

美国心理协会年度杰出贡献奖得主、美国国家科学院医学院
院士卡普兰教授的代表作

斯坦福大学、密歇根大学、马里兰大学、约翰霍普金斯大学
等400多所学校采用

PSYCHOLOGICAL TESTING

心理测验

原理、应用和问题

(英文版·原书第7版)

[美] 罗伯特 M. 卡普兰 (Robert M. Kaplan) 加利福尼亚大学洛杉矶分校 著
丹尼斯 P. 萨库兹 (Dennis P. Saccuzzo) 圣迭戈州立大学

郑日昌◎导读



机械工业出版社
China Machine Press

心理学英文版教材

PSYCHOLOGICAL TESTING

Principles, Applications and Issues

心理测验

原理、应用和问题

(英文版·原书第 7 版)

[美] 罗伯特 M. 卡普兰 (Robert M. Kaplan) 加利福尼亚大学洛杉矶分校 著
丹尼斯 P. 萨库兹 (Dennis P. Saccuzzo) 圣迭戈州立大学

郑日昌◎导读



机械工业出版社
China Machine Press

本书是非常有特色并在国际上享有盛誉的心理学教材,令人耳目一新。与其他教材相比,熔理论和实践于一炉,铸科学性与实用性为一体是本书的主要特色。

从内容上看,全书概括了心理测量与测验领域的重要进展和发展动态,既有经典测量理论又有当代测量理论,既有历史悠久的传统测验又有刚流行的测验和老测验的最新版本。概念阐释准确,理论论述严谨,贴近社会现实和生活实际,具有较强的可操作性。从形式上看,本书结构清晰新颖,资料丰富翔实,书中穿插许多典型案例和讨论专栏,既趣味盎然又有助于读者更好地理解一些深奥抽象的理论问题。全书在保持内在统一性的同时,各章节具有相对的独立性,教师可根据教学目的灵活安排。

本书适用于心理学专业人士。

Robert M. Kaplan, Dennis P. Saccuzzo. Psychological Testing, 7th edition.

Copyright © 2010 Cengage Learning Asia Pte Ltd.

First China Reprint 2010.

This edition is reprinted by China Machine Press, with license from Cengage Learning Asia Pte Ltd, for sale in the People's Republic of China (excluding Hong Kong SAR, Macau SAR, and Taiwan).

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

本书原版由圣智学习出版公司出版,此影印版由圣智学习出版公司授权机械工业出版社独家出版发行。此版本仅限在中华人民共和国境内(不包括中国香港、澳门特别行政区及中国台湾)销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可,不得以任何方式复制或发行本书的任何部分。

本书封底贴有Cengage Learning防伪标签,无标签者不得销售。

封底无防伪标均为盗版

版权所有,侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号:图字:01-2010-1053

图书在版编目(CIP)数据

心理测验:原理、应用和问题(英文版·原书第7版)/(美)卡普兰(Kaplan, R. M.), (美)萨库兹(Saccuzzo, D. P.)著;郑日昌导读. —北京:机械工业出版社, 2010.7

(心理学英文版教材)

书名原文: Psychological Testing

ISBN 978-7-111-31169-0

I. 心… II. ①卡… ②萨… ③郑… III. 心理测验—教材—英文 IV. B841.7

中国版本图书馆CIP数据核字(2010)第123022号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:许可 白春玲

北京京北印刷有限公司印刷

2010年7月第1版第1次印刷

214mm × 275mm · 34印张

标准书号: ISBN 978-7-111-31169-0

定价: 70.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

客服热线: (010) 88379210; 88361066

购书热线: (010) 68326294; 88379649; 68995259

投稿热线: (010) 88379007

读者信箱: hzjg@hzbook.com

导 读

Review

心理测验作为一种测量工具既是从事心理学研究又是将心理学应用于实际的重要手段,心理测验本身更是广泛应用于各行各业。

由于历史的原因,心理测验在我国被取缔多年。自20世纪70年代末,随着心理学在科学的春天中复苏,一些大学开设了心理测量课程,几本相关教科书也陆续面市。随着时代的发展,心理学越来越受到社会的重视,各行各业对心理测验的需求也与日俱增。

80年代伊始,笔者在北京师范大学率先开讲心理测量课。为了满足教学需要,编写了我国新中国成立后第一本心理测量教材,并于1987年在湖南教育出版社出版,1999年、2005年、2007年和2008年笔者又与几位学生和同行合作,分别在人民教育出版社、高等教育出版社、华东师范大学出版社、中国人民大学出版社出版了另外几本内容各有侧重的心理测验教材。

1985~1986年,笔者在美国访学进修期间曾拜读美国著名心理学家罗伯特 M. 卡普兰和丹尼斯 P. 萨库兹合著的《心理测验:原理、应用和问题》一书的第1版(1982年),这是一本非常有特色并在国际上享有盛誉的心理学教材,当时便有耳目一新之感。与其他同类教材相比,熔理论和实践于一炉、铸科学性与实用性为一体是本书的主要特色。

1998年,笔者应台湾五南图书出版公司之邀,与郑默合作翻译了此书第3版,主要在香港、澳门、台湾等地区发行。2005年,赵国祥等人又翻译了该书的第5版,由陕西师范大学出版社在中国内地发行。

本书作者罗伯特 M. 卡普兰是加利福尼亚大学洛杉矶分校(UCLA)健康服务系主任和医学院教授,1997~2004年,他是加利福尼亚大学圣迭戈分校家庭和预防医学系的教授和主任,是美国心理协会健康心理学分会、美国科学促进会J分会、国际生活质量研究会、行为医学学会和行为医学研究会等组织的前任理事长,美国胸科学会行为科学委员会前主席。同时,卡普兰博士是《健康心理学》现任主编、《行为医学年鉴》的前任主编,还是《美国心理学家》和四个其他学术刊物的副主编或顾问编辑。他还两次荣获美国心理协会健康心理学分会杰出科学贡献年度奖(1987年初级奖,2001年高级奖)、杰出研究学者奖(1988)和健康网络公司(Health Net)杰出学者奖(1991)、加利福尼亚大学125周年校庆最杰出校友奖、美国心理协会杰出学者奖、美国医学院校心理学家协会颁发的突出科学贡献奖、2003年行为医学学会国家领导奖、2004年生活质量研究国际协会职业成就总统奖、2006年行为医学学会杰出研究导师奖。卡普兰博士已独立或与人合作出版著作15本,发表论文400余篇。在科学文献索引(ISI)中,他是世界上被引用最多的作者(99.5%以上)。2005年,他当选为美国国家科学院医学院院士。

