

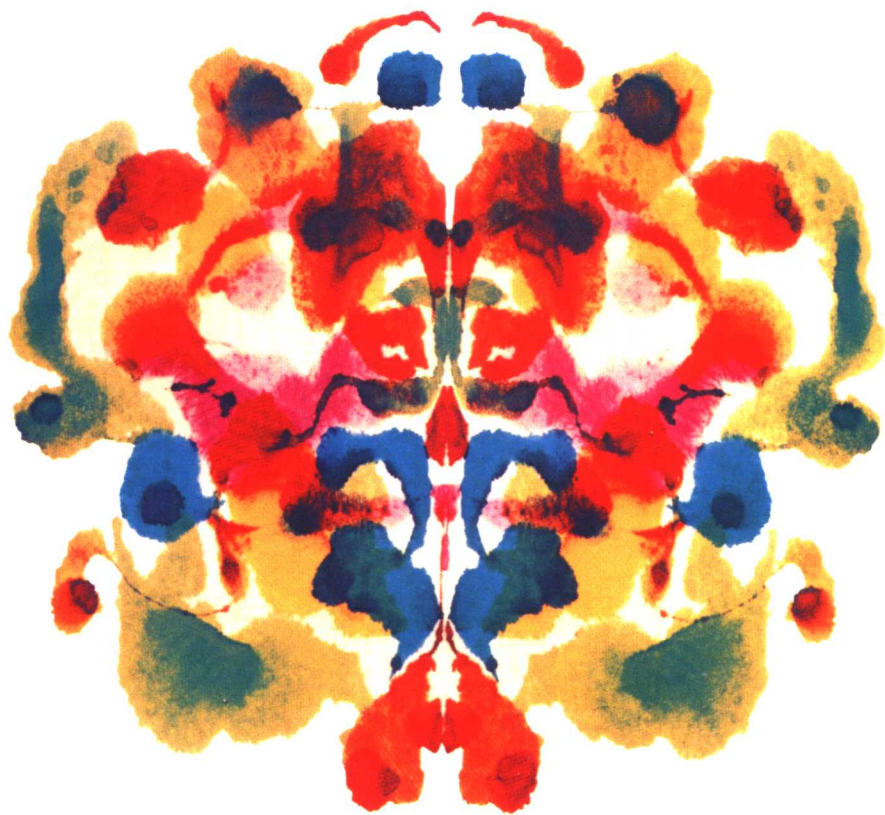
心理测验

原理和应用

第6版

Psychological Testing: Principles and Applications

[美] 凯温·R. 墨菲 查尔斯·O. 大卫夏弗 著 张 娜 杨艳苏 徐爱华 译



Kevin R. Murphy ◆ Charles O. Davidshofer

上海社会科学院出版社

Psychological Testing:
Principles and Applications

心理测验

原理和应用

第6版

[美] 凯温·R. 墨菲 查尔斯·O. 大卫夏弗 著
张 娜 杨艳苏 徐爱华 译

Kevin R. Murphy (宾夕法尼亚州州立大学)
Charles O. Davidshofer (科罗拉多州州立大学)



上海社会科学院出版社

图书在版编目(CIP)数据

心理测验:原理和应用/[美]墨菲,[美]大卫夏弗
著;张娜等译. —上海:上海社会科学院出版社,2006
ISBN 7-80681-811-1

I. 心... II. ①墨... ②大... ③张... III. 心理测验 IV. B841.7

中国版本图书馆 CIP 数据核字(2006)第 067594 号

Authorized translation from the English language edition, entitled PSYCHOLOGICAL TESTING; PRINCIPLES AND APPLICATIONS, 6th Edition, 0131891723 by MURPHY, KEVIN R.; DAVIDSHOFER, CHARLES O., published by Pearson Education, Inc., publishing as Prentice Hall, Copyright © 2005, 2001, 1998, 1994, 1991 by Pearson Education, Inc., Upper Saddle River, New Jersey, 07458.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

CHINESE SIMPLIFIED language edition published by PEARSON EDUCATION ASIA LTD., and SHANGHAI ACADEMY OF SOCIAL SCIENCES PRESS Copyright © 2006.
本书封面贴有 Pearson Education(培生教育出版集团)激光防伪标签。无标签者不得销售。

