

31

B 841.7
A 9

心理学丛书——实用心理系列

Tests & Examinations

Measuring Abilities and Performance

心理测验与考试

——能力和行为表现的测量

【美】 Lewis R. Aiken 著

张厚粲 译

 中国轻工业出版社

图书在版编目 (CIP) 数据

心理测验与考试：能力和行为表现的测量 / (美) 艾肯 (Aiken, L. R.) 著；张厚燊译. —北京：中国轻工业出版社，2002.4

(心理学丛书·实用心理系列)

ISBN 7-5019-3648-X

I. 心… II. ①艾… ②张… III. 心理测验—基本知识 IV. B841.7

中国版本图书馆CIP数据核字(2002)第009974号

版权声明

Copyright © 1998 by John Wiley & Sons, Inc.

All rights reserved

Authorized translation from the English language edition published by John Wiley & Sons, Inc.

策 划：石 铁

责任编辑：朱 玲 张乃东 责任终审：杜文勇

版式设计：刘智颖 责任监印：吴维斌

*

出版人：赵济清（北京东长安街6号，邮编：100740）

网 址：<http://www.chlip.com.cn>

电子信箱：wqtw@263.net

电 话：(010) 65262933

印 刷：北京天竺颖华印刷厂

经 销：各地新华书店

版 次：2002年4月第1版 2002年4月第1次印刷

开 本：850 × 1168 1/32 印张：9.5

字 数：150千字

书 号：ISBN 7-5019-3648-X/G · 314

定 价：16.00元

著作权合同登记 图字：01-2001-3252

· 如发现图书残缺请直接与我社发行部联系调换 ·

版权所有 · 翻印必究

译者序

在我国，现代心理学研究迄今已走过八十余年的曲折历程，中国心理学的国际地位不断提高，新世纪的第一个国际心理学大会将于2004年在北京举行。近年来，心理学不仅逐渐成为大学的一个热门专业，并鉴于心理学的重要意义和已取得的成绩，我国已将其列为一级学科，还被指定为全国大专院校的公共必修课。与此同时，社会各界也逐渐对心理学发生了浓厚的兴趣，特别是医学、司法、体育、学校、企业和政府部门，迫切希望了解心理与教育测量的相关理论及其在儿童教育、课堂教学实践，以及在诊断、司法鉴定、人事选拔、人员安置、员工培训等实践中的应用。

和许多科学理论与科学发明均起源于古代中国一样，虽然中国也是心理与教育测量技术的最早故乡，如能力测量、人格测量、学籍测量等现代心理与教育测量的主要领域在我国漫长的古代社会里均有所涉及，并取得过举世公认的成就。但由于众多因素特别是受中国人文传统的影响，现代心理与教育测量的理论和技术不是产生于中国，而是产生于工业革命后的一些西方国家。自20世纪初叶以来，西方心理与教育测量获得迅速发展。现代心理与教育测量在我国的发展则道路坎坷，自清朝末年传入中国，建国前持续了30多年的研究热潮，经过最初一代心理学家和工作者的

共同努力，在测验的引进、编制，以及测量学科的学科建设方面取得了非凡的成果，但自1949年以后的又一个30年间，由于深受前苏联心理学的影响，我国的心理与教育测量一度成为禁区，无人问津。直到1978年，随着心理科学在中国现代化进程中的地位得到重新肯定，心理与教育测量工作才得以恢复。迄今，我国的心理与教育测量在各主要领域取得了迅速发展，为心理与教育测量的学科建设和改进实际工作做出了一定的贡献，但同时也应该清醒地看到，我们与西方国家的差距。

在当前社会关注心理学、期待心理学的大好形势之下，为创建符合中国现实国情并服务于新世纪中国发展所需的心理与教育测量学，了解和借鉴西方心理与教育测量理论和研究成果更显重要与迫切。

在这种情况下，中国轻工业出版社精心选择了由西方新近出版、著名心理学家艾肯 (Lewis Aiken) 撰写的两本心理与教育测量方面的书籍来介绍给国内的广大读者，实在是再好不过的事，必将为心理学的应用推波助澜。我的研究生沈湘秦、赵海燕等积极参与了本书的初译工作，在此感谢他们。