丹尼斯 P. 萨库兹是美国圣迭戈州立大学的心理学教授和托马斯·杰斐逊法学院的兼职法学教授,是一位从事心理测验工作长达32年的学者和从业者,在该领域发表了大量的广受同行关注的文章并举行了许多专业讲座。萨库兹博士的研究得到了国家科学基金会、美国国立

精神卫生研究所、美国国家健康研究所、美国教育部、苏格兰典礼基金会和美国军队服务处的支持。他也是持有加州执照的心理学家和律师，通过了美国职业心理学委员会（ABPP）临床心理学的资格认证。此外，他还持有美国评估心理学委员会（ABAP）和美国法医委员会的证书。他以其在心理学领域的不同寻常的杰出贡献，成为美国心理学协会、美国心理学会和西方心理学协会的会士。萨库兹博士单独或与人合作发表300余篇广受同行关注的论文，出版了8本教科书和20本法律手册。他是一家将心理科学用于解决法律争议、进行法律研究的教育机构和法律心理公司的总裁与合伙人。

本书共21章，分为三部分。作者在系统阐述心理测量基本原理之后，分门别类地介绍了各种心理测验及其在不同领域的广泛应用，在此基础上深入浅出地讨论了心理测验在理论和应用方面存在的争议及相关法律和伦理问题，最后对心理测验的发展趋势做了科学展望。

从内容来看，全书概括了心理测量与测验领域的重要进展和发展动态，既有经典测量理论，又有当代测量理论；既有历史悠久的传统测验，又有刚流行的测验和传统测验的最新版本。概念阐释准确，理论论述严谨，贴近社会现实和生活实际，具有较强的可操作性。

从形式上看，该书结构清晰新颖，资料丰富翔实，书中穿插许多典型案例和讨论专栏，既趣味盎然，又有助于读者更好地理解一些深奥抽象的理论问题。全书在保持内在统一性的同时，各章节具有相对的独立性，教师可根据教学目的灵活安排。

本书是在海内外经过了众多知名大学反复使用、多次修订的较为完善的教材，将英文原版在我国出版可避免因翻译有误或不准确导致的以讹传讹、误人子弟。此书不但可供从事一般教学和研究工作的心理学者参考，更应成为心理测量课教师的案头必备；此书还可作为心理专业本科生的参考教材，只要达到大学英语四级水平，对照中文目录和术语表，读起来便不会有太大困难；对于心理学及相关专业的研究生，在以前学过的测量原理的基础上，可重点围绕本书案例和有争议的伦理与法律问题展开讨论，以加深理解、拓宽视野；而对于想去美国攻读心理学研究生学位的同学来说，只要认真读懂这本书，对顺利通过研究生入学考试（GRE）心理学专业试卷中令许多人头痛的测量学部分会大有裨益。

总之，学习本书不但能使你全面掌握心理测量原理和测验使用技能，而且能大大提高你的英语水平。

既然你已翻开此书，那就不妨读下去，相信你一定会获益匪浅！

郑日昌

2009年11月

于上海浦东

郑日昌

现任北京师范大学教授、博士生导师，中国浦东干部学院访问教师（中组部派遣）、中国社会心理学会常务理事、中国心理学会临床心理与心理咨询专业委员会副主任、中国性学会常务理事、教育部考试中心兼职研究员、教育部中小学心理健康教育专家指导委员会委员、教育部普通高等学校学生心理健康教育专家指导委员会委员、全国标准化技术委员会委员。主要研究领域为心理与教育测量、心理咨询与治疗、人力资源管理。代表作有《心理测量与测验》、《学校心理咨询》等。

