

实用心理 测验与评估

(原书第2版)

爱德华 S. 诺库格 (Edward S. Neukrug)

奥多明尼昂大学

(美)

R. 查尔斯·福西特 (R. Charles Fawcett)

著

弗吉尼亚大学

李原 孙健敏 何华敏 等译

Essentials of Testing & Assessment
A Practical Guide for Counselors,
Social Workers, and Psychologists, 2nd Edition



机械工业出版社
China Machine Press

实用心理 测验与评估

(原书第2版)

爱德华 S. 诺库格 (Edward S. Neukrug)

奥多明尼昂大学

(美)

R. 查尔斯·福西特 (R. Charles Fawcett)

著

弗吉尼亚大学

李原 孙健敏 何华敏 等译

Essentials of Testing & Assessment
A Practical Guide for Counselors,
Social Workers, and Psychologists, 2nd Edition B841.7 60



机械工业出版社
China Machine Press

社

es

图书在版编目 (CIP) 数据

实用心理测验与评估 (原书第2版)/(美) 诺库格 (Neukrug, E. S.), (美) 福西特 (Fawcett, R. C.) 著; 李原等译. —北京: 机械工业出版社, 2013.1
(心理学精品教材译丛)

书名原文: Essentials of Testing & Assessment: A Practical Guide for Counselors, Social Workers, and Psychologists

ISBN 978-7-111-40948-9

I. 实… II. ①诺… ②福… ③李… III. 心理测验 IV. B841.7

中国版本图书馆 CIP 数据核字 (2012) 第 309221 号

版权所有·侵权必究

封底无防伪标均为盗版

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: 01-2010-1050

这是一本兼具科学性与实用性、既通俗易懂又严谨规范的心理测量学教科书。本书要点明确、详略得当、深入浅出、案例丰富, 涵盖了心理测验与评估的经典理论和最新进展, 实用性强, 写作风格非常务实, 你会在阅读的过程中发现其趣味性。本书既适合于初学者, 也可作为研究生的教学参考书, 尤其适合于应用心理学专业的教学, 也适合咨询师、社会工作者及心理医生使用。

Edward S. Neukrug. R. Charles Fawcett. Essentials of Testing & Assessment: A Practical Guide for Counselors, Social Workers, and Psychologists, 2nd Edition.

Copyright © 2010, 2006 Brooks/Cole, Cengage Learning.

Original edition published by Cengage Learning. CMP Press is authorized by Cengage Learning to publish and distribute exclusively this simplified Chinese edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

All rights reserved.

本书原版由圣智学习出版公司出版。本书中文简体字翻译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地区)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。

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: 赵艳君 版式设计: 刘永青

藁城市京瑞印刷有限公司印刷

2013年2月第1版第1次印刷

185mm×260mm·15.5印张

标准书号: ISBN 978-7-111-40948-9

定 价: 49.00 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

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

PREFACE

前言

出版界有一个不肯告人的小秘密：一本书的第2版往往优于第1版。这本书同样如此，我们希望并相信如此！为什么优于第1版？首先，我们删除了一些晦涩难懂的句子和零星的我们确认是“错误”的地方。其次，我们吸取了读者的建议，增加了缺少的内容。增加的内容超出了我们的预期，具体包括以下内容：

- 第2章中的司法鉴定部分
- 第3章中关于漏报与误报的讨论
- 第3章中项目反应理论部分
- 第4章中关于累积分布的进一步讨论
- 第5章中估计的标准误
- 第6章关于美国教育进步评估的部分
- 第2, 5, 6章中对高风险测验的进一步讨论
- 第7章中非言语智力测验部分，新纳入了对非言语智力全面测验、普遍智力测验的讨论
- 第7章中神经心理评估部分，增加了对 Halsted-Reitan 和 Boston 过程方法的讨论
- 第9章中增添了新的临床评估，如 16PF、NEO PI-R、NEO-FFI、PAI 和 SASSI
- 第10章中基于绩效的评估部分