本书(《心理测验与考试——能力和行为表现的测量》)是其中的一本，全书行文简练而内容全面、心理认知领域的心理与教育评估，涉及测验的历史背景，心理测量学的理论与统计方法、测验的编制施测，还依次介绍了西方比较流行的智力测验、能力测验和心理测验等。此外，还附有一份心理学词汇表，是初学者的良师，也是专业人士的益友。

正如本书的作者艾肯博士所说的那样，这本书不仅适合作为心理学专业的学生、研究者的辅助教材与绝好的参考书，还是各

领域对心理测量感兴趣的读者从中获得关于心理测量的一般信息的最佳读本，即适合那些关心员工发展的人力资源经理或行政官员、关心学生认知能力及其发展倾向的教师、关心来访者认知能力及其发展趋势的咨询人员与专家阅读。

人文关怀是21世纪的主题，从某种意义上讲，心理学的繁荣与发展是实现人文关怀的必由之路。我寄望于有志于心理学的各界人士，特别是青年学生和专业工作者们，珍惜当前的宝贵时机，广泛汲取西方的理论与研究经验，共创我国心理学的辉煌！

译者

2002年1月

序

本书简要而又全面地介绍了认知领域的心理与教育评估，主要是为行为科学的研究者和应用人员编写的。这些人使用客观测验、答题测验、口头测验和操作测验作为对员工与学生进行选拔、安置、提升、资格确认，以及对其行为与认知所施加的教育、培训、干预的效果进行评价等方面的实践和理论研究。

无论是构建或施测自己测验的专家，还是出于理论或应用的目的（比如在公司中进行员工选拔、员工安置、绩效评估、项目评估，或是在学校中进行学生的甄别与安置、课程评估，以及进行组织规划等）而使用标准化测验的专职人员，都会发现本书资料丰富的特点。本书面向的读者包括各类组织中关心员工、关心学生或来访者的认知能力及其发展的人力资源总监、教师、行政官员和专家们。

实践证明，本书不仅可作为商业、心理学、教育、社会学、医学等方面的研究和测量课程的有益的补充教材，而且还是心理与教育研讨班的很好的辅助教材。相关专业的研究生在设计其在社会学方面、行为学方面或其它应用领域的论文、专题论述时，他们也会从本书的字里行间发现有关测量人类行为与成就方面的有用的、指导性的建议。

本书(《心理测验与考试——能力和行为表现的测量》)是这套心理测量丛书的第三本书,前两本即:《等级量表与检核表》(*Rating Scales and Checklists*, 本书未经翻译出版)和《问卷与调查表》[*Questionnaires and Inventories*, (Aiken, 1997b) 此书也已由中国轻工业出版社组织翻译出版]也同样涉及到了对人的能力与技能的评价。但前两本书并不专门解决认知测验或成就测验、特殊领域的特殊能力测验等问题,其重点主要是情感特性,如人格特质、态度和观念这类问题,而感觉运动能力与技能只是点到为止。而在这本有关认知测验的书中,重点恰是对本套丛书提及的心理测验与评估做的全面论述。

在结构上,本书和前两本书保持了一致,即:正文部分同样分为八个章节,外设附录(心理学词汇表)。(英文原版附带有一盘与书中内容配套的计算机软盘,该软盘含有几十个关于测验的结构、施测、计分、标准化和评价的程序,还有一些评定操作、项目和产品的其它方法)在内容上,本书也和前两本书的风格类似,全书不仅涉及了概念性或理论性的知识,还包括各种测验在广阔领域中的实际应用知识。书中详细地介绍了书面、口头和操作性的成就测验,以及一般认知能力(智力)测验和特殊能力测验。此外,在全书的结尾部分提供了全书的词汇表。这些辅助资料不仅有助于读者理解各章的概念、原理和细节问题,也有助于读者在理解书中固有内容的基础之上进一步提高。

艾肯 (Lewis Aiken)

