

戴忠恒 编著

高等学校文科教材

心理与教育测量



华东师范大学出版社



● 教育科学丛书



2 032 9595 0

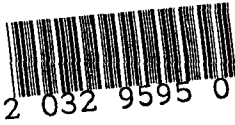
高等学校文科教材

心理与教育测量

戴忠恒 编著



华东师范大学出版社



心理与教育测量

戴忠恒 编著

华东师范大学出版社出版

(上海市中山北路 3663 号)

新华书店上海发行所发行

吴县光福印刷厂印刷

开本: 850×1168 1/32

印张: 10.25

字数: 246千字

1987年3月第一版

1987年3月第一次印刷

印数: 1—15,000本

统一书号: 7135·179

定价: 1.80元

6-12-13

前 言

心理与教育测量是本世纪初兴起的一门学科。它以心理和教育统计学为基础，论述各类心理与教育测验的一般编制原理及实施、记分、解释的标准化的方法。各类心理与教育测验在鉴别智愚、选拔人才、指导就业、临床诊断和预测、评价学生学业成绩以及实验设计等方面，均能发挥其工具之效能。时至今日，随着该学科的发展，各类心理与教育测验已不仅在心理与教育诸领域中显示出它的理论和应用的价值，而且在诸如小儿科、精神病科、社会学、法学和企业管理等其它应用领域，也愈益受到人们的重视并获得广泛的应用。

本教材的目标在于引导读者掌握编制心理、教育测验的原理与技术，并能正确地使用各类标准化测验，为日后从事心理与教育的科研以及各种临床诊断打好必要的基础。

这一教材是笔者根据四年来在华东师大讲授心理测量和从事心理、教育测验等工作的经验而写成的。1982年应教学之急需，曾打印过一份讲义，虽经我师左任侠教授审阅过初稿，但由于笔者的水平限制，未能恰当地将左先生的许多精辟见解付与笔端，致使在章节的编排和内容的取舍等方面，不尽令人满意。此次付印，对原讲稿作了相当的变动和增删。

今天拙作能得以由讲义而衍成此书，又承蒙左先生抽暇审阅原稿并为本书的出版撰写序言，在此谨表示至诚的谢意。

由于笔者的专业和文字水平的限制，错误疏漏之处在所难免，企望前辈、同行、学友，不吝赐教。

戴忠恒

一九八五年六月于华东师范大学

序

人们观察天体,可以借助望远镜;探究细胞等微细结构,可得显微镜之便。然而在进行人的个别差异的研究时,为了测定个体的能力或人格特征,目前行之有效的方法还只能以心理测验作为其工具。

我国自隋唐至清末的科举考试制度,可谓首开心理测验的先河,至于使心理测验达于科学的境地则始于法国的比纳。1905年因他与西蒙首创智力测验而著称于世。自此,心理测验迅速在世界范围得到广泛应用和发展。在我国自20年代起,陈鹤琴、廖世承、艾伟等心理测验工作者,先后出版了有关心理与教育测验的专著。时隔半个多世纪以后的今天,测验技术已有长足发展,不过多少有点令人遗憾,国内迄今尚无较完整地反映当今测验新进展的著作问世。这势必影响到我国心理与教育测验事业的发展。目前,在我国测验运动刚刚复苏之际,有关引导从事这方面专业工作的人掌握编制心理、教育测验的原理和技术以及指导他们正确使用各类标准化测验的著作诚属必需。戴忠恒君积累多年从事教学及测验编制、修订的经验写成这本《心理与教育测量》,并由华东师范大学出版社出版,在填补心理测验的文献的空白上作出一定的贡献,这是令我高兴的事。