我们还调整了本书中的其他很多方面。例如，我们根据 ACAs 伦理法案的最新版本对相应的内容做了更新，扩展了关于 ADA 和 FERPA 的部分。为了方便读者检验测验的好坏，我们以表格形式归纳了测验的7条标准，并将伦理法案移至附录 B。我们对历史数据和事件部分增加了一些内容，增加了有关如何在测评中使用计算机的内容，扩展了有教无类法案 (No Child Left Behind) 和高风险测验部分。第7章中删除了 CPI 的内容，将环境评估移至第10章。当然，我们将所有的测评工具都更新至修订后的最新版本。

尽管这一版的内容有所增加和改变，我们仍然相信这本书所涵盖的关于测量和评价的内容可以在一个学期内学完，不需要更多的时间，因为这是一本精要。但是，请不要被精要二字所迷惑。精要也包含了极其丰富的内容！本书的写作风格是非常务实的，相信你会在阅读的过程中发现本书的趣味性。此外，我们安排了一些有助于大家理解的故事和插图，相信本书会为你提供心理测验与评估课程所需的全部信息。

本书分为四篇：第Ⅰ篇，历史和当前职业问题；第Ⅱ篇，测量的价值和测量统计学；第Ⅲ篇，常用的评估技术；第Ⅳ篇，诊断和撰写评估报告。以下是本书包含的四篇内容和附录概览。

第Ⅰ篇包含两章：第1章，测验与评估的历史；第2章，评估中的伦理、法律和职业问题。这两章是对测验与评估领域的概览，是评估的历史与当前困扰我们的问题之间的一个衔接。第1章从测验与评估的区别谈起，并指出当前这些概念的定义如何与其历史发展直接相关。随后，我们回顾了从古代评估到现代评估工具的发展历史。在此过程中，我们重点列举了一些对评估测验的发展起到重要作用的人物，并讨论了出现的一些有争议的问题。本章后面部分还考察了当前评估工具的种类，包括能力测验（成就与性向测验）、人格测验和非正式评估。在本章末尾，我们指出了今天使用评估工具时仍会面临的很多问题。

第2章是关于评估他人时经常遇到的复杂伦理、法律和职业问题。我们首先肯定了伦理决策的复杂性，进而指出一些重要的伦理问题，例如对测验和评估十分重要的规范 and 标准。紧接着，我们讨论了做出明智的、合乎伦理的决策的重要性，列举了一些影响测验使用和评估过程的法律以及以往的诉讼案件。本章以对一些重要职业问题的讨论作为结束，包括评估行业协会、行业协会的认证标准、相对较新但发展迅速的司法鉴定领域、作为整体过程的评估、跨文化评估中的问题以及坚持测验和评估过程的重要性。

第Ⅱ篇讲到测验的价值和测验统计学。组成这三章的三章探究了如何设计测验，及如何对测验进行评分和解释测验分数。这几章均以某种方式使用测验统计学来解释所呈现的概念。在这部分，我们说明了收集和解释测验数据是一个需要细致规划的过程及其方法，它包括理解人与人之间差异的科学方法。

第3章讨论了测验的价值，考察了测验开发和测验使用中的四个重要领域：效度，一个测验是否测出了应该测的内容；信度，一个个体在测验上获得的分数，是否反映了他的真实情况；跨文化的公平性，一个个体获得的分数是不是个体的真实反映，而不受测验所固有的文化偏见的影响；实用性，在特定情境下，使用某一测验是否有意义。在考察了这4个因素之后，为确保测验的价值，我们在分析的基础上总结出选择一个测验时应该遵循的5个步骤。本章还介绍了两种统计方法相关系数和判定系数的解释。我们之所以在这一章介绍这些统计方法，是因为这些方法对于理解本章和后面章节的内容至关重要。