全书各部分简介

Introduction

第一部分 原理

第一部分系统阐述心理测验的基本原理，即测验发展的理论基础，包括第1~7章。

第1章导论，着重介绍了心理与教育测验的基本概念、测验的类型和测验发展历史上的重大事件。

第2章测验常模与基本统计，介绍了心理量表的特征和种类、描述统计的基本概念以及测验常模的含义及编制过程。

第3章相关与回归，阐释了分数散点图和相关的含义、相关和回归的区别、运用回归直线进行分数预测的程序及多元分析的方法。

第4章信度，阐释了误差的概念和来源、真分数和信度的含义、估计信度的方法以及如何使用信度信息。

第5章效度，重点介绍了效度的定义、类型、估计方法及信度与效度的关系。

第6章测验项目的编写和评价，介绍了测验项目的主要形式、编写原则以及分析项目难度和区分度的方法。

第7章测验的实施，主要讨论测验主试与被试的关系以及期望效应怎样影响测验成绩，计算机辅助测验的优点及行为评估所要注意的问题。

第二部分 应用

第二部分翔实介绍在教育、商业、工业、军队和临床等领域广为运用的主要心理测验及评价技术，包括第8~18章。

第8章访谈技术，阐释了访谈的交互性质、访谈的类型、有效访谈的原则、访谈的误差来源及如何提高访谈技能。

第9章智力理论和比奈量表，介绍了比奈和斯皮尔曼等心理学家是如何界定智力的，以及比奈智力量表的演变。

第10章韦克斯勒智力量表，详尽介绍了韦克斯勒量表的起源及向儿童和幼儿的拓展，测验的结构、解释方法、心理测量学特征及评价。

第11章教育和特殊教育中的其他个体能力测验，主要介绍了用于婴幼儿、残疾人等特殊人群的一般个体能力测验。

第12章教育、公共服务及军队中所使用的标准化测验，介绍了团体能力测验的主要特征，从幼儿园到大学的主要团体能力测验及在商业和工业领域的广泛应用。

第13章临床和咨询中的应用，介绍了结构化人格测验的编制策略及运用每种策略所编制的典型测验。

第14章投射性人格测验,介绍了投射测验的理论假设,罗夏墨迹测验、主题统觉测验及其他投射法测验。

第15章计算机和基础心理科学测验,介绍了认知行为评价的原理和方法以及心理生理法,讨论了计算机在现代心理测验中的作用。

第16章咨询心理学中的测验,介绍了影响最大的几个职业兴趣量表,讨论了在工作安置中个性特征测量的有关问题。

第17章健康心理学和保健测验,重点介绍了临床神经心理学评定以及生活质量评定的常用方法。

第18章工业和商业场所的测验,重点介绍了测验在人事选拔中的应用及人职匹配、工作分析等有关理论问题。

第三部分 问题

第三部分介绍了心理测验领域中存在的主要争议问题,展望了测验未来的发展趋势,包括第19~21章。

第19章测验偏向,主要讨论了测验对少数族群成员的公平性问题,并对解决方法提出了建议。

第20章测验与法律,介绍了美国政府控制测验使用的法律以及影响心理测验的主要诉讼案件。

第21章心理测验的伦理及未来,介绍了测验在专业、道德、社会等方面涉及的伦理问题,以及当前和未来的发展趋势。

序 言

Preface

心理学是一个充满魅力的广阔领域。心理学家工作在各种不同的场合，从学校、诊所到基础研究实验室，从制药公司到私人跨国公司。尽管工作环境不同，所有的心理学家至少在以下两方面存在共同之处：都在研究行为；都在某种程度上依赖于对行为的测量。本书探讨的就是一种特殊类型的测量工具——测量人类所有行为特征的心理测验。

本书是作者们长期合作的结果。作为心理测验开发和使用的积极推动者，我们深感遗憾的是，有太多的大学本科生将心理测验视为枯燥乏味、与他们的目标或职业兴趣无关的一门课程。相反，我们认为心理测验是一个充满魅力的领域。尽管由于一些挑战、新进展和争议，心理测验处于不断的变迁之中，但它在心理学的发展历史中的地位牢不可破。心理测验方面的书应该加强而不是减弱学生的兴趣。因此，我们提供了有关心理测验和测量原理的多方面概述，以一种特殊的风格来吸引当代大学生。

为了理解心理测验的应用和相关问题，学生必须学习一些以初级统计学知识为基础的基本原理。因此，进行一些复习并且认真阅读第一部分，有助于理解第二部分中讨论的有关测验应用的问题。第三部分论述了一些影响测验未来发展的的问题，包括测验焦虑、测验偏差以及与测验有关的法律问题。心理学应用的前景如何，或许就看心理学家们有无应对这些挑战性问题的能力了。

贯穿全书，我们呈现了一系列重点讨论和重点范例。书中的各个部分都是通过举例，或提供对于某特定问题的翔实讨论来进行阐述。我们还使用“日常生活中的心理测验”的专栏来解释统计方法等材料。

更加强调应用

当今的学生往往青睐非正式的讨论和个性化的例子。因此，我们决定借鉴各领域的叙述模式，以一种非正式的风格写作本书。但是，由于测验是一个严肃而复杂的领域，而且许多学者和专家在这个领域存在着重大的分歧，因此，我们对这些测验上的争议进行了更正式的讨论和更详尽的说明。

《心理测验：原理、应用和问题》第1版诞生于1982年，至今已过去了20多年，世界在许多方面发生了变化。例如，个人计算机在1982年还是一种新事物，大多数学生和教授从未听说过电子信箱或互联网，心理测验的应用远远少于今天。然而另一方面，心理测验的原理却一直没有太大变动。因此，新版本对有关原理的章节只做了一些改进和完善，而后面有关应用和问题的章节则有很大的变化。

不仅心理测验领域发生了变化，我们当中的一位作者罗伯特 M. 卡普兰在经历了大半生医学院教授的职业生涯后，现在就职于公共健康学院。另一位作者丹尼斯 P. 萨库兹完成了法学院的学业，现在既担任心理学教授也兼职做法律教授。在保持心理学家身份的同时，我

们也有机会探索心理测验在医疗、公共卫生、教育和法律等方面的前沿应用。相对于以往所有版本，第7版进一步阐明了心理测验在各种领域的广泛应用。

在第7版中，我们对应用领域的主题有所发展。第11章论述心理测验在教育及特殊教育中的应用；第12章着眼于标准化测验在教育、公共服务和军队中的应用；第13章讨论心理测验在临床和咨询场所中的使用。

计算机时代使心理测验发生了彻底的变革。我们通过讨论计算机自适应测试和项目反应理论来处理原理章节中的一些问题。在第15章，我们讨论了心理科学在计算机时代的应用；第16章讨论了心理测验在咨询心理学领域的应用，主要集中于兴趣测验；第17章探讨了心理评估在健康心理学、医学和保健领域的迅速发展；第18章回顾了心理测验在工业和商业场所的应用。

最后几章涉及的是有关心理测验延续至现代的争议问题，反映了社会正义、法律和道德的新发展。

本书的第1版是在打字机上完成的，当时还没有广泛使用文字处理软件，很少有教授和学生使用个人计算机。本书的早期版本还为准备在大型通用计算机上进行统计分析提供了说明。在第3版出版的时候，大学生们还没有广泛使用互联网。现在，几乎所有的学生都有机会使用互联网，因此我们提供了通用互联网资源，并且对使用计算机实施测验进行了深入讨论。

第7版的结构：对教授使用的提示

《心理测验》一书的7个版本的编写出版工作历时超过25年，是一项既具有挑战性也具有回报性的工作。我们深感荣幸的是，遍布全世界的数百所学院和大学的数百名教授采用了我们的教科书。一些教授建议我们重新编排这本书，以便于他们在课堂上使用。为了方便各种不同的使用目的，我们尽量使各章保持足够的独立性，这样教授就可以按他们选择的顺序进行教学。当然也可以按我们编排的章节顺序进行教学。