上海市版权局著作权合同登记号 图字:09-2005-590 号

心理测验:原理和应用

作 者:[美]凯温·R. 墨菲、查尔斯·O. 大卫夏弗

出 品 人:朱金元

责任编辑:唐云松

封面设计:闵 敏

图书策划:德瑞姆图书工作室

组稿编辑:周 杨

出版发行:上海社会科学院出版社

上海淮海中路 622 弄 7 号 电话 63875741 邮编 200020

<http://www.sassp.com> E-mail:sassp@sass.org.cn

经 销:新华书店

印 刷:上海长阳印刷厂

开 本:787×1092 1/16

印 张:33.5

插 页:3

字 数:700 千字

版 次:2006 年 8 月第 1 版 2006 年 8 月第 1 次印刷

ISBN 7-80681-811-1/B·018 定价:58.00 元

版权所有 翻印必究

出 品 人：朱金元

组稿编辑：周 杨

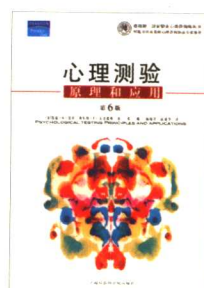
美术编辑：闵 敏

责任编辑：唐云松

责任编辑信箱：rmcbst@yahoo.com.cn

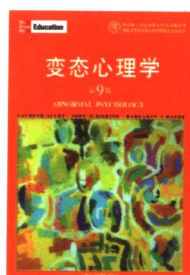
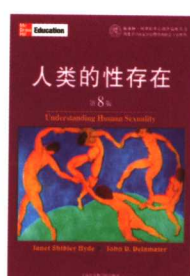
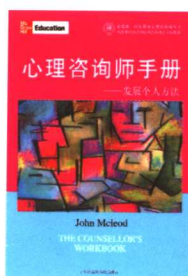
德瑞姆·国家职业心理咨询师丛书

第二辑



德瑞姆·国家职业心理咨询师丛书

第一辑



序 言

测验是用来做出决定的。这个简单的句子说明了本书的主题和基本原理。通常,心理测验都被介绍为一门枯燥的、技术性的和抽象的学科,其实不然。心理测验对各种类型的重要决定有着重要影响。在某些情境下,比如军中服役,测验是进行选拔和分类选择的唯一可行的方法。从这点来看,心理测验无论是好是坏,它都影响着人们的生活。在必须做出许多有关个人的重要决定(比如大学录取、工作安置、临床评估)的情境下,需要对心理测验有着深入的了解。

在一个测验课程中,大多数学生已经参加了许多测验而且在将来会参加更多的测验。这些学生对测验有着真正的、实际的兴趣。他们也可能怀疑心理测验的价值和精确性。当你教授一门测验课程的时候,你的大部分工作是对心理测验的优缺点进行充分且公正的评价。我们相信,通过关注测验对决策的影响,这能被很好地完成。在学习测验名称的细目清单时,学生通常表现出很少的兴趣,但是,当他们知道测验如何进行操作、使用测验与不使用测验相比会达到更好还是更坏的决策的时候,他们会非常感兴趣。

本书分为四大部分,在第一部分(第一章到第三章)中,我们介绍心理测验的概念,讨论测验对社会的影响。第二部分(第四章到第十章)讨论心理测量的原理和分析测验使用的技术。第三部分(十一章到十七章)讨论测验的发展,并特别关注了认知能力、兴趣和人格等领域。第四部分(十八章到二十一章)讨论了心理测验在做出有关个人重要决定时的运用。

本书存在其他几本测验书所没有的特征。第一,我们没有引入许多晦涩的测验。相反,我们关注主要类型测验中那些被广泛使用并且专业浅显的例子。在本书中,你不会碰到在心理测量课堂之外从未听到的许多测验,坦白地说,我们找不到任何你应该听到这些测验的理由。第二,我们没有引入几个虽然重要但与决策制定关联较小的主题。因此,我们不会对诸如态度测量或价值测量等主题进行详细论述。

本书包含了某些我们认为是有特点的而且对学生和教师都有用处的特征。第一,每一章都包含了一个或更多的标题为“关键性讨论”的简要部分。关键性讨论部分介绍了引发争议的各种类型的问题(例如,“黑人应试者的智力分数应该基于白人常模吗?”见