第4章从原始分数开始了解统计知识。在本章中我们指出原始分数一般很少提供关于一组分数有意义的信息，我们提供了如何操作这些分数以使得一组数据有意义的方法。因此，在本章中我们考察了如何运用下列概念来帮助我们理解原始分数：频数分布，直方图和频数多边形，累积分布，正态曲线和偏态曲线，集中趋势的测量如平均数、中位数和众数，以及离散程度的度量如全距、四分位差和标准差。

第5章讨论了如何创建新分数来解释测验数据，考察原始分数如何被转化为导出分数从而使个体能更容易地理解他们的原始分数的意义。我们首先区分了常模参照与标准参照测验并讨论了导出分数：百分位数；标准分，包括 z 分数、 t 分数、离差智商、标准九、斯登

分数、正态曲线等值、大学入学考试分数（如 SAT、GRE 和 ACT）和出版商型分数；发展常模，如年龄对比和年级当量。本章接着考察了测量的标准误和估计的标准误，两者都提供了理解个人“真分数”所落范围的方法。本章最后讨论了称名、顺序、定距和比率尺度，探讨了每个尺度的独特属性并指出不同的测量工具要用不同的尺度。

第Ⅲ篇考察一些常用的测验和在不同领域中使用这些测验的程序，如同每章标题描绘的那样，第6章教育能力评估，第7章智力和认知功能评估，第8章职业和职业评估，第9章临床评估，第10章非正式评估。

第6章考察了通常在学校实施的测验，以测量学生学到了什么以及他们能学到什么。首先我们明确了教育能力测验的目的，然后介绍了各类教育能力测验，例如成套调查测验、诊断性测验、准备性测验及认知能力测验。紧接着，我们考察了每个种类中一些更为常用的测验。在整章中我们提供了关于高风险测验的一些注意事项和由于实施《有教无类法案》所引发的重要问题。

第7章探究了与智力和认知功能相关的测验。本章我们介绍了智力测验和神经心理学评估的简史。此外，我们定义了智力测验和神经心理学评估，并描述了这类评估的一些差异。本章也介绍了一些比较流行的智力模型以及一些比较常用的智力测验和神经心理学评估。

第8章考察了有助于个人对职业或职业生涯进行决策的各种测验。我们首先考察了兴趣调查表，这是一种概括性地考察个人对工作环境的好恶和人格性向的评估方法。接着探讨了多元能力性向测验，这种测验有助于个人识别那些对于他们的职业选择非常重要的技巧和能力范围。最后考察了一些更流行的针对某种能力领域的特殊能力性向测试，例如文书能力、机械能力、艺术能力和音乐能力。

第9章考察了临床评估测验的使用过程。本章首先强调了这种评估应用的广泛性，可作为临床医生或研究者的一项重要工具。然后定义了临床评估，并围绕该评估可能遇到的重要问题展开论述。本章的重点是对较著名、较常用的客观人格测验和投射临床评估程序做了综述。

第10章是非正式评估程序。我们首先给出非正式评估的定义，然后介绍了许多不同类型的非正式评估技术，包括观察法、评定量表、分类方法、环境评估、履历和个人文档，以及基于绩效的评估。最后我们提供了关于非正式评估测验价值的一个讨论。

第Ⅳ篇探究了如何进行诊断（第11章），这是个体评估的一种类型，然后又描述了编写一份全面测评报告的过程，其中部分内容涉及诊断的使用（第12章）。

第11章首先讨论了诊断的重要性，然后描述了《诊断和统计手册》第4版修订版（DSM-IV-TR）。我们简要介绍了DSM-IV-TR的历史，并讨论了做出诊断所使用的五轴。我们还展示了一个五轴的诊断是如何做出的，而且提供了一些案例研究作为附加练习。在本章结尾，我们讨论了在整个评估过程中诊断的重要性。