有一些教授希望强调心理测验学方面的问题，那他们可以选择第1~7章，接下来是第19章和第20章。然后，他们可以从应用部分返回某些章节。而有些学校要求预修复复杂的统计课程，那么第2章和第3章可省略不讲。有些教授强调应用，他们可以选择第1~5章，然后直接进行第二部分，有些教授仅采用本书的几个章节。虽然第9~13章最有可能被用做基础课程，但第14~18章也饶有趣味，请予保留。第17章和第18章是心理测验扩展出的较新领域。最后，第19章和第20章既可以在课程结束时使用，也可以在课程开始时使用。例如，一些教授喜欢在第5章后就跳到第19章和第20章。

术语表

Glossary

A

- achievement 成就** 过去学会的。
- acquiescence 默认** 同意或认可测验项目为真的倾向。
- adverse impact 不利影响** 任何测验用于职业选拔时, 申请者中少数民族被拒绝的比例稳定地高于主流群体的系统效应。
- age differentiation 年龄分化** 年龄大的儿童比年龄小的儿童能力更强, 基于这一事实而产生的区别。
- age scale 年龄量表** 项目按照年龄水平分组的一类测验。(例如比奈量表, 在特定年龄的一组有代表性的儿童中, 2/3或3/4的儿童能成功通过的项目, 组成该年龄水平的测试项目。)
- anxiety 焦虑** 一种以烦恼、忧虑和紧张为特征的不愉快的情绪状态。
- aptitude 能力倾向** 学习某种特定技能的潜能(如音乐的能力倾向)。
- assessment 评估** 用来评价个体以便描述其当前机能状况并预测其将来机能的一种程序。评估过程中常常会使用测验。

B

- basal 基准** 正确反应数量的最低标准所达到的水平。
- basal age 基准年龄** 在斯坦福-比奈量表上, 被试成功通过所有测验项目的最高年龄水平。
- base rate 基础率** 在决策分析中, 随机选择的人预期在某个标准上成功的比例。
- biserial correlation 二列相关** 表示一个连续变量和一个人为二分变量之间关联程度的指标。

C

- category format 分类形式** 一种评定量表形式, 通常采用从1~10的分类级别。
- ceiling 高限/天花板** 测验可能得到的最高分。当一个测验太容易时, 许多人可能得到最高分, 从而无法区分高水平 and 较低水平的被试。
- class interval 组距** 频数分布的横轴的单位。
- closed-ended question 封闭式问题** 访谈中能明确回答的问题(如“是”和“否”)。这类问题一般要求受访者

回忆一些东西。

- coefficient alpha α 系数** 一种估计信度的通用方法, 类似于 KR_{20} 公式, 但允许项目有一个值, 而不要求采用0和1记分。
- coefficient of alienation 不相关系数** 在相关和回归分析中, 两个变量之间不具有关联性的指标。
- coefficient of determination 决定系数** 相关系数的平方; 当Y是X的函数时, 对由X造成的Y的变异, 在Y的全部变异中所占比例的估计。(反之亦然。)
- concurrent validity evidence 同时效度证据** 指测验和效标在同一个时间点施行时所获得的效标效度证据。
- confrontation 面质** 指出差异或矛盾的陈述。
- construct validity evidence 构想效度证据** 一种通过一系列研究来确定测验意义的方法。为了评价结构效度证据, 研究者在定义某些测验构想的同时要开发工具来测量它。如果能够观察到测验与其他测量之间存在相关, 就可以为测验的意义提供证据。参见会聚证据和区分证据。
- content validity evidence 内容效度证据** 表明测验内容代表了它想要包含的概念范围的证据。
- convergent evidence 会聚证据** 得到的证据表明测验具有与其他测量同样的特性, 是结构效度证据的一种形式。
- correction for attenuation 衰减校正** 在估计测验与其他测验之间的相关时, 对低信度导致相关减少所做的校正。衰减校正公式是用来估计当变量完全可靠时这种相关是多少。
- correlation coefficient 相关系数** 用来描述两个变量之间联系的方向和大小的数学指标。相关系数的取值范围在-1.0~1.0之间。
- criterion-referenced test 标准参照测验** 用来描述个人对于明确规定了掌握标准的某种技能、作业或知识掌握情况的测验。标准参照测验的内容受到其明确定义的目标的限制。
- criterion validity evidence 效标效度证据** 表明测验分数与对所感兴趣的事物的精确测量相互一致的证据。对所感兴趣事物的测量就称为效标。
- cross validation 交叉验证** 用于最初研究样本不同的样本评价一个测验或者回归方程的过程。

D

deciles 十分位数 把频数分布划分成相等的十份的各个点。
descriptive statistics 描述统计 用于提供对所搜集的数量信息进行简明描述的一系列方法。

developmental quotient, DQ 发展商数 在格赛尔发展量表中, 根据对儿童与成熟相联系的行为发生与否的评估而确定的测验分数。

dichotomous format 双选形式 一种项目形式, 每个项目只有两个可选项。

differential validity 差异效度 同一个测验对于不同人群的意义不同的程度。例如, 一个测验可以有效地预测白人学生在大学里的成功, 对非裔美国学生却不行。

discriminability 区分度 在项目分析中, 与某种标准相比, 某个项目的作用多么好。例如, 可以根据项目把测验上得分高低不同的两个组区别开的程度对它们进行比较; 区分度的指标是被试在项目上的表现和在整个测验上的表现之间的相关。

