		(443)

第一章 绪 论

“一棵树上没有两片完全相同的叶子”，世界上很难找到两个完全相同的人。当我们放眼周围世界的时候，就会发现，人与人之间的差异是非常明显的。有人身材魁梧，有人个头矮小；有人聪智伶俐，有人愚蠢笨拙，有人天生丽质，有人相貌平平，……在身体和外貌上的差别显而易见，在心理过程和心理特征上的差别也是如此。这种人与人之间的不同，我们称之为个别差异。人们在谈论个别差异的时候，总得使用描述这些差异的名词或术语，如身高、长相、智力等等，同时还要赋予这些名词以一定的差异数量，它们有些是准确的数值，如一米五，90分等，有些则是不准确的描述，但实际隐含着比较和差别意义。

在日常生活中使用不准确的描述是可以接受的，但对于心理科学研究来说，这种主观色彩浓厚、不易实证的描述远远不符合现代科学的要求。我们不能只说这人聪明，那人愚蠢；这人孤僻，那人开朗，我们必须回答这些行为特征反映什么样的心理品质，它的性质如何，如何定义，它的相关变量以及如何定量描述和研究。解答这些问题，就需要借助心理测量和评估的知识。本书的目的就是从方法论的角度，全面介绍心理测量的原理、原则与实施方法。定量的方法有很多，我们将以其中应用最广、影响最大、效果相对较好的心理测验为主。

第一节 心理测验的历史

温故而知新。简单回顾心理测验的产生和发展，对于了解测

量的原理和心理测验的发展趋势，吸取历史的经验教训，颇为有益。有史以来，先哲们就以各种方式论述过人们在能力、个性、行为表现等许多方面的差异，中外皆然。可以说，心理测验虽只是世界上最年轻的学科之一——心理学中较年轻的分支，但其思想和实践源远流长。我国始于汉代、兴于隋唐的科举取士制度就被中外学者公认为最早的心理测量的实践。在古希腊，测验作为教育的附加物，即测智力技能又测运动才干。在中世纪，欧洲的某些大学已开始使用考试方法。但科学的心理测验则是在工业革命成功后的19世纪的欧洲发展起来的。

一、中国古代心理测量思想

我国古代具有非常丰富的心理测量思想，而且具有非常鲜明的特色，这个丰富的宝藏需要我们去挖掘。

关于对心理现象可测性的认识，孟子就曾有见解：“权，然后知轻重；度，然后知长短。物皆然，心为甚”，这明确指出心理能力和心理特征与物体一样具有可测量的特性。孔子曾根据自己的观察将人分为中人，中人以上和中人以下三个类别，并说“中人以上，可以语上也；中人以下，不可以语上也”，“唯上智与下愚不移”，这与现代测量理论中的命名与分类量表的思想相一致。

三国时期（公元前3世纪）刘劭著有《人物志》一书。由于该书对人物的研究颇有独到之处，1937年，美国的施罗克（J. K. Shryock）曾将它翻译成《人类能力研究》在美国发表。刘劭将人分类为圣贤、豪杰、傲荡、拘慤，即“心小志大者圣贤之伦也；心大志大者豪杰之雋也；心大志小者傲荡之类也；心小志小者拘慤之人也。”

关于心理测量方法的认识方面，刘劭在《人物志》中提出的是观察的方法，他认为要“观其感变，以审常度”，意思是根据一个人的行为变化便可推测他的一般的心理特点，他还提出通过词的回答（“应赞”）来观察人的智力。

公元6世纪初叶，南朝人刘勰在《新论·专学》中提到“使左手画方右手画圆，无一俱成”，“由心不两用，则手不并运也”，这恐怕应算是最早的“分心测验”了。

6世纪中叶以来我国江南的风俗“抓周”(周岁试儿)，也有点类似目前的婴儿发展测验。本世纪20年代美国心理学家葛塞尔(A. Gesell)才用类似的方法在实验室记录了幼儿的动作和适应行为等方面的发展状况。