第12章主要讲述来访者访谈客户的过程及评估报告的撰写。我们首先从定义评估报告的目的开始，然后说明当你将“垃圾（信息）”放进报告中时，你会得到一个很差的报告。

然后讨论准确识别客户问题广度及深度的重要性，这样有助于选择适当的评估流程，以便为诊断报告搜集信息。本章还区分了三种访谈：结构化、非结构化、半结构化。指出近年来计算机在测评报告的撰写中发挥着越来越重要的作用。最后，本章对评估报告的种类进行了细致的描述，并以一个虚拟的来访者为例介绍了如何撰写一份报告的案例。最终形成的评估报告全文可见附录 F。这个例子将有助于你自己撰写评估报告。

在本书的最后，你会发现 6 个附录，目的是帮助你提升对本书中所涉及的一些问题的深入认识。附录 A 提供了主要专业协会和有关它们道德准则的网站；附录 B 提供了美国咨询协会和美国心理协会道德准则评估部分的内容；附录 C 提供了教育领域公平测验实践的规则；附录 D 是补充的统计方程式，以帮助你理解测验统计；附录 E 是 z 分数的百分数转化；附录 F 提供了一个评估报告样本，希望你撰写测验报告有所帮助。

我们认为，本书全面而不繁冗、可读性强，充满趣味性，提供了对测验和评估的概论性介绍。我们希望，读完本书你能对测验和评估有全新的认识。

推荐阅读

■ **心理学与人类世界：无处不在的心理学**

作者：（美）罗伯特 S. 费尔德曼 ISBN: 978-7-111-34800-9 定价：65.00元

■ **人生的心理智慧（原书第9版）**

作者：（美）韦恩·韦登 ISBN: 978-7-111-33058-5 定价：65.00元

■ **左看右看心理学**

作者：（美）戴维 N. 萨特勒 ISBN: 978-7-111-31163-8 定价：45.00元

■ **心理学的50大奥秘**

作者：（美）斯科特 O. 利林费尔德 等 ISBN: 978-7-111-37788-7 定价：39.00元

■ **发展心理学——孩子的成长**

作者：（美）丹尼斯·布伊德 等 ISBN: 978-7-111-34108-6 定价：65.00元

■ **发展心理学：探索人生发展的轨迹**

作者：（美）罗伯特 S. 费尔德曼 ISBN: 978-7-111-34764-4 定价：70.00元

■ **你一生的故事——走近发展心理学**

作者：（美）斯潘塞 A. 拉瑟斯 ISBN: 978-7-111-33699-0 定价：59.00元

■ **打开“积极心理学”之门**

作者：（美）克里斯托弗·彼得森 ISBN: 978-7-111-31422-6 定价：65.00元

■ **社会心理学（原书第12版）**

作者：（美）罗伯特 A. 巴伦 ISBN: 978-7-111-35320-1 定价：65.00元

■ **教育心理学**

作者：（美）莉萨·博林 等 ISBN: 978-7-111-36547-1 定价：69.00元

■ **人格心理学：经典理论和当代研究（原书第4版）**

作者：（美）霍华德 S. 弗里德曼 ISBN: 978-7-111-33921-2 定价：70.00元

■ **工业与组织心理学（原书第4版）**

作者：（美）保罗 E. 斯佩克特 ISBN: 978-7-111-31391-5 定价：70.00元

■ **心理学：主题曲和变奏曲（英文版·原书第8版）**

作者：（美）韦恩·韦登 ISBN: 978-7-111-32965-7 定价：92.00元

■ **认知心理学**

作者：（美）罗伯特·索尔所 等 ISBN: 978-7-111-31144-7 定价：70.00元

■ **实验心理学**

作者：（美）巴里 H. 坎特威茨 等 ISBN: 978-7-111-31168-3 定价：62.00元

■ **心理测验：原理、应用和问题**

作者：（美）罗伯特 M. 卡普兰 等 ISBN: 978-7-111-31169-0 定价：70.00元