目 录

第一章 历史背景与专业问题	1
历史背景	2
测验分类和信息汇总	15
心理测验的标准、伦理与法律问题	19
小结	31
第二章 心理测量学的概念与方法	33
测量与统计	33
经典测量理论与方法学	43
项目反应理论	60
小结	65
第三章 测验编制、施测与计分	69
测验的编制	69
测验的施测	82
测验计分与解释	89
小结	97

第四章 标准化成就测验	101
基础知识	102
成套成就测验	108
单学科测验	112
商业和职业背景测验	121
小结	126
第五章 口头与操作成就测验	129
口头成就测验	131
操作测验和档案夹	135
小结	148
第六章 单项与多项能力倾向测验	151
单项能力倾向测验	156
多项能力倾向测验	167
能力倾向测验的效度	185
小结	190
第七章 一般心理能力测验	193
定义、设计与应用	197
个体智力测验	201
韦氏成人智力量表、韦氏儿童智力量表和 韦氏幼儿智力量表	205
韦氏成人智力量表第三版 (WAIS- III)	209

团体智力测验	221
小结	229
第八章 认知发展与障碍的评定	233
婴幼儿个体发展的测量	234
躯体残障者的测量	240
神经心理测量	246
小结	258
词汇表	261

第一章 历史背景与专业问题

测验，这个词在字典里至少有十几种解释，这里所指的测验，含义并没有那么复杂，它的意思很明确，就是指心理测验。那么何谓心理测验呢？心理测验由一组题目、问题或者能够引发个体特定反应的任务组成，这些反应与我们所要测量的个体心理特质、能力或成就有关。用来测量人格特质的测验通常称之为人格测验、人格量表或人格问卷。在本系列丛书的另一书——《心理问卷与调查表》中我们详细阐述了这几种测量工具。有时，这种类型的测验工具也被视为情感性测验、非智力性测验或非认知性测验。这类测量工具测量动机、兴趣、态度、个人气质与其它个性特征或特质。认知能力测验则测量智力活动的过程与结果，比如记忆、理解、推理和问题解决。这一类型的测验多半为纸笔测验，当然，也有一部分为口头测验或者操作性测验。认知能力测验可以进一步划分为成就测验和能力倾向测验。前者考察个体在某一学科或其它学习领域当前的知识水平与技能水平，后者探究个体习得某一特定的知识或技能的潜在能力。成就测验着眼于过去与现在——个体已经学到的或取得的知识技能；而能力倾向测验关注将来——在适当的教育和训练条件下个体能够达到什么水平。能力这个词内涵很广，包含了智力、成就与能力倾向，因此，所有这些认知测量通常

都可以称为能力测验，其中智力测验测量一般心理能力，能力倾向测验测量特殊能力。

对大多数人而言，测验并不是个令人愉快的词。只要在某个领域，个体的表现存在着一个接受标准，每年有成千上万的人不得不在测验的海洋里浴血奋战，努力展示自己在某个具体学科、某一个既定职业或任何其它相关领域内的知识和技能。人们总是把测验和考试看成自己前进的障碍，认为测验妨碍了自己在班级、组织、职业或任何其它特定团体中的自我实现。当然，并非所有的测验都是用来进行决策，如决定是否把个体纳入某种教育、训练或者职业。很多时候我们用认知能力与技能测验来对特定领域、工作或者职位的个体进行分类和安置，以使得这些个体可以找到自己的最佳位置。还有一些时候，我们用测验来确定个体是否具备某种资格、是否可以给其颁发某种执照，或者个体当前的知识技能水平是否可以令人接受。还有一些测验的目的是诊断，相应领域的工作者借助这些测验确定某一个体或群体是否存在特定问题、是否具有特殊的需要，以便制定相应的干预或者治疗计划。

历史背景

在人类历史上，很早就出现了通过对各种任务的操作来考察其能力的“测验”。比照现在的标准，这些“测验”大多数不具备足够的科学性，但是它们确实具有一定的效力。否则，这些测验就不会存在那么长的时间了。任何人只要粗略地瞄一眼一群人如何完成某个任务，就可以马上发现，做一件事情可