我国的“七巧板”在某些方面可作为创造力测验的一种方法。利用大小形状不同的七个小板块，能够组成上百种生物或实物的复杂图形。现在的智力测验和创造性测验中也在使用类似的方法。

我国的科举制度在心理测量历史上地位突出。汉代有荐官制度，隋唐至清末盛行科举制，人才选拔悉依考试结果为准则。科举制传入欧洲后，很受西方新兴资产阶级的欢迎，并用于他们的官吏考试制度中。

我国古代的心理测验思想具有明显的东方特点：首先，它们都是描述性的，不定量的，其次，我国古代往往将心理特点(能力、人格、智力等)与道德观念联系起来。我们不难发现我国的测验思想是分类思想，即把个体分成几种不同的类别，而每个类别在人们的评价中是有层次高低、好坏善恶等感情色彩的。例如后汉班固(公元32—92年)的“九品量表”分古今人物为上上(圣人，如周公、武王)、上中(仁人，如伯夷、叔齐)、上下(智人，如仓颉)、中上，中中，中下，下上，下中，下下(愚人，如蚩尤、妲己、赵高)。这种分类将智力与道德、心理能力与社会评价等混杂在一起，带有明显的主观色彩。

二、心理测验的早期探索

心理测量的产生源于对个别差异的兴趣，从科学的观点看，这种兴趣在中世纪的欧洲是几乎不存在的。当时的社会结构中，一个人的行为活动被他所出身的社会阶层所支配，没有多少个人

的表现和发展的自由，直到16世纪，新兴的资产阶级逐渐占领了历史舞台，促使个人主义、艺术和科学的复兴，这就不仅使人们对个人的天赋及发展产生了兴趣，而且它所带来的政治和经济的繁荣使得这种个别差异的研究更加实际，更有可能。这就是心理测量产生的社会思想背景。

1. 测验产生的实际需要

测验产生的最初原因是对智力落后和精神病人治疗的需要。我们知道，在19世纪以前，智力缺陷者和精神病人被认为是恶魔缠身，采用嘲弄、禁闭、甚至拷打的方式对待他们。后来欧洲和美洲对低能者的态度发生了变化，建立了许多收容所，而且已经意识到必须确定一种鉴别各种心理疾病的统一标准以及管理和教育的方法。

首先需要区别的是智力落后和精神病。法国医生艾斯克罗（J.E.D.Esquiril）是第一个作出这种区别的人。他认为，精神病的显著标志是情绪障碍，有可能伴随或可能不伴随智力上的损害，而智力落后的主要特征是从出生或从婴儿期就表现出来的智力缺陷。他还指出，从接近正常到最严重的白痴，智力落后有许多等级。在试图发现这种区分不同等级的智力落后的系统中，艾斯克罗观察到个体使用语言的能力最依赖于他们的智力水平。有意思的是，这和现在使用的区别智力落后的标准大致相仿，而且现今的许多心理测验也包含许多言语测验内容，言语能力在智力概念中也占着很大的比重。

另一位值得一提的先驱是法国医生沈干（E.Seguin），他的老师意太（Itard）是第一个尝试将狼孩培养成正常儿童的人，但他没有成功。沈干承其师业，开始训练智力落后儿童。1837年他建立了第一个专门培养智力落后儿童的学校。1846年发表《白痴：用生理学方法进行诊断和治疗》一书。1848年，他移居美国，他的思想得到广泛的接受。沈干采用的感觉训练方法和肌肉训练方法至今仍在全世界范围内的智力落后儿童教育机构中应

用，其中的感觉辨别和运动控制方法已成为现今的操作测验或非言语智力测验的组成部分。

在艾斯克罗和沈干的工作开创半个世纪以后，法国心理学家比内（Alfred Binet）提出，那些不能适应正常学校学习的儿童可在退学前被鉴别出来，并将他们分别安排在特殊班级里进行学习。比内和他的儿童心理学研究会（Society for the Psychological Study of the Child）的同事促请法国公共教育部改善智力落后儿童的学习条件。结果，一个专门研究公立学校中智力落后儿童管理办法的委员会成立了，比内也是其中一员，这项任命产生了心理测验历史上的重大事件——第一个智力测验量表的诞生。