第五章),或者举例说明对该章中所介绍的概念的运用(例如,“运用项目反应理论检测测验偏向”,见第十章),或者提供对一个熟悉材料的不同观点(例如,“申请者眼中的人事选拔”,见第十九章)。这些关键性讨论或许能够为进一步的课堂讨论、学期论文或计划提供思想火花,如此设计是为了帮助学生将该课程的材料与他们其他的兴趣和知识进行整合。

第二,本书的第二部分到第四部分(也就是第四章到第二十一章)包括了“问题及解答”,我们在这些章节的开始提出与该章中介绍材料相关的问题(例如,“确定测验的潜在益处”,见第九章;“评价的不一致意味着什么?”见第二十一章),在每一章的末尾,我们重新提出这个问题并揭示一个潜在的解决办法。这个问题包括真正的测验使用者所面临的各种疑问和问题,还包括一个解决办法,它告诉学生如何使用每一章的材料解决一个真正而且重要的问题。

当你通读本书,你会逐渐注意到,总体而言,我们对心理测验抱有乐观的态度。在有些领域中,技术跟不上理论的发展,基本理论存在缺陷,或者测验运用弊大于利。然而从总体上来说,心理测验通常会为制定重要决定提供最公平和最精确的方法。这个结果部分地反映出测验的优点,也部分地反映出竞争的缺点(例如,面试、推荐信和临床直觉)。无论如何,我们相信,心理测验的确有所贡献,而且很可能会伴随我们很长时间。我们希望,这本书能够帮助你的学生更好地了解心理测验的优势和缺点。

在准备这个版本的时候我们做了许多改变。首先,我们讨论了几个现在广泛运用的测验版本,这些测验包括斯坦福-比奈测验、韦克斯勒儿童智力量表、韦克斯勒学龄前和学龄初期儿童智力量表和认知能力测验。最后,我们增加了从本书上个版本面世后出现的大量研究参考文献。

感谢

本书如果没有得到我们许多同事的帮助和鼓励是不会出版的。我们特别感谢弗兰·克兰蒂,他鼓励我们写出第一个版本,并向我们建议了这本书的大体框架。我们感谢丹尼斯·罗伯茨,他在本书的早期阶段与我们合作,对本书的大体框架和关注点做出了巨大的贡献。我们也感谢那些评论家,他们在本书每个版本的发展过程中向我们提供了非常有价值的反馈,他们是丹尼斯·斯文尼,宾夕法尼亚州加利福尼亚大学;K. 伊莱恩·饶尔,MTSU; Jo. 安·李,夏洛特北卡罗莱纳州大学。最后,我们感谢那些帮助我们形成观点和表述的同事,以及那些在最初的四个版本中给我们指出错误和疏漏之处的同事。他们包括金尼特·克利夫兰、乔治·桑顿、库尔特·格兴阁和那些向我们提供有用建议的众多师生。我们也向为我们挑出错误、提供其他的研究解释方式和在我们准备这个最新版本时向我们提供改善建议的几个同事表示感激。

凯温·R. 墨非

查尔斯·O. 大卫夏弗

目 录 *Contents*

序 言	1
-----	---

第一部分 心理测验的绪论

1 测验和测量	1
心理测验——一个定义	2
测验的类型	5
测验和决策——心理测验的使用	7
心理学家的测验活动	9
有关测验的资料	10
测验的标准	11
关键性讨论:与心理测验相关的伦理原则	14
关键性讨论:心理测验的其他选择	15
总结	16
关键术语	16
2 定义和测量心理属性:能力、兴趣和人格	17
心理属性和决策	17
智力——一般的心理能力	18
关键性讨论:智商还是情商?	31
兴趣	32
人格	35
关键性讨论:能力、兴趣和人格特征之间的关系	41
总结	42
关键术语	43
3 测验和社会	44
决策类型	44
社会担忧	46