discriminability analysis 区分度分析 见区分度。

discriminant analysis 区分分析 一种多元数据分析方法, 其目的是发现变量之间的线性联系, 以便最好地描述如何将一组人分到不同类别。

discriminant evidence 区分证据 得到证据表明测验测到了与其他已有测验所测的不一样的东西, 是构想效度证据的一种形式。

distractor 干扰项 多重选择考试中的不正确或者不给分的选项。

drift 偏离 在行为研究中, 观察者偏离其在培训时学到的行为定义而形成自己的特别定义的倾向。

dyslexia 诵读困难 一种特殊的阅读障碍, 表现为阅读能力退步。

E

EEOC guidelines 就业机会均等委员会指导纲要 就业机会均等委员会 (Equal Employment Opportunity Commission, EEOC) 为了确保雇用的公平而制定的一套程序, 讨论了用于雇用选拔的心理测验在信度和效度上的最低要求。

estimated learning potential, ELP 学习潜能估计 在桑珀 (SOMPA) 系统中, 根据儿童社会经济背景调整在韦氏儿童智力量表-修订版 (WISC-R) 上的得分, 以此代替智商。

evaluative statement 评价性陈述 在访谈中做出的表示判断或评价的陈述。

expectancy effect 期望效应 实验或者测验的结果受到实验者或施测者想要发现什么的期望影响的一种倾向 (在

心理学家Rosenthal对此问题进行了重要的研究之后, 这种效应也被称为罗森塔尔效应)。

F

face validity 表面效度 一个测验中的项目表面上看来是有意义的或有关的程度。由于表面效度不能作为推论的基础, 所以实际上它不能作为效度的证据。

factor analysis 因素分析 一套对多元数据做分析的方法, 目的是将一巨大的变量组合为线性的相关矩阵缩减为较少的变量。

false negative 错误否定 在测验-决策理论中, 某一个案例被测验判断为否定类别, 但其正确类别为肯定。

false positive 错误肯定 在测验-决策分析中, 某一个案例被测验判断为肯定类别, 但其正确类别为否定。

four-fifths rule 4/5规则 联邦机构用来判断雇用机会是否均等的规则。在进行雇用选拔时, 对于任何种族、性别或民族的人群, 如果其录取率低于录取比率最高的群体4/5 (80%), 就可以认为该选拔程序具有不利影响。

frequency distribution 频数分布 把一次测量的各个分数值进行系统排列, 以便反映该测量的每一分数发生的频繁程度。

G

general cognitive index, GCI 综合认知指数 在麦卡锡儿童力量表 (McCarthy Scales of Children's Abilities) 中使用的一种平均数为100, 标准差为16的标准分数。

group test 团体测验 一种可以同时对不止一个人施测的测验。

H

hit rate 命中率 在测验-决策分析中, 测验能够准确地预测其成功或者失败的个案比例。

hostile statement 敌意性陈述 访谈中表示愤怒的陈述。

human ability 人的能力 表明个体已经学到的或者具有可能导致特定行为的才能的行为, 包括成就、能力倾向和智力。

I

individual test 个别测验 一种在一段时间内只能对一个人施测的测验。

inference 推论 以证据为基础, 对不能直接观察到的东西所做出的合乎逻辑的推理。

inferential statistics 推论统计 根据所观察的小团体 (样本) 进行推论的多种方法, 随后把这些推论应用到更大的团体 (总体)。一般说来, 研究者想要了解较大的团

体,但不可能对所有成员进行必要的观察。

intelligence 智力 独立于先前学习的一般性的潜能。

intelligence quotient, IQ 智商 表示智力测验结果的单位。

智商就是个体的智力年龄(MA,由测验测定)与其实
际年龄或自然年龄(CA)的比率: $IQ = MA/CA \times 100$

intercept 截距 当 $X=0$ 时,二维图形在 Y 轴上的取值。在回
归分析中,该点就是回归直线与 Y 轴的相交点。

interquartile range 中四分间距 第25百分位和第75百分
位之间的分数距离。

interval scale 等距量表 一种可以用来对顺序对象分等的
量表,其中的每个单位都反映了同等程度的所测特性。

interview 访谈 通过谈话、讨论或者直接提问收集信息的一
种方法。

ipsative score 个人基准记分法 根据相对比较而非绝对
标准呈现测验结果。个人基准记分法是将个体与他/她
自己进行比较,这样就提供了每个人的自我参照框架。

isodensity curve 等密度曲线 在散点图(二维散点图)
上包含了组成某一团体的特定比例的个案的一个椭圆。

item 项目 要求人做出反应并可以记分和评价的特定刺激。

item analysis 项目分析 用来评价测验项目的一套方法。
最普通的方法包括对项目难度和项目区分度的评估。

item characteristic curve 项目特征曲线 作为项目分析方
法一部分的图形。在 X 轴上表示测验总分,在 Y 轴上表示
被试通过某项目的比例,就可以得到该项目的项目特征
曲线。

item difficulty 项目难度 对项目究竟有多难进行评估的一
种项目分析形式。最常用的难度指标是正确回答项目的
被试所占的百分比。

item discriminability 项目区分度 参见区分度(discrim-
inability)。

K

Kuder-Richardson 20 库存德-理查逊20公式 一个估计
测验内部一致性的公式。该方法相当于测验项目所有可
能分半相关的平均。 KR_{20} 公式应用的前提条件是所有项
目都采用0或1记分。

L

Likert format 李克特量表 态度量表的一种形式,被试采
用以下等级来表明他们赞同某个陈述的程度:非常不同
意、不同意、既非不同意也非同意、同意、非常同意。