CONTENTS

目 录

前 言

PART1 第 I 篇 历史和当前职业问题

第 1 章 测验与评估的历史 / 3

- 1.1 测验与评估的区别 / 3
- 1.2 评估的历史 / 4
- 1.3 评估个体时应考虑的问题 / 14
- 本章小结 / 14

第 2 章 评估中的伦理、法律和职业问题 / 17

- 2.1 评估中的伦理问题 / 17
- 2.2 评估中的法律问题 / 23
- 2.3 职业性问题 / 26
- 本章小结 / 30

PART2 第 II 篇 测量的价值和测量统计学

第 3 章 测验的价值：效度、信度、跨文化公平性和实用性 / 35

- 3.1 相关系数 / 35
- 3.2 效度 / 37
- 3.3 信度 / 42
- 3.4 跨文化公平性 / 46
- 3.5 实用性 / 48
- 3.6 选择和实施一个好的测验 / 49
- 本章小结 / 51

第4章 统计概念：寻找原始分数的意义 / 54

- 4.1 原始分数 / 54
- 4.2 频数分布 / 55
- 4.3 直方图和频数多边形 / 56
- 4.4 累积分布 / 57
- 4.5 正态曲线和偏态曲线 / 58
- 4.6 集中趋势的测量 / 59
- 4.7 离散量数 / 61
- 4.8 记住“人” / 64
- 本章小结 / 64

第5章 统计概念：创建新分数解释测验数据 / 66

- 5.1 常模参照和标准参照 / 66
- 5.2 常模比较和导出分数 / 68
- 5.3 常模参照分数总结 / 75
- 5.4 测量的标准误 / 76
- 5.5 估计的标准误 / 78
- 5.6 测量尺度 / 79
- 本章小结 / 81

PART3 第三篇 常用的评估技术

第6章 教育能力评估：成套调查测验、诊断性测验、准备性测验以及认知能力测验 / 86

- 6.1 定义教育能力评估 / 86
- 6.2 成套调查测验 / 87
- 6.3 诊断性测验 / 92
- 6.4 准备性测验 / 96
- 6.5 认知能力测验 / 98
- 6.6 助人者在教育能力评估中的角色 / 104
- 6.7 对教育能力评估的最终思考 / 105
- 本章小结 / 105

第7章 智力和认知功能评估：智力测验和神经心理学评估 / 110

- 7.1 智力测验简史 / 111

- 7.2 智力测验的定义 / 111
- 7.3 智力模型 / 111
- 7.4 智力测验 / 116
- 7.5 神经心理学评估 / 122
- 7.6 在神经心理学和认知功能评估中辅助者的角色 / 125
- 7.7 关于智力和认知功能评估的最后思考 / 125
- 本章小结 / 126

第8章 职业和职业评估：兴趣调查表、多元能力性向测试以及特殊能力性向测试 / 129

- 8.1 定义职业和职业生涯评估 / 129
- 8.2 兴趣调查表 / 130
- 8.3 多元能力性向测试 / 136
- 8.4 特殊能力性向测试 / 139
- 8.5 职业评估中帮助者的角色 / 142
- 8.6 有关于职业评估的最后思考 / 142
- 本章小结 / 142

第9章 临床评估：客观和投射人格测验 / 146

- 9.1 定义临床评估 / 146
- 9.2 客观人格测验 / 147
- 9.3 投射人格测验 / 160
- 9.4 助人者在临床评估中的角色 / 165
- 9.5 对临床评估的最终思考 / 165
- 本章小结 / 166

第10章 非正式评估：观察法、评定量表、分类方法、环境评估、履历和个人文档以及基于绩效的评估 / 170

- 10.1 定义非正式评估 / 170
- 10.2 非正式评估的类型 / 171
- 10.3 非正式评估测验的价值 / 184
- 10.4 在非正式评估使用中帮助者的角色 / 186
- 10.5 关于非正式评估的最后思考 / 186
- 本章小结 / 187