关键性讨论:在能力—测验争论中的科学和政治——贝尔曲线 50

关键性讨论:高赌注测验 54

关键性讨论:西里尔·波特先生的奇怪案例 57

分析测验的社会影响 57

关键性讨论:对社会问题是否有技术性解决办法? 60

总结 60

关键术语 61

第二部分 心理测验的原理

4 测量和统计中的基本概念 62

问题及解答——预测未来的表现 62

心理测量 62

评价心理测验 67

统计概念 69

关键性讨论:测量量表和统计 76

总结 77

问题及解答——简要回顾 78

关键术语 78

5 量表、转化和常模 80

问题及解答——比较量表 80

转化 81

平衡量表分数 85

常模 88

期望表 94

关键性讨论:黑人应试者的智商分数应该基于白人常模吗? 97

总结 99

问题及解答——简要回顾 99

关键术语 100

6 信度:测验分数的一致性 102

问题及解答——评价一个测验 103

测验分数一致性和不一致性的来源 103

- 信度的一般模型 105
- 估计信度的简单方法 107
- 信度估计和误差 112
- 测验分数的概化 113
- 关键性讨论:概化和信度理论 115
- 总结 116
- 问题及解答——简要回顾 117
- 关键术语 117
-
- 7 使用 and 解释关于测验信度的信息 119
 - 问题及解答——你应该重新进行这个测验吗? 119
 - 使用信度系数 120
 - 影响信度的因素 124
 - 信度中的特殊问题 127
 - 测验应该多可靠? 132
 - 关键性讨论:精确但狭隘的——带宽—保真度困境 134
 - 总结 134
 - 问题及解答——简要回顾 135
 - 关键术语 136
-
- 8 测量的效度:内容和构念导向的验证策略 137
 - 问题及解答——这个测验测量什么?
 - 验证策略 138
 - 评估测量的效度 139
 - 以内容为导向的验证策略 140
 - 关键性讨论:表面效度 145
 - 以构念为导向的验证策略 146
 - 内容效度和构念效度 154
 - 关键性讨论:我们如何知道智力测验是否真正测量了智力? 155
 - 总结 157
 - 问题及解答——简要回顾 157
 - 关键术语 158
-
- 9 决策效度:效标关联效度 159

问题及解答——确定测验的潜在益处	159
决策和预测	160
准则	161
效标关联验证策略	161
构念效度和效标关联效度	166
解释效度系数	168
关键性讨论:价值和验证	176
总结	177
问题及解答——简要回顾	177
关键术语	178
10 项目分析	180
问题及解答——你应该放弃哪个项目?	180
项目分析的目的	180
迷惑项分析	181
项目难度	183
项目鉴别力	185
项目特征之间的互动	187
项目反应理论	189
关键性讨论:运用项目反应理论检测测验偏向	196
总结	197
问题及解答——简要回顾	197
关键术语	198
第三部分 能力、兴趣和人格测量的发展	
11 测验发展的过程	200
问题及解答——你应该相信自己的判断还是测验?	200
编制测验	201
关键性讨论:如何正确地侦查反应定势?	208
关键性讨论:不同的测验形式只能提供一种解释信息吗?	210
常模化和标准化测验	211
测验的出版和修订	214
关键性讨论:应该建立测验的全国性、地域性或本地性的常模的?	215