M

McCall's T 麦柯尔T分数 一种平均数为50,标准差为10
的标准分数体系。 Z 分数经过简单的线性转换就可以得

到麦柯尔 T 分数($T=10Z+50$)。

mean 平均数 某个变量一系列得分的算术平均数。

measurement error 测量误差 测验的观测分数中既不是
真分数也不是你想要测量的品质的那些成分。

median 中数 频数分布中表示第五十个百分点的那个点。

mental age 智力年龄 表示智力测验结果的一种单位。这
种单位的基础是把个体在测验上的表现与某一特定实足
年龄的团体在测验上平均表现进行比较。

multiple regression 多元回归 一种多元数据分析方法,
用来评估一个连续的结果变量与两个以上有线性联系的
预测变量之间的关系。

multivariate analysis 多元分析 用来考察三个以上变量
之间联系的一整套数据分析方法。

N

nominal scale 称名量表 一个随意给物体指派数字的体
系。从称名量表中得到的数字不能进行数学运算。例如,
足球运动员的球衣背上的数字就是一种称名量表。

normative sample 常模样本 由在标准条件下接受测验,
也就是完全按照测验手册规定的指导语、形式和一般程
序施测的个体所组成的对照团体(也称标准化样本)。

norm-referenced test 常模参照测验 一种根据与常模团
体的关系评价个体的测验。

norm 常模 一组在标准化条件下接受测验的被试的表现
的概要总结。一个常模通常包括参照团体的平均数和
标准差,以及如何把原始分数转换为百分等级的有关
信息。

O

one-tailed test 单测检验 对虚无假设的某一方向的检验。
实验者强调特定一端的虚无分布应该做单测检验以确定
拒绝此虚无假设的区间。例如,一个研究减肥的实验者
可能认为B组比A组体重减轻更多,只有当B组体重减轻
的量在统计上不大可能比A组大时才能拒绝虚无假设。
如果A组比B组减轻更多,则不能拒绝虚无假设。

open-ended question 开放性问题 在访谈中一般不能做
出特定回答的问题。这类问题要求被访谈者自发地做出
反应。

ordinal scale 顺序量表 一种可以用来对物体或者个体排
列顺序的量表。

P

parallel forms reliability 平行复本信度 用来评价与使用
一套特定的项目相联系的误差的一种信度评估方法。按
照同样规则编制的两个测验相互等值,这两个测验之间

的相关就是平行复本信度的估计。

pearson product moment correlation 皮尔逊积差相关
两列连续变量之间相关的指标。

percentile band 百分带 最有可能代表被试真实分数的百分位数范围。首先得到所得分数上下各一个测量标准误差范围内的分数所构成的区间，然后将其转换为百分位数。

percentile rank 百分等级 低于某一特定分数的所有分数所占的百分比。

performance scale 操作量表 由要求被试做一些事而不是回答问题的任务组成的一种测验。

personality test 人格测验 测量个体明显的或者内隐的性情（即个体在任一给定情况下表现出某种特定行为或者反应的倾向）的一种测验。人格测验测量人的典型行为。

point scale 点量表 一种给每个题目指定分数（如0、1、2）的测验。在点量表中，所有涉及特定内容的项目可以集合到一起。

polytomous format 多选项形式 客观性测验的一种形式，其中每个题目都有三个以上的可选择项。最常见的是多重选择测验。也可写作 **polychotomous format**。

predictive validity evidence 预测效度证据 测验对未来某一时间的效标分数预测有效性的证据。

probing statement 追问性陈述 在访谈中要求被访谈者提供比他/她已经说出的信息更多的信息的一种陈述。

projective hypothesis 投射假说 认为个体在试图理解一个模糊或者含义不明确的刺激时，其解释会反映其需要、情感、经历、先决条件和思维过程等。

projective personality test 投射性人格测验 一种刺激或要求的反应不明确甚至二者都含糊不清的测验。投射测验隐含的普遍观点是，一个人对含糊不清的刺激所做的解释反映了其独特的性格特征。

prophecy formula 预言公式 斯皮尔曼和布朗提出来的一个公式，当采用分半法计算测验信度并且每一半测验都是整个测验长度的一半时，可以用这个公式来校正对信度的低估。也可以用来估计要达到期望的信度水平所必须增加的测验长度是多少。

psychological test 心理测验 一种测量人的外显的（可观察的）和内隐的（个体内的）特征的手段。心理测验可以测量过去、现在或者将来的人的行为。

psychological testing 心理测试 心理测验的运用，包括各种可能的使用、应用以及心理测验的基本概念。

Q

quartiles 四分位数 把频数分布分为相等的四个部分的点。

R

randomly parallel test 随机化平行测验 指从一个范围或体系的项目中连续地随机抽取项目组成的多个测验。

ratio scale 比率量表 一种有绝对零点（该点没有所测特性）的等距量表。

reactivity 反应性 在行为研究中，当观察者知道其工作受到监控时引起的量表信度提高的现象。

reassuring statement 安慰性陈述 访谈中，使人感到舒适或者被支持的陈述。

receptive vocabulary 接受的词汇 在皮博迪图画词汇测验中，对语言智力（既通常说的理解语言的能力）的非语言估计。

regression line 回归线 在散点图上通过一组点的最佳拟合直线。

reliability 信度 一个分数或测量在多大程度上没有测量误差。理论上，信度就是真分数变异与观测分数变异的比率。可以采用一系列相关方法估计这个比率，如 α 系数、 KR_{20} 公式、重测法和复本法。

representative sample 代表性样本 由具有与测验应用对象类似特征的个体组成的无偏向或随机抽取的样本。

residual 残差 从回归方程得到的预测值与观测值之间的差异。

response style 反应风格 按照固定方式对测验项目进行反应而不考虑其内容的一种倾向。

restricted range 范围受限 在相关和回归中，都是用一个测量的变异预测另一个测量变异，如果其中任何一个测量的变异是有限的，那么所得相关便可能较低。例如，GRE分数与一个精英项目中的学生学习成绩之间的相关可能较低，因为这些参与这一项目的学生GRE分数变异很小，而对于所有大学的全体学生而言，（GRE和学习成绩之间的）真实相关则可能较高。