当然戴君之书,未必尽善尽美,我们希望今后能有更多不同体系、各具特色的专著问世,以繁荣我国心理与教育测验事业。

对从事心理测验的人来说,掌握测验原理和技术自属重要,但遵守其职业道德更具特殊的重要性。非如此,测验不能发挥其对社会应有的效能。美国心理学会曾通过《心理学家的道德

标准》条例。作者把它译出,附于书末。它或许对我们能起暮鼓晨钟的作用。

左任侠
一九八五年五月

目 录

第一章	心理测量的发展史	(1)
第一节	心理测量的早期尝试与先驱者的探索	(2)
第二节	比纳—西蒙智力量表的产生及其特点	(5)
第三节	心理测量的蓬勃发展及应用	(13)
第二章	测验的性质	(22)
第一节	测量	(22)
第二节	测验的定义	(32)
第三节	正确使用测验,防止滥用	(38)
第四节	测验的实施与被试的主观状态	(39)
第五节	测验的种类和功用	(42)
第三章	智力测验	(46)
第一节	智力的性质	(46)
第二节	个别智力测验的举样	(66)
第三节	团体智力测验的举样	(78)
第四章	教育测验	(86)
第一节	综合成就测验	(86)
第二节	分科成就测验	(94)
第三节	品质表的编制	(100)
第四节	教师自编测验	(104)
第五节	教育测验在教育评价中的作用	(108)
第五章	人格测验	(115)
第一节	自陈量表	(116)
第二节	投射技术	(127)
第三节	人格评定量表和情境测验	(137)

第六章	态度测量法	(143)
第一节	态度的性质及态度量表的功用	(143)
第二节	等距量表法	(146)
第三节	李克特量表法	(152)
第四节	哥特曼量表法	(155)
第七章	编制测验的方法	(160)
第一节	测验目的的确立和材料的选择	(160)
第二节	测题编制的技术	(166)
第三节	预试和测题的分析	(178)
第四节	测题的选择、编排以及测验的标准化	(182)
第五节	测验的基本特征的鉴定	(189)
第八章	信度	(192)
第一节	信度的理论	(193)
第二节	误差的来源	(198)
第三节	估计信度的方法	(202)
第四节	影响信度系数的因素	(210)
第五节	测量的标准误差	(213)
第九章	效度	(219)
第一节	效度的含义与类型	(219)
第二节	准则关联效度	(223)
第三节	内容效度与结构效度	(230)
第四节	效度的应用	(239)
第十章	项目分析	(247)
第一节	项目的难度	(247)
第二节	项目的鉴别力	(256)
第三节	项目的相关	(259)
第四节	一个项目分析的实例	(264)
第十一章	测验量表与常模	(267)
第一节	标准化的样组	(267)

第二节	原始分数与导出分数	(270)
第三节	测验量表	(272)
第四节	关于某些常模表和侧面图的显示	(294)
附 录	心理学家的道德标准	(298)
参 考 文 献	(310)
附 表	正态曲线的面积和纵轴	(311)

第一章 心理测量的发展史

心理测量的思想和实施在中国由来已久。我国的古代考试制度是现代心理测量的渊源，任何《心理测量年表》都以此为开端。

二千多年前，孔子曾说过：“惟上智与下愚不移。”“中人以上，可以语上也；中人以下，不可语上也。”当时虽记载不详，但可推知孔子是根据某种法则（规则）把人类的智力分为上、中、下三个类别，而后把属于某个类别的个体一一枚举其中。这完全符合现代人的测量的见解（见第二章史蒂芬斯的测量定义）。

孟子向齐宣王说：“权，然后知轻重，度，然后知长短，物皆然，心为甚，王请度之。”由此可知，在孟子看来心与物二者都具有一种可以测量的特性，这种思想较我们所熟知的名言“凡物之存在必有其数量”（桑戴克，1918），“凡有数量的东西都可以测量”（麦柯尔，1922）早二千多年之久。

汉代至魏晋，在人材选求上盛行推荐制。隋唐至清末 1300 多年科举制通行，对人才的选拔悉依考试结果为准则。自从利玛窦 1582 年来我国之后，我国的文物制度逐渐传入欧洲。科举制度一经传入欧洲，备受启蒙思想家伏尔泰的赞扬。

于是，夺取政权后的法国资产阶级于 1791 年参照我国科举制建立了自己的文官考试制度。随后，英美两国也相继建立。它们为受官任职制定了准绳，打破了门第或名人推荐的限制，扩大了资产阶级进入政府的通道。

古代波兰大学法学考试在 1219 年开始运用口试。1599 年耶稣会颁布了笔试条例，但只是在中学实行。1787 年普鲁士毕业考试是论文式的作业考试。这是十八世纪西欧的考试状况。