总结 216

问题及解答——简要回顾 218

关键术语 219

12 测验实施和解释的计算机化 220

问题及解答——哪种类型的测验最适合于你? 220

在测验中使用计算机:一种分类法 221

关键性讨论:因特网上的测验信息 222

计算机化的测验实施 223

计算机化的自适应测验 225

基于计算机的测验解释 228

检测无效的或不寻常的反应类型 231

关键性讨论:巴纳姆效应和计算机化的测验解释 232

总结 233

问题及解答——简要回顾 233

关键术语 234

13 能力测试:个别测验 235

问题及解答——使用测验来诊断阅读缺陷 235

测验者的作用 236

一般心理能力的个别测验 237

特殊能力的个别测验 252

关键性讨论:对来自于不同种族群体的被测者进行测试 253

总结 254

问题及解答——简要回顾 255

关键术语 256

14 能力测试:团体测验 257

问题及解答——这个测验对你来说合适吗? 258

团体测验的优势和劣势 258

一般认知能力的测量 259

学业测验 263

关键性讨论:大规模的大学入学测验是如何开始的? 268

多元能力倾向成套测验 274

关键性讨论:反对大规模测验的情形 280

总结 281

问题及解答——简要回顾 282

关键术语 282

15 能力测验中存在的问题 283

问题及解答——测验中存在公平和歧视吗? 284

心理测验中的偏向和公平性问题 284

关键性讨论:判断项目偏向有意义吗? 296

文化简化测验 296

关键性讨论:智力和适应行为 304

智商的遗传性、一致性和可变性 304

关键性讨论:智力和社会阶层 308

关键性讨论:启智方案 313

总结 314

问题及解答——简要回顾 315

关键术语 316

16 兴趣测验 317

问题及解答——测验和职业咨询 317

斯特朗兴趣量表 318

库德兴趣量表 332

关键性讨论:一种新颖的兴趣测量方法:库德职业搜索与个人匹配 340

职业评价问卷 342

杰克逊职业兴趣调查表 346

关键性讨论:在人事选拔中应该使用兴趣测验吗? 352

总结 353

问题及解答——简要回顾 354

关键术语 355

17 人格测验 356

问题及解答——人格测验能够帮助做出选拔决策吗? 356

人格测验的发展 357

人格的客观测量 358

关键性讨论: CPI 是否被用来测定人们的类型?	362
人格投射测量	370
关键性讨论: 可以对人格测验的各分量表进行独立解释吗?	378
关键性讨论: 投射测验、人格量表和智力测验的信度和效度	379
总结	380
问题及解答——简要回顾	381
关键术语	383

第四部分 利用测验制定决策



18 测验与教育决策	384
问题及解答——你应如何解释一个低于平均分的分数?	384
在教育机构决策中的应用	385
关键性讨论: 全国教育发展评估	390
在个体决策中的运用	391
关键性讨论: 什么是公文包测验? 它可以用于教育评估吗?	394
关键性讨论: 什么是标准参照测验?	395
总结	396
问题及解答——简要回顾	397
关键术语	398
19 工业领域中的心理测量	399
问题及解答——对一个大的申请者群体的筛选	399
人事测验的范围和影响	400
绩效预测	401
对人事测验的比较性评估	415
关键性讨论: 笔迹学能预测职业成就吗?	416
人事测验的效度	417
雇用测验的社会及法律背景	420
关键性讨论: 申请者眼中的人事选拔	466
诚实评估	423
工作绩效及有效性的测量	425
评判测量	428
总结	437

问题及解答——简要回顾 439

关键术语 439

20 诊断测验：在临床中的运用 441

问题及解答——对分测验成绩的比较 441

明尼苏达多相人格调查表 442

MMPI-2 448

本德尔—格式塔测验 449

离散分析 452

诊断分类系统 455

关键性讨论：职业道德与临床测验 459

神经心理评价 460

总结 465

问题及解答——简要回顾 466

关键术语 466

21 多元法评估中的问题 468

问题及解答——评估的不一致意味着什么？ 468

临床评价的实质 469

对临床判断的研究 473

结构化评价程序 480

关键性讨论：法庭中的心理评估 489

总结 490

问题及解答——简要回顾 491

关键术语 492

附录 A 四十个代表性测验 493

附录 B 心理学家的道德准则和行为规范 495

主题索引 510

第一章 测验和测量

心理测验(psychological test)这个术语使我们想起很多相冲突的形象。一方面,这个术语会使一个人想起经常在电视、电影和大众文学上出现的各种类型的测验,在那里,一个病人会回答诸如“你恨你的母亲多长时间了?”等问题,这样做可以揭示出其人格中没有向临床医生透露的方面。另一方面,心理测验可能涉及很长的多项选择题,这些选择题类似于成千上万的高中生在参加大学入学考试时回答的题目。另一种类型的心理测验是在《读者文摘》上刊登的自我测验,它声称可以告诉你的婚姻是否处在危险的边缘、你是否与你身旁同事一样感到焦虑,或者你是否应该转换工作或你的生活方式。

一般而言,心理测验既像我们的第一个例子所展示的那样并不神秘,也像我们的最后一个例子所展示的那样并不零乱。心理测验更多是系统应用一些相关的简单原则,它试图测量出个人的特征,这些特征在描述或理解个人行为的时候应该是很重要的。本书的目的就是描述出心理测量(psychological measurement)的基本原理和测验的主要类型及其应用。我们不会机械地详细描述测验理论,也不会叙述(甚至不会提及)现在可以获得的所有不同的心理测验。相反的,我们的目标是提供一些信息,使得我们能够利用这些信息对在教育、工业和临床实践等领域中的心理测验及其应用进行可行的评估。