2. 实验心理学的影响

实验心理学诞生于德国，代表人物为费西纳、冯特、艾宾浩斯，他们证明了心理现象可以定量表达和实证研究，并创造了进行定量研究的心理学概念体系。不过，19世纪的实验心理学家一般说来并不关心个别差异的研究，他们的主要目的是试图发现人类行为的一般规律，个别差异或者被忽略或者被认为是误差的来源。对相同的刺激，不同被试的反应不同，这种差异被称为个体变量。对个别差异的这种态度在冯特所创建的实验室中也同样存在。他们所从事的工作大部分是有关视觉、听觉及其它感觉和简单反应时的实验，这种强调感觉的现象对早期的心理测验是很有影响的。

必须指出，实验心理学对心理测验的发展有积极的推动作用。例如，实验心理学家非常强调严格控制实验条件，指示语不同就会加快或减慢被试的反应，颜色的亮度也会影响到被试所感受的刺激等。这种使所有被试在尽可能标准化的条件下实验的思想被心理测验学家所沿用，并且测量的标准化程度已经成为今天心理测验的重要指标。

3. 高尔顿的贡献

第一个直接推动测验运动的学者是英国生物学家高尔顿(S-

ir Francis Galton)。作为达尔文的表兄弟，他深受进化论思想的影响，提出了能力遗传说，并主张改善种族，实行优生。他在调查遗传问题的过程中，认识到需要测量在血缘上有关和无关的人们的特征，以确定其相似的程度。基于这一观点，他促使许多教育机构对他们的学生进行系统的人类学记录。1884年，在伦敦国际博览会上，他还专门设立了一个“人类测量实验室”，参观者付三个便士，即可得到身高、体重、阔度等身体素质和视听觉的敏锐度、肌肉力量和反应时等感觉运动功能的测验结果。此后，实验室迁至伦敦南肯辛顿博物院，持续6年之久，积累了9,337人的材料，这可以算是第一个大量系统地测量个别差异的尝试。

高尔顿在他的人类学实验室中发明了许多测验的仪器，有些至今仍十分有效，如长度视觉辨别所用的高尔顿棒，听力实验所用的高尔顿笛，等等。高尔顿相信洛克的“一切知识来自感官”的思想，指出白痴在冷、热、痛的感觉辨别上也是有缺陷的，因而认为感觉辨别能力“基本上是心智能力中最高能力。”

高尔顿也是评定量表和问卷法应用的先驱，他也经常应用自由联想技术。另一个最重要的贡献是他发展了分析个别差异材料的统计方法，他所选择和修改的数学分析技术，使一些没有受过数学训练的人也能使其测验结果定量化。这些工作后来由他的学生加以完善，其中最突出的就是皮尔逊(Karl Pearson)，他找到了相关概念的计算方法，形成了矩阵相关的理论。

4. 卡特尔和早期心理测验

在心理测验史上另一位著名代表人物是美国心理学家卡特尔(James M. Cattell)。他将实验心理学和年轻的测验运动结合起来。卡特尔早年从师冯特，获得莱比锡大学博士学位。尽管冯特不太情愿，卡特尔还是完成了他的关于反应时的个别差异的论文。在剑桥逗留期间，卡特尔与高尔顿过从甚密，他的测量个别差异的兴趣得到了强化。返美之后，卡特尔积极致力于心理实验室的建立和测验运动的传播。