以好到什么程度，又可以差到什么程度，也可以知道个体表现上的差异有多大。由此出发，把个体的表现从最差到最优进行排序，并且借此设计一套系统评估个体的表现也不是什么特别高深的事情。我们可以通过控制行为情境的特定方面，如测验材料、时间、地点、在场的其他人等等来提高评估过程的一致性与准确性。这样，一个根据个体在特定任务上的表现对个体进行评估的正规测验很快就出现了。

最早的非正式选拔测验——“苹果测验”(the Apple Test)记载于圣经的创世纪(Genesis)一节中，这个测验只有两个参与者(亚当和夏娃)，但失败者的下场是很悲惨的。圣经里还有另一个例子记载在Judges7: 3-7，这个例子里，君主对吉迪恩说：

你这里的人太多了，我无法把米甸(人名)送到这么多人的手里。为了以色列不向我夸耀他用自己的力量拯救了自己，现在你要向大家宣布：“那些怕得浑身战抖的人都回去吧，请真正的勇士留下吧。”结果有22000人离开了，还剩下10000人。

但是君主对吉迪恩说：“人还是太多了，现在把这些人赶到水里。我会到那里替你进行一下筛选的。如果我说，这个人可以跟着你，他就去；如果我说，这个人不能跟着你，他就不去。”

于是吉迪恩把人们领到了水里。在那里君主告诉他：“把那些像狗一样舔水的人和跪下去喝水的人分开”。300个人用手捧起水然后用舌头舔着喝水，其他人则跪下去喝水。

君主对吉迪恩说：“这300个舔水的人交给你，我会把神子们交给你们；其他人就别理他们了，让他们哪来哪去吧。”

在这个小故事里，君主知道那些在喝水的时候像狗一样用舌头舔的人更加警觉或机警，几百个机警的人要胜过一大群粗心大意的乌合之众。

比这略微正式的心理测量可以追溯到4000多年前。公元前2200年，中国皇帝设立了一种文官考试制度，这种制度通过口头考试选拔政府官员，每三年举办一次。大约2000年后，汉朝才开始对其官员实行笔试考试，内容包括公法、军事、农业、税收、地理和文章、诗歌等。中国古代文官考试制度的最后一次完善发生在公元1370年，这种考试十分重视对儒家经典的学习与钻研，其最终形式共分为三个阶段，第一个阶段是初试，通过初试的学子才可以参加下一阶段的地区考试，通过地区考试的学子才可以参加第三阶段的全国考试。参加初试的学子中只有不到1%的人可以顺利通过三个水平的考试。英国、法国和德国都在19世纪仿效中国建立了自己的文官考试制度，但中国在20世纪初期废除了科举制度。

欧洲国家在引入正式考试上要比中国晚得多。其中最早的是1219年Bologna大学采用的法律考试。这些早期的大学考试大多数是口头测试，正式的笔试直到16世纪才出现。表1.1列出了心理和教育测量早期历史上的几个重要的里程碑。尤其值得一提的是1845年在波士顿中学最早采用的印刷测验，和1864年制定的论述题的客观评价标准，以及几年之后先驱者所做的一般心理能力或智力测量的研究。

表 1.1 心理和教育测验史大事记

公元前 2200	中国建立文官考试制度
公元 1219	欧洲的大学里出现正式考试
1540	大学里出现书面考试
1599	杰休茨出版发行了书面考试的规则
1833	在英国出现竞争性文官考试制度
1845	在霍拉斯·蔓的指导下波士顿学校董事会最先采纳了印刷的 笔试考试
1864	乔治·费舍，一位英国校长，设计了一套包括例題和相应答 案的量表，作为评估学生论文的标准
1866	O.E. 赛格温发表了第一本论述智力障碍者的评估与治疗的学 术专著
1869	高尔顿按照天赋对人进行分类，开创了对个体差异进行科学 研究的先河
1870	竞争性文官考试被引入美国
1884	高尔顿在伦敦的国际健康博览会设立了人体测量实验室
1888	J.M. 卡特发表了关于美国儿童的拼写能力的研究结果
1893	约瑟夫·贾斯特罗在芝加哥的哥伦比亚博览会上展示了自己的 感觉运动测验。
1897	J.M. 赖斯发表了关于美国儿童拼写能力的研究发现。
1904	斯皮尔曼提出了心理能力的二因素论。
1905	比内-西蒙智力量表的第一个版本出版发行。
1908	比内-西蒙智力量表的修订版本出版发行；J. C. 斯通和S. A. 考特斯的客观性计算测验问世。
1908-1914	桑代克编制了一套计算、字迹、语言和拼写的标准化测验。
1915	编制出首个多项选择测验，茨萨斯阅读测验开始应用，斯坦 奎斯特机械能力测验出版发行。
1916	推孟出版了斯坦福-比内智力测验。奥提斯在斯坦福-比内 智力测验修订版的基础上编制了第一个团体智力测验。
1917	最早的团体智力测验：陆军甲种测验和陆军乙种测验被编制