到十九世纪，欧美先进的资本主义国家逐步建立了学校考试制度，并把考试同升学、就业、免费、奖学金等联系起来。然而到该世纪的后期，这些国家由于小学普及，普通中学与职业中学也有普及的趋势，从而面临着选择最合适的人接受更多教育的问题；另一方面又由于工业生产的发展，分工日益精细，也就有一个选择最合适的人去接受最合适的职业训练，以便为资产阶级训练出创造最大利润的熟练工人的问题。解决上述两个方面的问题，显然是文官考试制度和学校考试制度所不能胜任的。因而，按照人的智力高低作为选拔人才的标准就提到了心理学家的议事日程上。而实验心理学的诞生和个别差异的研究成果为心理测验奠定了基础。

第一节 心理测量的早期尝试 与先驱者的探索

一、冯特的实验心理学

实验心理学的鼻祖冯特(Wundt)，1879年在德国莱比锡大学设立第一所心理实验室，用生理学和物理学的仪器及其方法来作心理实验，从此心理学开始成为一门实验科学。冯特当时的兴趣是想发现人类行为的一般规律。后来发现各被试观察力不同，词的联想不同，反应能力也不同，因此，就发生了个别差异的研究。要研究个别差异，必须有测量个别差异的工具，所以个别差异的研究引起了心理测验运动。

二、高尔顿的理想

英国优生学家高尔顿(Galton)在1869年出版了《遗传与天才》一书。书中认为智力由遗传而来，而且人类智力的分布是常态的，一半的人归属于平常，四分之一的人是在平常人之上，四分之一的人是在平常人之下。他很想着手编制一种测量人类智力的量表，不过他始终未能着手具体编制。十四年后他的《人类

才能及其发展的研究》一书出版。在这本著作中，他首先提出“测验”、“心理测量”这两个术语。高尔顿在 1884 年创设“人类学测量实验室”于国际卫生展览会内，后迁至伦敦南肯辛顿博物院內，只要付三个便士就能受测，并把测量结果记入卡片。该实验室记载的资料有：身高、体重、阔度、呼吸力、拉力和压力，手击效率，听力、视力、色觉及个人其他资料。在该实验室开办的六年内共测量了 9337 人。高尔顿有从这些生理的、感觉的和知觉的材料里求出人类许多品质的设想，特别是推出智力的设想，这已为心理测量学的发展史所否定。然而，根据这些资料他提出的相关概念（后来由他的学生皮尔逊完成相关计算公式），却是心理测量的重要工具，为日后心理测量的研究工作铺平了道路。

三、卡特尔的个别差异研究

对个别差异研究最有贡献的要推美国人卡特尔 (Cattell J. McK)。他是冯特的学生，又跟高尔顿在一起工作过。返美后，继承冯特和高尔顿二人的学说，以各种实验来研究个别差异。1890 年，他在心理杂志上发表了一篇重要文章《心理测验与测量》。文章说：“心理学者不立根基于实验与测量上，决不能有自然科学的准确。”又说：“如果我们规定一个一律的手续使在异时、异地得出的结果可以比较、综合，则测验的科学和实用的价值都可以增加。”这些是编制测验的一些重要观念。不过他所发表的测验都属于感觉能力和动作过程，脱胎于冯特的实验。

现将他在《心理测验与测量》一文中所讲的第一套测验介绍如下：

1. 握力测量。用握力计测量肌肉的力量。
2. 动作速度测量。被试的右手和臂从静止状态，在一平面上尽量以最快速度移动五十公分。
3. 触觉两点阈测量。
4. 引起痛觉的最低点的测量。用一硬质的橡皮带加压到被试的前额，视被试忍受痛苦的最低点，结果可以从指示器上获

知。

5. 辨别重量最小差别的能力的测量。
6. 对声音反应时的测量。
7. 说出四种混杂在一起的颜色名称速度的测量。
8. 把一根 50 公分的线平分之二,测量被试精确性。
9. 对 10 秒钟时间判断的测量。
10. 复述听过一次的字母数目的测量。