PART4 第IV篇 诊断和撰写评估报告

第11章 评估过程中的诊断 / 192

- 11.1 诊断的重要性 / 192
- 11.2 什么是 DSM-IV-TR / 193
- 11.3 做出诊断 / 199
- 11.4 评估过程中关于 DSM-IV-TR 的最后思考 / 200
- 本章小结 / 200

第12章 评估报告过程：与来访者访谈并撰写评估报告 / 203

- 12.1 评估报告的目的 / 203
- 12.2 为报告收集信息：无用输入，无用输出 / 204
- 12.3 结构化、非结构化及半结构化访谈 / 205
- 12.4 选择一个合适的评估工具 / 206
- 12.5 撰写报告 / 207
- 12.6 总结评估报告的撰写 / 214
- 本章小结 / 216

附录 A 心理健康专业协会及其道德准则的网站 / 219

附录 B 美国咨询协会及美国心理学会道德准则的评估部分 / 221

附录 C 教育领域公平测验实践的规则 / 227

附录 D 补充的统计方程式 / 231

附录 E z 分数的百分数转化 / 232

附录 F 评估报告 / 233

译后记 / 237

第 I 篇

历史和当前职业问题

本书第 I 篇包括两章：第 1 章，测验与评估的历史；第 2 章，评估中的伦理、法律和职业问题。这两章是对测验与评估领域的概览，是评估的历史与当前困扰我们的问题之间的一个连接。

在第 1 章，我们将回顾测验与评估的历史，以此说明过去的历史是如何影响当前评估方法的。本章从测验和评估的区别谈起，并指出目前这些概念的定义如何与其历史直接相关。随后，我们将回顾评估的历史，从古代评估历史到现代评估工具。在此过程中，我们将重点列举一些对测验和评估的发展起到重要作用的人物，并讨论一些有争议的问题。

本章末尾我们将考察当前评估工具的种类，包括能力测验（认知领域的测验）、人格测验（情感领域的测验）和非正式评估程序。我们还将介绍本书是如何组织上述各种评估类型的。本章将以今天使用评估工具时仍将面对的很多问题作为结束。

第 2 章将考察评估他人时经常面对的复杂的伦理、法律和职业问题。我们首先指出指导精神健康从业者实践的伦理规范，特别强调美国咨询协会和美国心理学会（APA）的规范。这些规范对评估具有十分重要的作用，本章就对这些重要规范的评估方面进行简要介绍。接下来，我们总结了很多已有的指导实践者实施、计分、解释评估过程的标准。此部分最后讨论了伦理决策问题。

本章的后续部分，我们将考察很多影响测验使用的法律和已往的诉讼案件。这些法律约束大都涉及保密性、公正性、测验价值（信度、效度、跨文化公平性和实用性）的问题。本章的最后部分首先强调了评估领域的两个专业协会：咨询和教育评估协会和美

国心理学会第 5 分会：评价、测量与统计分会。然后，我们将介绍制定评估领域课程标准的三个组织，包括美国心理学会、美国学校心理学家学会（NASP）和咨询与相关教育项目认证委员会（CACREP）。随后，我们将讨论评估中日趋热门的次级专业：司法鉴定。在本章中，我们强调只有将评估视为一个整体过程，评估效果才能达到最佳。接下来，我们将探索通过评估技术改善咨询过程的途径。本章结尾将讨论测验中有关偏差的跨文化问题。

测验与评估的历史

“准备好纸笔，开始考试！”这句话仍然让我感到后背发凉。数以百万计的学生一想到考试就感到害怕，这可不是一幅美好的画面。这些有时让人感到害怕的经历有价值吗？你读过本书后将会得到答案。而测验是如何产生的呢？这个问题将在本章得到解答。

(爱德华·诺库格)

