

MEASUREMENT AND EVALUATION  
IN PSYCHOLOGY AND EDUCATION

Eighth Edition

# 教育评价

教育和心理学中的测量与评估

第八版

[美] Robert M. Thorndike 著  
Tracy Thorndike-Christ 著  
方群 吴瑞芬 陈志新 译



Pearson



1917

商务印书馆  
The Commercial Press

MEASUREMENT AND EVALUATION  
IN PSYCHOLOGY AND EDUCATION

Eighth Edition

# 教育评价

教育和心理学中的测量与评估

第八版

[美] Robert M. Thorndike 著  
Tracy Thorndike-Christ 译  
方群 吴瑞芬 陈志新

 商务印书馆  
The Commercial Press

2018年·北京

图书在版编目(CIP)数据

教育评价:教育和心理学中的测量与评估:第八版/  
(美)罗伯特·M.桑代克,(美)特雷西·桑代克-克莱斯特著;方群,吴瑞芬,陈志新译.—北京:商务印书馆,  
2018

ISBN 978-7-100-15882-4

I. ①教… II. ①罗… ②特… ③方… ④吴…  
⑤陈… III. ①教育评估 IV. ①G40-058.1

中国版本图书馆 CIP 数据核字(2018)第 036296 号

权利保留,侵权必究。

教育评价

——教育和心理学中的测量与评估

(第八版)

[美]罗伯特·M.桑代克 著  
[美]特雷西·桑代克-克莱斯特  
方群 吴瑞芬 陈志新 译

---

商务印书馆出版

(北京王府井大街36号 邮政编码100710)

商务印书馆发行

北京市艺辉印刷有限公司印刷

ISBN 978-7-100-15882-4

---

2018年10月第1版

开本 787×1092 1/16

2018年10月北京第1次印刷

印张 34½

定价:88.00元

Authorized translation from the English language edition, entitled MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION, 8th Edition, by THORNDIKE, ROBERT M.; THORNDIKE-CHRIST, TRACY M., published by Pearson Education, Inc., Copyright © 2010, 2005, 1997 Pearson Education, Inc., publishing as Merrill.

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 THE COMMERCIAL PRESS LTD., Copyright © 2018

本书简体中文版由培生集团授权商务印书馆独家出版发行, 未经出版者书面许可, 不得以任何方式复制使用。权利保留, 侵权必究!

# 前 言

本书第一版于 1955 年出版时,我才 11 岁,对这本书的贡献也很有限,尽管我父亲需要用他的测试做试验时,我的确充当过他的“小白鼠”。接着出版的第二版和第三版中,我对该书还是没有实质性的内容输入,但 1970 年的时候,我开始参照第三版的内容对我自己班上的学生进行测量。我刚对心理测量学知识有了一定程度的了解,我父亲罗伯特·L. 桑代克就常跟我讨论有关心理测量的诸多问题,因此我对第三版中内容的一些意见和看法对第四版中的内容也有一些间接的影响。第四版是在我父亲退休后出版的。到第五版的时候,我接过了父亲肩上的大部分担子,虽然他仍然在校订过程中起着十分重要的作用。乔治·坎宁安说伊丽莎白·哈根对第五版的贡献尤其显著,但对于前面这五个版本的编写,她都为我们提供了许多可靠的教育素材和资源。到了第六版,我让我的女儿特雷西·桑代克-克莱斯特也参与了进来。她是一位教育心理学家,主要为我们提供一些测试的应用材料。有关临床试验的想法是马克斯阿琳·路易斯提出的,他们为我们校订了有关性格测试那一章。此外,还有里克·伊顿巴赫帮忙校订了有关测试的特殊情形和评估中的专业性和合法性问题这两章。马克斯和阿琳也为第七版和第八版的顺利出版出了许多力,这两版是我和我女儿特雷西合著的,而关于这本书后续版本修订的担子也将会落在她的身上。我父亲去世已经十八年了,但他对这本书的影响仍然显而易见,不论是其中一些仍留有他独特文字魅力的内容阐述上,还是他对我的思维方式和专业知识的积累方面潜移默化的影响。谢谢你,爸爸!

罗伯特·M. 桑代克博士  
荣誉教授

在本书第一版出版后的五年多里,有关教育心理测量领域的理论和实践都发生了翻天覆地的变化。每一个体,无论其肤色、性别、残疾与否或任何其他方面的特征差异,其权益是否得到了保障引起社会各界的广泛关注,以致有人希望并开始尝试为

每一个孩子提供最佳的教育体验。还必须保证整个社会内的每一个人都具有接受高等教育和求取理想职位的平等机会。但即使发生了如此大的变化(或许也正是因为这些变化),人们对客观公正及具有较高可靠性和相关度信息的渴求依然强烈,这些信息主要用于进行相关决定时的参考。在1955年,这些测试能够为我们提供诸多有用信息,并且成为了二十一世纪初一个更加重要的信息来源。

这本书的前七个版本帮助了三代人,教育家和心理学家们,学会巧妙应用测试信息的一些方法和技巧;而这第八版也是本着这一目的出版的。第八版延续了将测试作为获取信息来源的传统,这些信息可以有效辅助我们做出相关决定。测试存在的意义即帮助人们做决定,且这些决定的质量取决于对这些信息的充分解析和评估,这是我们一以贯之的重点。测试本身设计不够完善会导致测量过程中出现许多问题,但即便是精心设计的测试,如果使用不当,也一样会出现很多问题。我们旨在为测试使用者提供所需的相关信息和使用技巧,以免在应用测试和解析数据时出错。