1890年,卡特尔发表《心理测验与测量》于《心理》杂志上,这是心理测验(mental test)第一次出现于心理学文献中。卡特尔写道,“心理学若不立足于实验和测量上,决不能够有自然科学的准确性”。同时他又指出,“心理测验如果有一个普遍的标准,则其科学和实用价值将增加不少”。该文主要报告了他在宾夕法尼亚大学实验室对大学生实施一套测验以确定他们的智力水平的结果。测验系个别实施,包括肌肉力量、运动速度、痛感受性、视听敏感度、重量辨别、反应时、记忆力以及类似项目。由此可见,卡特尔与高尔顿的观点基本相同,认为智力功能的测量可以通过感觉辨别力和反应时测验来获得。卡特尔对测验的这种看法在当时是有道理的,因为在那个时候,简单的功能已经能够准确和可靠地测量,但是复杂功能的客观测量几乎是不可能的事。1894年,卡特尔以他的测验施测哥伦比亚大学学生,结果发现其测验成绩与学业成绩相关很低。

类似卡特尔这样的测验在19世纪的后十年得到发展,开始应用于学校儿童、大学生及各种各样的人群。1893年在芝加哥的哥伦比亚展览会上,贾斯特罗(J. Jastrow)展出一套15个测验,被试可以进行感觉、运动和简单知觉过程测验并与一定的常模比较。不过,试图评价这些测验的研究报告所述结果却不令人满意。这些测验的结果彼此相关不高,与教师评定的学生智力水平无关或相关不高,与学业成绩相关也不高。

由当时欧洲学者所编制的类似测验却倾向于包括较复杂的功能。克雷丕林(E. Kraepelin)对精神病患者的临床检查有兴趣。1895年,他制订了一个测验来测量他所认为的个体特性的基本因素。该测验主要是算术测验,用来测量应用效果、记忆和对疲劳与分心的易感性。在此之前,克雷丕林的学生厄恩(A. Oehrn)曾编制一套测验,包括知觉、记忆、联想和运动机能的测量。另一位德国心理学家艾宾浩斯则用算术运算、记忆广度和语句填充测验施测学校儿童,其中只有最复杂的语句填充测验与儿

童的学业成绩十分吻合。

如同克雷丕林，意大利心理学家费雷尔（G.C.Ferrair）和他的学生主要对病理学研究中的测验应用感兴趣，1896年，他们发表了从生理测量到理解广度和图片解释的一系列测验。

在1895年发表的文章中，比内和亨利（V.Henri）批评这些测验在感觉方面做得太多，而且过份集中于简单、特殊的能力。他们进一步阐述道，对更复杂的功能的测量要求准确程度很高是不必要的，因为在这些功能上的个体差异非常大。他们提出了一些较为严格的测验，包括记忆、想象、注意、理解、暗示性、审美等。从这里我们可看到著名的比内智力量表的萌芽。

三、心理测验的发展及应用

任何科学领域，在其拓荒时代总有许多人在不同国家中成为该学科的先驱者，而其中影响最大的一个或几个人就会被认为是该学科的开创者。在心理测验领域，高尔顿、卡特尔和比内就是这样的人物，其中比内被认为是心理测验的鼻祖。

1. 比内和第一个智力测验

比内1857年生于法国尼斯，青年时从医，继而对心理学产生兴趣。从此，他和他的同事们就一直致力于智力测量的研究。他们尝试了许多方法，甚至包括头盖骨、面貌和书写笔迹的测量。研究的结果逐渐使他们意识到：只有直接的，即使是粗糙的方法，才最有希望有效地测量复杂的智力功能。1903年，他发表《智力的实验研究》一书，提出了广义的智力概念。他认为，智力包括一切高级的心理过程，表现在推理、判断以及运用旧知识解决新问题的能力上。他以自己的两个女儿为被试，尝试使用一些测验，如词语填充，图片解释等。1904年，一个特殊的机会使得比内的思想可能付诸实践。当时法国巴黎公共教育部委托许多医学家、教育家和其它科学家组成一个委员会，专门研究公立学校中智力落后儿童的教育方法。作为委员之一，比内主张用测验的