在本章中，我们将考察测验与评估的历史。首先，我们将探讨测验与评估的区别，指出目前关于这些概念的定义如何与其历史直接相关。然后，我们将回顾评估的历史，从古代评估历史一直追溯到现代评估工具的发展。在这个过程中，我们将重点列举评估测量领域的一些先驱，讨论曾经出现过的一些有争议的问题。本章末尾，我们将考察当前评估工具的类型，并在本章结尾指出当前仍困扰评估工具使用的问题。

1.1 测验与评估的区别

如今，评估一词包含着为获取某人的信息而进行的一系列宽泛的评价程序 (Hunsley, 2002)。评估程序包括临床面谈；非正式评估技术，如观察法、评定量表、分类方法、环境评估、记录表和个人文档、基于绩效的评估；人格测验，如客观人格测验、投射人格测验、兴趣调查表；能力测验，如成就测验和性向测验 (见图 1-1)。

定义

◆ 评估 (assessment)

一系列宽泛的评价程序。

◆ 测验 (tests)

评估的子集，是指根据收集到的数据计算总分的评估技术。

测验是评估技术的子集，是指通过收集整体数据 (例如，选择考试中答对的题目的总分) 得到总分的评估技术。评估程序可以是正式的，即评估程序是经过深入研究的、科学合理的、有效的和可靠的；也可以是非正式的，尽管也能获得一些有价值的信息，但它不如正式评估严谨。

一般而言，用于评估个体的程序越多，获得关于被测者的一个清晰的图像的可能性越大。因此，当要做的决定事关来访者一生的重要事情时，经常考虑使用多种评估程序 (Halperin & McKay, 1998; Joint Committee on Testing Practices, 2004; Meyer et al., 2001)。本书中，我们将考察一系列正式和非正式评估程序，所有这些评估程序都可以用于决策



图 1-1 评估程序

过程。首先，让我们从源头开始，看看过去的事件如何引领我们走向当前使用的评估工具。

提示

◆ 通常要考虑使用多种评估程序。

Sun, 2002)。该测验在数百个小隔间和小屋里进行，选拔过程繁重累人。在考试的强烈压力之下，甚至有考生过劳而死（Cohen & Swerdlik, 1999）。这项测验最终于1905年被废止。另一个古代评估的例子出现在柏拉图（Plato, 公元前428—公元前327）的著作中，希腊人筛选国家公务员时，要评估人们的智力和体力（Doyle, 1974）。

1.2 评估的历史

评估的历史可以追溯到古代，但现代评估产生于20世纪早期。让我们先来考察一下评估领域在过去几个世纪发生的一些变化。

1.2.1 古代历史

评估存在了多久？我们可以说当上帝要求亚伯拉罕（Abraham）杀死他的儿子以撒（Isaac）时，他的信心就得到了评估。毫无疑问，非正式评估远早于亚伯拉罕而存在。中国政府常常被称赞于大约公元前2200年开发了第一个广泛应用的测验，以评估个体是否适合政府工作岗位（DuBois, 1970; Higgins &

1.2.2 现代测验发展的先驱

随着19世纪实验研究和对照研究在科学领域的广泛应用，医生和哲学家开始将这些研究原理用于人身上，尤其是认知功能领域。例如，法国医生让·埃斯基罗尔（Jean Esquirol, 1772—1840）在法国精神病医院研究智力迟钝患者，使用语言来鉴定病人的智力水平（Zusne, 1984; Drummond, 2006）。当时智力迟钝患者被叫做“白痴”（Esquirol, 1838, p. 76），直到埃斯基罗尔认为智力迟钝是一种疾病，他的工作通常被认为是后来称为言语智力评估的开端。同一时期，另一位