Rosenthal effect 罗森塔尔效应 参见期望效应（expectancy effect）。

S

scaled score 量表分数 韦克斯勒测验的一种平均数为10、标准差为3的标准分数。

scales 量表 把测验项目的原始分数同某种理论和经验分布联系起来的工具。

scatter diagram 散点图 一种表明两个变量之间关系的图形。从每个人身上得到一对观测值，然后标绘在由变量X和变量Y组成的二维空间内。

selection ratio 录取率 在测验决策分析中被选中的申请者所占的比例。

self-report questionnaire 自我报告问卷 一种问卷, 提供有关个体的陈述, 并要求被试用某种方式回答每个陈述, 如“是”或“否”。

shrinkage 收缩率 很多时候, 为某一团体建立的回归方程会被用来预测另一个团体内被试的表现。这种做法倾向于过高估计第二团体内相关的强度。这种从原来样本得到的相关强度在第二个样本中减少的量称做收缩率。

social ecology 社会生态学 一个涉及环境与行为之间的关系、描述行为背景以及其他相关主题的研究领域。

social facilitation 社会助长作用 人们做出与其周围榜样类似行为的倾向。

Spearman's rho 斯皮尔曼 ρ 计算两列等级变量相关的方法。

split-half reliability 分半信度 一种把测验分为两半的信度估计方法。由于长度缩短为一半, 对两半测验之间的相关需要进行校正, 才能作为信度的估计。

standard administration 标准施测 测验手册中规定的测验施程序。

standard deviation 标准差 离均差的平方和的平均值(即方差)的平方根, 一般用来衡量分数分布的变异性。

standard error of estimate 估计标准误差 一个表示回归方程准确性的指标。它等价于回归分析中残差的标准差。估计标准误差越小, 预测越准确。

standard error of measurement 测量标准误差 一个表示测验或者测量中的误差大小的指标。测量标准误差就是采用同一个测验得到的一系列观测值的标准差。

standardization sample 标准化样本 由在标准条件下接受测验, 也就是完全按照测验手册规定的指导语、形式和一般程序施测的个体所组成的对照团体(也称常模样本)。

standardized interview 标准化访谈 按照手册或者程序指导书明确规定的标准条件进行的访谈。

stanine system 标准九分制 把数字1~9分派到测验分数上的一个系统, 是由美国空军发展出来的一种平均数为5、标准差大约为2的标准化九级分数分布。

state anxiety 状态性焦虑 对情景的一种情绪反应, 在不同的情景中表现不同。

stress 紧张(应激) 对引发需要、受到强制或者得到机会的情景产生的一种反应。

structured personality tests 结构化人格测验 提供一个通常自我报告式的陈述(如“我喜欢摇滚乐”), 要求被试在两个或者更多的反应选项(如“是”或“否”)中进行选择的测验。有时也称为客观性人格测验(objective personality tests)。

subtest scatter 分测验离散程度 在韦氏测验中, 分测验的变异性大小。

T

Taylor-Russell tables 泰勒-罗塞尔预期表 依据测验所提供的信息量超出偶然所得信息的多少来评价测验效度的一系列表格。

test 测验 一种用来量化行为的测量工具。

test administration 施测 实施测验的行动。

test administrator 施测者 负责测验实施的人。

test anxiety 测验焦虑 在接受测验的情景下发生的焦虑。

test battery 成套测验 将分数共同用于评价个体的测验集。

test-retest reliability 重测信度 一种估计由于时间取样或者者在两个不同的时间点施测同一个测验所引起的测量误差有多少的方法, 通常采用同一个测验的两次不同施测的成绩之间的相关来估计。

third variable 第三变量 一个可以解释另外两个变量之间的显著关系的变量。

tracking 轨迹 与同龄人相比, 保持同等发育或表现水平的趋势。

trait anxiety 特质性焦虑 一种反映个体在面临压力性情境时反应强度差异的人格特征。

trait 特质 个体独立于情境的持久的、稳定的特点。

T score T分数 在MMPI中使用的一种平均数为50、标准差为10的标准分数(参见麦柯尔T分数)。

true score 真分数 如果没有测量误差, 在测验或者测量上获得的分数。实际上, 真分数只能估计而不能直接观测到。

two-tailed test 双尾检验 对虚无假设的无方向的检验。同强调特定方向的单测检验相反, 双测检验是用于同时评价样本分布的高低两端所观察到的显著差异是否是由机遇引起的。

U

understanding response 理解性反应 在访谈中一种表示理解的陈述。也称为共情反应(empathy response)。

unstructured interview 无结构访谈 没有任何具体的或者特定的问题或问题顺序的访谈。

V

validity 效度 一个测验测到它想测的特征的程度。效度的类型包括内容效度、效标效度和构想效度证据。

variance 方差 离均差平方和的平均值, 标准差的平方。

目 录 Contents

导读

全书各部分简介

序言

术语表

第一部分 原 理

第1章 导论1

1.1 基本概念4

1.1.1 什么是测验5

1.1.2 测验的类型6

1.2 本书概要8

1.2.1 心理测验的原理8

1.2.2 心理测验的应用8

1.2.3 心理测验的问题9

1.3 历史回顾9

1.3.1 早期发展9

1.3.2 达尔文和个体差异10

1.3.3 实验心理学和心理物理测量10

1.3.4 智力的演变和标准化成就测验11

1.3.5 人格测验：1920~194014

1.3.6 人格测验新方法的出现16

1.3.7 测验地位快速变化时期16

1.3.8 当前的环境18

1.4 小结19

第2章 测验常模和基本统计20

2.1 我们为什么需要统计20

2.2 测验量表21

2.2.1 量表的特征22

2.2.2 量表的类型23

2.2.3 可进行的运算24

2.3 频数分布24

2.4 百分等级27

2.5 百分位数30

2.6 描述分数分布31

2.6.1 平均数31

2.6.2 标准差32

2.6.3 Z分数33

2.6.4 标准正态分布34

2.6.5 麦柯尔T分数40

2.6.6 四分位数和十分位数41

2.7 常模42

2.7.1 年龄常模44

2.7.2 轨迹45

2.7.3 标准参照测验46

2.8 小结49

第3章 相关与回归52

3.1 散点图53

3.2 相关54

3.3 回归55

3.3.1 回归直线55

3.3.2 最优拟合直线56

3.3.3 相关系数的统计显著性检验62

3.3.4 如何解释回归图形63

3.4 其他相关系数65

3.5 相关分析中的术语和争议67

3.5.1 残差67

3.5.2 估计的标准误68

3.5.3 决定系数68

3.5.4 不相关系数68

3.5.5 收缩率68

3.5.6 交叉验证69

3.5.7 相关-因果问题69

3.5.8 第三变量解释70

3.5.9 范围受限	70
3.6 多元分析	71
3.6.1 一般方法	71
3.6.2 多元回归应用举例	72
3.6.3 判别分析	73
3.6.4 因素分析	73
3.7 小结	76
附录3A 回归方程和相关系数的计算	77
3A.1 回归方程的计算	78
3A.2 相关系数的计算	79