在卡特尔看来,这些测题都与人们的高级心理活动有关,或与人们的神经活动有关。譬如,压力测量是测量人们的意志的控制和情绪激动的;辨别重量最小差别的能力是测量人们耐心的;再如引起痛觉最低点的测量,可以用作神经系统的诊断。

1894 年,卡特尔首先用各种测验法测量哥伦比亚大学学生,使测验走出实验室。无论是他自己做的测验得分,还是他的学生威斯拉(Wissler)做的测验得分,都与学业成绩相关极低。其实从现在的观点看来,这是必然的结果。学业的优劣与智力的高低有密切关系,而智力的高低却不是感觉能力,动作过程所能测量的。卡特尔的这个结果启发了后来的研究者应从别的途径探索测量智力差异的方法。

上述三位学者只是早期心理测验先驱者的代表人物,其他尚有不少人物对心理测验早期尝试产生不小的影响。在这里要特别提及如下几位:

E. 克雷匹林(Kraepelin), 1894 年建议在精神病理学上使用测验,并用心理测验的方法研究心智正常的人和病态的人。他的学生厄恩(Oehrn)编了一套测验,将能力分成四种:知觉、记忆、联想和运动机能。

艾宾浩斯(Ebbinghaus), 认为智力的高下在于综合能力的差异。他于 1896 年首创填充法来测量学生的智力,这大大有别于卡特尔的低级心理过程的测验。填充法是现在常用的测题格式,很简单,可是,在当时却是一个创举。

K. 皮尔逊(Pearson),在 1896 年对高尔顿的相关的概念予以数学上的突破,找出了计算的公式。尽管法国数学家 A·布拉维在 1846 年已求出基本定理,但对高尔顿的问题解决得那么好,则应归功于皮尔逊的积矩相关的理论。

1904 年查尔斯·E·斯皮尔曼(Spearman)发表了他著名的论文《普通智力》,此文奠定了他的智力结构二因素论的理论基础,为日后比纳智力测验的编制提供了基础。

在法国,比纳(A·Binet)正在多方面探索测量智力的方法。譬如,1894 年至 1898 年,他与人合作发表了七篇论文,其中有一篇是关于测验的。在该文中他们建议测验应测量记忆、表象、想象、注意、理解、暗示、美的欣赏、道德情感、意志力和运动技能。智力测验的创始人,也同那些先驱者一样在探索!

第二节 比纳—西蒙智力量表的产生及其特点

比纳是发明智力测验常模量表的第一人。他于 1857 年生于法国,起初学习法律,继而改习医学,后来致力于实验心理学的研究。他和费雷合作发表了一系列有关《动物吸引力》的实验报告,开辟了一条使催眠能够为实验心理学者所利用的道路。他和费雷合著的《推理心理学》发表后,又有一连串关于特殊思维的实验研究作为续编。例如,对有关心理有缺陷的人、下棋的人以及快速计算的人思维活动的研究。1903 年他的另一本著作《智力的实验研究》问世,在此书中他所讲的智力是广义的,包含一切高级心理过程,他以自己的两个女儿做被试,所用的测验有填充、图片解释等。由于这些工作,他认为智力是极其复杂的能力,非简单的方法所能测量,重要的是直接面对广泛而复杂的课题,而不可通过狭窄而简单的方法去研究广泛而复杂的课题。就是本着这样的精神,他设想着智力测量的问题,这是他超越前

人的地方。

但是比纳的成功是经过许多尝试,走了许多歧路而得到的。在1900年,比纳花了很多时间去测量儿童的头盖骨,目的是比较聪明与愚笨儿童头盖骨的大小。结果表明聪明儿童大体上稍占优势,但差别很小,交掩很大,在个体智力的评定上完全无用,不久比纳放弃了这条探索道路。

一、比纳—西蒙智力量表

1904年,法国教育部委派许多教育家、医学家组成一个委员会,研究公立学校内低能儿童班级的管理问题。比纳也是委员之一,他主张用一种测验的方法去辨别心理缺陷的儿童。经过他与西蒙(Simon)的精心研究,次年在《心理年报》上发表了一篇论文,题为《诊断异常儿童智力的新方法》。我们现在称之为1905年量表。1908年和1911年,他对此又加以修正,称为1908年量表和1911年量表。现分别介绍如下:

(一) 1905年量表

这个量表共包括30个测验项目,具有如下两大特点:(1)项目种类繁多,可以测量智力的多方面的表现。只有对一个人的行为反应进行多方面的观察,才能有效地和可靠地评判他的智力高低。(2)测验项目的排列,由浅而深,使同一量表可以测量智力高低不同的儿童。虽然这个量表没有用年龄作为测量单位,但读此论文就可以知道这时年龄量表观念已略具雏形。在这篇文章里,比纳已指明三岁、五岁、七岁、九岁、十一岁儿童在他的量表里通过什么项目。譬如,能力平常的三岁儿童,大约可从第一测题做到第九测题。以下是该量表的大概情形:

1. 看:用一支点燃的火柴在儿童面前慢慢移动,看儿童眼睛是否能跟随物体移动。

2. 握物:把一块木头与儿童的手掌或手背接触,看这个儿童能否拿着它不致掉下,并且放在嘴里。这是测量儿童是否能通过触觉的刺激唤起抓握的反应。

3. 用视觉的刺激唤起抓握的反应：把一块木头放在儿童面前，看他能否去拿它。

4. 食物之辨认：给予一块糖和一块木头（糖须用巧克力）。

5. 搜寻食物：用纸在儿童面前包好一块糖，看他能否将纸包打开把糖吃掉。

6. 执行简单的命令和模仿简单的手势：叫小孩关门或坐下，视其是否照办。

（根据其测量结果重度低能，即白痴只能做到上面测题）

7. 物件之认识：指出身体上各部位之名称和说出经常见到或听到的物品。

8. 图片之认识：给儿童看一图片，然后询问他图片上某物在何处。

9. 列举图片中的人物和物体：儿童须在图片中列举常见的物品和人物。

10. 比较两条线的长短：假如三次都能指出何者长，何者短，就算通过。

11. 重述两个数目字。

12. 比较两个重量：一个三克重，另一个十二克重。

13 抵抗暗示能力：问以看不见的物件，如一幅图画上没有茶杯，但问茶杯在哪里？或用无意义的字询问图片中的物体；比较线的长短，头三根线是不一样长的，后三根线是一样长的。

14. 解说物品和人物的名称：匙、房屋、狗、妈妈。

（中等的低能只能做到这儿为止，五岁正常儿童也只能做到这儿为止）

15. 重述句子：每个句子有15个字母，先说一次给儿童听，然后要其背诵。

16. 说出二物的分别：苍蝇与蝴蝶，布与纸等。

17. 记忆图片中的物件：给儿童一张图片，内有十三个普通物件，令其看了三十秒钟后，再背出，背出愈多愈好。

18. 根据记忆重画图片：给儿童看过几种图形，而后要他画出一种。

19. 重述数目：能背出六位数算通过。

20. 说出二物的相同点。

21. 比较线的长短：先让儿童看三公分长的线条，然后要他从另一图中找出相同长度的线段。

22. 比较重量：把五个重量不同(18、15、12、9和6克)的物品依次排列。

23. 重量记忆：做对第22题以后，拿走一个物品。问被试哪个物品被拿走了？

(轻度的低能只能做到这里为止)

24. 说出同韵的字。

25. 填充：每一个句子缺少一个字，要被试填入使其成为有意义的句子。

26. 用三个字造句。

27. 问题的理解：从易到难共有15句。例如，你在睡觉时做什么？假如回答“作梦”就对了。

28. 交换长短针的位置：答出交换之后是什么时刻。

29. 剪纸：将正方形纸折成许多折，然后剪去其中一个角，问打开之后有几个洞，而且画出洞的位置。

30. 抽象名词之意义：例如革命与进化、总统与皇帝有何不同？

(二) 1908年量表

1905年量表发表后，经德可乐里(Decroly)和德冈(Degond)在比利时应用，证实它的信度是高的，但是，德可乐里也指出，这个测验的内容大都取自资产阶级生活环境，因而资产阶级的子女的测验成绩往往高于贫苦人家子女。比纳和西蒙根据使用的结果对这个量表进行了修订。1908年他又和西蒙发表一篇极其重要的著作，题为《儿童智力之发展》，在此文中他俩