方法来发现和鉴别智力落后儿童，但遭到许多反对。比内并不气馁，他仍与西蒙（Th. Simon）合作，完成了世界上第一个智力测验量表——比内-西蒙量表（Binet-Simon Scale）。1905年他在《心理学年报》上发表《诊断异常儿童的新方法》一文介绍此量表，因而历史上称为1905年量表。

1905年量表共有30个问题，题目由易到难排列。难易程度的确定系由对50名3岁~11岁的正常儿童和部分智力落后儿童和成人的测试结果分析而得。该量表的设计包括广泛的功能，但主要强调的是判断、理解和推理能力，亦即他所认为的智力的基本组成成份。因此，尽管也有一些感知觉的测验，但言语部分比当时的其它测验所占比例要大。不过当时的比内-西蒙量表还很粗糙，尚不能准确地表示总分。

1908年，比内发表修订后的比内-西蒙智力测验量表，删掉了1905年量表中不合适的题目，增加了一些新题，使总数达到59个。所有测题都按年龄分组，组别从3岁到13岁。年龄水平根据300名正常儿童的测试结果来确定，3岁组的题目是那些能被80%—90%的3岁儿童所能接受的测验，余组类推。此外，测验成绩用“智力水平”（Mental Level）表示，其通常的翻译和提法是智力年龄（mental age），表示儿童能完成何种年龄水平的儿童能完成的测验。值得一提的是，比内一直避免使用智力年龄一词，因为他觉得智力年龄所隐含的意义是无法证实的。他倾向于中性的智力水平一词。

比内-西蒙量表的第三次修订本于比内不幸逝世的1911年发表。这次修订没有重大变化，只是改变了几种年龄水平分组，并扩展到成人组。比内英年早逝（终年54岁），确是心理测验运动的一大损失。

早在1908年量表发表之日起，比内-西蒙量表就吸引了全世界心理学家的注意，各种文字的翻译本和修订本相继出现，其中最为著名的当属1916年出版的由美国斯坦福大学的推孟（Lewis

Terman)教授修订的斯坦福-比内量表(Stanford-Binet Intelligence Scale)。该量表首次采用了智力商数(intelligence quotient)概念,它表示的是智力年龄和实际年龄的比值,简称智商(IQ)。斯坦福-比内量表在世界上广为流行,我国也有修订本,至今仍在使用。另一个有特殊用途的比内-西蒙量表修订本是库尔门(F. Kuhlmann)1912年修订的库尔门-比内量表,他将适用范围下延到3个月年龄水平,这可以算是世界上最早的学前儿童和婴儿智力测验。

2. 团体测验

比内式的智力测验都是个别测验(Individual test),每次只能施测一个人。这种测验要求被试作口头报告或熟练地运用测验材料,因此其时间是固定的。另外,主试还必须经过严格的训练。这种测验在临床个案研究中很有用,但不适合于大规模团体施测。

团体测验也是由于实际需要产生的。1917年美国决定参加第一次世界大战。美国心理学会任命以伊尔克斯(R. M. Yerkes)为首的委员会研究为战争服务的方式。他们认为,选拔和分派官兵的任务必须考虑到他们的一般智力水平,不过军队有一百多万人,实践这种想法,只能采用大规模的团体施测方法。军事心理学家使用了许多有用的材料,特别是欧提斯(Arthur Otis)的团体智力测验,这是欧提斯当推孟的研究生时设计出来的,其中多项选择题和其它客观题形式为首次采用。测验最后编制修订成著名的军队 α 和军队 β 测验,即陆军甲种测验和陆军乙种测验,前者为文字测验,后者为非文字测验,是为文盲和外籍不懂英文的新兵设计的,二者都适合于大量团体的测验。在1917年3月至1919年1月间,共有200多万官兵进行了测验,积累了大量的材料。

战后不久,陆军甲种和乙种测验被迅速应用于美国社会。这不仅因为它适合于大量群体,也由于它的实施方法简便易行,主试不须经过太多的训练就可胜任。在20年代,智力测验运动出现