- 出来，用来甄选新兵。
- 1919 西肖尔的音乐能力倾向测验出版发行。
- 1923 斯坦福成就测验第一版出版发行。
- 1924 衣阿华学能测验出版发行。
- 1926 学能测验 (SAT) 首次被使用。
- 1927 格塞尔发展量表出版发行。
- 1932 梅特波利帝安成就测验出版发行。
- 1937 斯坦福-比内智力测验修订版出版发行。
- 1938 O. K. 伯罗斯出版了第一期心理测量年鉴；基本心理能力测验、瑞文推理测验、本德尔完形视觉运动测验出版发行。
- 1939 韦-拜智力量表发表。
- 1941 空军学员分类成套测验被编制，并用在陆军航空兵飞行学员身上。
- 1947 能力倾向区分性测验出版。
- 1949 韦氏儿童智力量表出版发行。
- 1968 韦氏学前儿童智力量表出版发行。
- 1970 至今 计算机在测验的编制、施测、计分、分析和解释方面发挥越来越重要的作用。
- 1971 (美国) 联邦法院要求用于人事选拔的测验必须与工作有关 (Griggs v, Duke Power)。
- 1974 韦氏儿童智力量表修订版和霍尔斯特德-里顿神经心理成套测验出版发行。
- 1980 至今 研究者提出项目反应理论。
- 1983 考夫曼儿童评估系统 (K-ABC) 出版发行。
- 1985 教育和心理测验标准出版发行。
- 1986 斯坦福-比内测验第四版出版发行。
- 1995 心理测量年鉴第十二期出版。
- 1997 韦氏成人智力量表第三版出版发行。
-

智力的测量

人们通常把高尔顿 (Galton) 在 19 世纪后期对智力的研究与著作视为现代心理测量的开端。高尔顿使得当时的人们相信,智力是一种遗传的能力,可以通过一系列简单的感觉运动来进行测量,他相信这些感觉运动测验能够考察个体的各种智力成分。在评价高尔顿所做的智力测验研究之前,有一点必须注意,那就是,在高尔顿所处的时代,心理测量与实验研究正在迅猛地发展。然而,当时的科学水平还非常原始,无法从哲学上深入解释大脑的功能,也无法科学地阐明大脑和神经系统如何发挥作用。当时对“心理”的流行观点认为心理源自对感觉印象的精致化,所以,心理活动中包含的内容必须首先出现在感觉系统里。

这样看来,高尔顿选取一系列的感觉运动测验来测量个体的智力,尤其是天才的智力,这样的做法也就并不奇怪。这些测验考察的内容包括感觉阈限、视觉和听觉反应时间、运动速度、肌肉力量、视觉和听觉的敏度、痛觉敏感性、重量阈限等。此外,高尔顿还获得了大量的身体特征数据,如身高、体重、头长、头宽、胳膊跨度、中指的长度、下臂的长度、肺活量等。根据卡特尔 (McKean Cattell) 的说法,魏斯勒 (Clark Wissler) 的一个经典研究发现高尔顿的这些测验和大学分数相关非常低,甚至测验之间彼此的相关也很低。无论如何,再用当时在心理学实验室里发明的那种陈旧的测验方法来测量智力是不可能了。

尽管高尔顿或卡特尔在测验上的失败冷却了人们编制单纯