第4章 信度

4.1 信度的历史和理论	82
4.1.1 误差的概念	82
4.1.2 斯皮尔曼的早期研究	82
4.1.3 测验分数理论的基础	83
4.2 领域抽样模型	84
4.3 项目反应理论	85
4.4 信度模型	86
4.4.1 误差来源	87
4.4.2 时间取样: 重测法	87
4.4.3 项目取样: 复本法	89
4.4.4 分半法	89
4.4.5 KR_{20} 公式	91
4.4.6 α 系数	92
4.4.7 差异分数的信度	93
4.5 行为观察研究中的信度	94
4.6 与信度评估方法相关的误差来源	97
4.7 信度信息的使用	100
4.7.1 测量标准误和弹性码尺	100
4.7.2 什么样的信度是可信的	101
4.7.3 面对信度低怎么办	102
4.8 小结	104
附录4A 测验的两部分方差不同时利用 α 系统来估计分半信度	105
附录4B 利用 KR_{20} 计算信度	105

第5章 效度

5.1 效度的定义	108
-----------------	-----

5.2 效度的各方面	109
5.2.1 表面效度	109
5.2.2 与内容有关的效度证据	109
5.2.3 与标准有关的效度证据	111
5.2.4 与构想有关的效度证据	120
5.2.5 信度与效度的关系	126
5.3 小结	126

第6章 测验项目的编写和评价

6.1 项目编写	128
6.1.1 项目形式	129
6.1.2 其他可能性	136
6.2 项目分析	139
6.2.1 项目难度	139
6.2.2 区分度	140
6.2.3 项目特征图	142
6.2.4 不同测量的链接	147
6.2.5 标准参照测验的项目	148
6.2.6 项目分析的局限性	149
6.3 小结	149

第7章 测验的实施

7.1 主试和被试	151
7.1.1 主试和被试的关系	151
7.1.2 主试的种族	152
7.1.3 被试的语言	154
7.1.4 施测者的培训	154
7.1.5 期望效应	154
7.1.6 对反应强化的影响	156
7.1.7 计算机辅助测验的实施	158
7.1.8 施测模式	159
7.1.9 被试变量	160
7.2 行为评估的方法学	161
7.2.1 反应性	161
7.2.2 偏离	162
7.2.3 期望	162
7.2.4 欺骗	162
7.2.5 评定误差的统计控制	163
7.3 小结	164

第二部分 应用

第8章 访谈技术	165	离差智商 (SB-LM)	198
8.1 作为测验的访谈	168	9.6 现代比奈智力量表	199
8.2 访谈的交互性质	168	9.6.1 第4和第5版比奈智力量表的模型	199
8.3 有效访谈的原则	169	9.6.2 1986年修订版的特征	199
8.3.1 合适的态度	169	9.6.3 2003年第5版的特征	202
8.3.2 应当避免的反应	170	9.6.4 2003年第5版的心理测量学特性	203
8.3.3 有效的反应	171	9.6.5 中位数效度	204
8.3.4 保持互动顺利的反应	172	9.7 小结	204
8.3.5 对理解的测量	175	第10章 韦克斯勒智力量表: WAIS-III, WISC-IV和WPPSI-III	205
8.4 访谈的类型	176	10.1 韦克斯勒智力量表	206
8.4.1 评价性访谈	177	点量表和操作量表的概念	206
8.4.2 结构化临床访谈	177	10.2 从韦克斯勒-贝尔韦智力量表到 WAIS-III	208
8.4.3 个案历史访谈	181	10.3 量表、分测验和WAIS-III的指数	208
8.4.4 心理状态检查	182	10.3.1 语言分测验	209
8.4.5 提高访谈技能	183	10.3.2 原始分数、量表分数和言语智商	212
8.5 访谈的误差来源	183	10.3.3 操作分测验	213
8.5.1 访谈的效度	183	10.3.4 操作智商	214
8.5.2 访谈的信度	185	10.3.5 全量表IQ	214
8.6 小结	186	10.3.6 指数分数	214
第9章 智力理论和比奈量表	188	10.4 韦克斯勒智力测验的解释特征	215
9.1 智力的界定问题	188	10.4.1 言语智商和操作智商比较	215
9.2 比奈的测验建构原理	190	10.4.2 模式分析	216
9.2.1 原理1: 年龄分化	190	10.4.3 假设性案例研究	216
9.2.2 原理2: 一般心理能力	191	10.5 韦克斯勒成人量表的心理测量学特性	217
9.3 斯皮尔曼的一般心理能力模型	191	10.5.1 标准化	217
9.3.1 一般心理智力 (g) 的含义	192	10.5.2 信度	217
9.3.2 智力的gf-gc理论	192	10.5.3 效度	218
9.4 早期的比奈量表	193	10.6 对韦克斯勒成人量表的评价	218
9.4.1 1905年比奈-西蒙量表	193	10.7 WAIS-III的向下拓展: WISC-IV和WPPSI-III	219
9.4.2 1908年量表	194	10.7.1 韦克斯勒儿童智力量表WISC-IV	219
9.5 推孟的斯坦福-比奈智力量表	195	10.7.2 韦克斯勒幼儿与学龄初智力量表 WPPSI-III	222
9.5.1 1916年斯坦福-比奈智力量表	195	10.8 小结	223
9.5.2 智力商数 (IQ)	196		
9.5.3 1937年量表	196		
9.5.4 1960年修订版斯坦福-比奈量表和			