在一个心理测验文本中我们首先应该提出的第一个问题是:“心理测验为什么重要?”对于这个问题可能有好几种答案,但是,我们认为,最好的答案存在于组成本书核心主题的简单陈述中:测验是用来帮助做出有关个人重要决定的。大学的招生官员在决定是录取还是拒绝申请者之前会参考测验分数。临床心理学家在为一个人选择治疗方案的过程中会参考各种类型的客观测验和投射测验。军事部门在决定一个士兵能够胜任任何种工作的时候会将测验分数作为一种辅助手段。测验被运用到人事选择领域、职业合格证和许可证领域等工作世界中。几乎每个读本书的人都至少参加过一次标准化心理测验。这种测验的分数或许已经在你做出你生活重要决定时产生了某些影响。因此,心理测验是一个具有很强实践性价值的领域。

心理测验用来测量各种类型的特质——智力、动机、七年级数学的掌握程度、职业偏好、空间能力、焦虑、形状知觉以及无数的其他特质。不幸的是,所有心理测验共有的一

个特征是精确度有限。它们很少能够为对人类行为产生重要影响的各种变量提供精准的、明确的测量。因而,心理测验并不对个人做出精确的判定提供一种基础。事实上,没有一种方法能够保证完全的精确。因此,虽然心理测验并不完美,但是,国家科学院的一个小组得出这样的结论:心理测验是为对个人做出可行的判定提供必要信息的最好、最公正也是最经济的方法(Wigdor & Garner, 1982a, 1982b)。国家研究员小组做出的这个结论形成了贯穿本书的另一个重要的主题。因此,虽然心理测验并不完美,但是心理测验是对个人做出重要判定的最好、最公正也是最准确的方法。

心理测验具有很大的争议性。自20世纪20年代以来,公众对测验特别是对标准智力测验的使用争论就非常激烈(Cronbach, 1975; Haney, 1981; Scarr, 1989)。①大量的文献,既有普及性的,又有专业性的,在处理诸如测验偏差和测验公正等问题。已经通过的联邦法案和州立法案号召使用最少的能力测验,要求测验具有现实性,有些条款谈及要付出各种努力使测验具有规则,增加公众对测验的发展和应用的认可度。测验和测验程序在法庭上已经受到了挑战,这种挑战往往是成功的。

心理测验不仅是重要的和具有争议性的,而且它还是具有高度专业化和有些技术化的事业。在许多自然科学中,测量是一个相对直接的过程,包括评估客体的物理特征,例如高度、重量和速度。②然而,很大程度上,诸如智力和创造性等心理属性并不能使用与测量物理属性相同的方法来进行测量。心理属性在简单的、物理的方面并不明显,它们只在个人行为方面是明显的。此外,行为很少只反映心理属性,它还反映生理的、心理的和社会方面的属性。因此,心理测量不会像物理测量那样简单和直接。合理地评价心理测验需要熟悉心理测量的专业化方法。

本章对心理测量进行一般性的介绍。首先,我们界定测验这个术语,讨论一下该定义的几种含义。然后,我们简要描述一下可得到的各种类型的测验,讨论一下在教育、工业和临床情景下做出决定时采用的测验方式。我们也会讨论一下有关测验的资料来源以及测验的标准、道德规范、法规。

心理测验——一个定义

心理测验的多样化令人眼花缭乱。几千个不同类型的心理测验可以在说英语的国家中通过商业途径获得,而且,毫无疑问的,许许多多其他类型的心理测验也在世界其他地方出版。这些测验从人格测验到自测智力测验,从学校考试到知觉测验,各种各

① 《美国心理学家》杂志中的特殊问题一栏在1965年11月和1981年12月刊登了对本次争论中众多问题的精彩概括。

② 注意,物理测量并非静态的也非简单的。有人建议根据光从一点到另一点经过的时间重新界定长度的基本单位,如米(Robinson, 1983),这种建议是在重新界定物理测量基础方面不断进步的一种体现。