法国人埃都阿德·谢根 (Edouard Sequin, 1812—1880) 也研究智力迟钝患者。他认为如果孩子的智力缺陷与生理问题并存, 其预后更差。他认为医生应该“留意活蹦乱跳的、双手‘无意识的繁忙’的、嘴角‘无意识地’流出唾液的、外表‘光泽亮丽而无实质内容’的、话语‘有限’或‘重复’的孩子” (Zenderland, 1987, p. 54)。最终, 谢根开发了增强其病人运动控制和感觉辨别力的拼图板。有人认为谢根是“操作智商” (performance IQ) 测量的先驱 (DuBois, 1970), 其开发的拼图板直到今天仍应用于操作导向的智力测验中。

提示

- ◆ 让·埃斯基罗尔: 提出利用言语来判断智力。他是“言语智商”的先驱。
- ◆ 埃都阿德·谢根: 开发了增强运动控制的拼图板, 他是“操作智商”的先驱。
- ◆ 弗朗西斯·高尔顿: 考察了感觉运动反应与智力间的关系。

同时, 受查尔斯·达尔文 (Charles Darwin, 1809—1882) 进化论的影响, 19 世纪中期科学家专注于理解人类物种的发展 (Juve, 2008; Kerr, 2008)。例如, 达尔文的表弟弗朗西斯·高尔顿 (Sir Francis Galton, 1822—1911) 对人与人之间的差异十分着迷。高尔顿假设反应时间更快、理解能力更强的个体拥有更高的智商。在好奇心的驱使下, 他检验了这些个性特征之间的关系, 他的研究推动了其他人发展出一个新的统计学概念——相关系数, 用以描述变量之间关系的强度 (Dubois, 1970; Kerr)。计算相关系数已经成为测验发展和完善的一项重要工具。

威廉·冯特 (Wilhelm Wundt, 1832—1920) 是另一个对人性着迷的科学家, 创建了他称为生理心理学的“科学新领域”。1875 年前后, 冯特在莱比锡大学创建了第一个用于实验研究的心理学实验室 (Nicolas, Gyselinck, Murray, & Bandomir, 2002)。该实

验室中, 大量实验研究了听觉、视觉和其他感官对刺激的反应时间 (Watson, 1968)。很多与冯特一起工作的学生促进了生理心理学新时代的兴盛。例如, 詹姆斯·麦基恩·卡特尔 (James McKeen Cattell, 1860—1944) 是冯特的一位博士生, 后来受到高尔顿的启发, 成为最早使用统计学概念来理解个人的心理学家之一 (Goodwin, 2008; Roback, 1961)。卡特尔的研究重点是测验智力功能, 他以创造了心理测验一词而著名。另一个著名人物 G. S. 霍尔 (G. S. Hall, 1844—1924) 也和冯特一起工作, 并最终在约翰·霍普金斯大学建立了自己的实验室。霍尔是很多其他伟大的美国心理学家的导师, 是美国心理学会的创建者, 并于 1892 年担任美国心理学会首任主席 (Benjamin, 2008)。

提示

- ◆ 威廉·冯特: 创办了第一个心理实验室。
- ◆ 詹姆斯·卡特尔: 将统计学引入心理测验, 创造了“心理测验”一词。
- ◆ G. S. 霍尔: 早期实验心理学家, 美国心理学会首任主席。

1.2.3 能力测验的出现

受到理解人性的新科学方法的影响, 20 世纪初研究者着手开发能够科学测量个体能力的工具。这一时代见证了能力测验的出现, 包括个体智力测验、神经心理评估和团体能力测验。

1. 智力与认知功能: 智力测验与神经心理评估

1904 年, 巴黎教育局找到阿尔弗雷德·比奈 (Alfred Binet, 1857—1911), 期望他能开发出一种可以帮助智力逊常儿童进入合适的学校学习的智力测验。比奈和他的同事泰奥菲·西蒙 (Theophile Simon) 激烈地批评一些将儿童诊断为智力缺陷者的做法, 他们共同开发了一种个别测验式量表, 该量表通过评估儿童对一系列不同任务 (例如, 跟踪光