## 第八版中的变化

在该版本中,我们对所有与教育决策有关的章节内容进行了彻底的更新,以相应地提高书中相关内容对美国教育体系中的标准、评估程序和问责制的重视。美国人口的文化和语言的丰富多样性仍在不断提高,评估教育体系中英语学习者能力水平的需要也在日渐增强。因此,我们在书中也相应增加了一些内容,主要关于英语水平、测量一般英语水平对象的学习能力及对上述提到的针对特殊群体中英语语言学习者的特殊教育评估。在所有有关教育决策的探讨中,我们都提高了对标准参照、基于课程和更真实的评估开展教学的重视。

本书第八版的内容由三大部分变成两大主要部分。和前面的几个版本一样,第一部分包括第一章到第六章,主要是有关测试中一些技术问题的总体概述,任何要根据测试信息做决定的人都应该对此部分的内容有一个基本的了解。其中第一章是一些基本理论,接下来的一章关于一些必不可少的定量研究概念,还有一章主要介绍参照系(常模参照、自比参照、标准参照及潜在倾向),接下来的内容有关测量程序的可靠性和有效性,最后一章是对一些实际操作中的问题的总结,包括如何找到有关一些特定测试的信息材料等。第二部分从第七章到第十四章,主要处理具体测试操作中与测量程序相关的问题。第七章涉及教育领域内特定类型的决定和做出该类型的决定所需的特定类型的信息。第八章是第八版中整个重新改写过的一章,阐述一些针对特殊人群的评估程序中有关心理测量、合法性和道德伦理的问题,包括残障学习者

和非英语母语背景的英语语言学习者。第九章和第十章均与传统的测试、表现和质量评估有关,这些测量信息或许对做出某些非统一的或非标准化的教育决策具有一定的参考价值。第十一章介绍了一些测量个人态度及制作评分量表和检查表的方法。第十二章关于能力倾向测试。第十三章关于标准化的成绩测试,我们根据一些用于测量一般认知能力和学习技能的专业测试的最新修订版对这一章的内容也做了较大幅度的更改。最后,第十四章中主要是对一些关于测量兴趣爱好、性格特征和顺应调节的方法程序的回顾。

# 致 谢

在此,我们想感谢以下同事在我们完成这本书的过程中提出的诸多富有建设性和启发性的独到见解,他们是:内布拉斯加-林肯大学的詹姆斯·A. 伯瓦尔德;弗吉尼亚理工学院的罗伯特·J. 哈维;美国西东大学的玛丽埃塔·佩斯金;乔治梅森大学的查尔斯·L. 托马斯和得州理工大学的阿曼达·威廉姆斯;我由衷地感谢所有人,还有我的妻子埃尔娃。她前七次都参与了著书和出版的全过程,并自始至终对我予以支持和理解。

罗伯特·M. 桑代克

我要感谢我的父亲罗伯特·M. 桑代克,感谢他给我这个机会继续为传承了三代的家族“结晶”的成长贡献自己的绵薄之力。我由衷地感激他给予我的全部支持,不论是个人情感上的,还是专业知识和技能上的。他无比耐心,在写作本书过程中给予我充分理解。他是我生命中最重要启蒙导师。能够和他共事是一种恩赐,能够被信任和委以跟进后续几个版本的重任更让我倍感荣幸和自豪。

我还要感谢西华盛顿大学特殊教育系的同事,特别是詹妮·帕克。我们经常交流想法,经过热烈讨论,我在有关对有特殊需要的学生进行评估这方面的观点才逐渐明朗。

最后,我想感谢我的丈夫史蒂夫和儿子麦卡伦,感谢他们的支持和鼓励,以及对我经常长时间待在自己的工作室里,一门心思扑在这本书上的理解和包容。谢谢你们!

特雷西·桑代克-克莱斯特



# 目 录

## 第一部分 技术问题

第一章 测量中的基本问题 .....	3
一、引言 .....	3
二、历史简介 .....	4
1. 早期阶段 .....	5
2. 繁荣阶段 .....	6
3. 第一批判阶段 .....	7
4. 测试组合阶段 .....	7
5. 第二批判阶段 .....	8
6. 问责制时期 .....	8
三、决定的类型 .....	9
四、测试与决定 .....	10
做决定过程中的价值取向 .....	11
五、测量程序的具体步骤 .....	13
1. 特征识别与定性 .....	13
2. 确定隔离和表现特征的操作程序 .....	15
3. 属性的量化 .....	16
4. 测量过程中的问题 .....	18
六、一些测量中的当前问题 .....	19
1. 少数族裔个体测试 .....	20
2. 侵犯隐私 .....	21
3. 使用常模对照组 .....	22
4. 影响测试分数的其他因素 .....	23

5. 受试者的权利和责任 .....	23
总结 .....	24
习题 .....	24
推荐阅读 .....	26
<b>第二章 测量与数字</b> .....	<b>28</b>
一、如何理解测试分数 .....	28
二、测量量表 .....	31
三、频率分布表的准备 .....	33
1. 分组频率分布 .....	34
2. 累积频率分布 .....	39
3. 图示法 .....	40
四、集中趋势测量 .....	43
1. 众数 .....	43
2. 中值 .....	44
3. 百分位数值 .....	46
4. 算术平均值 .....	47
5. 集中趋势与分布形态 .....	49
五、差异性测量 .....	51
1. 全距 .....	51
2. 半四分位数间全距 .....	51
3. 标准差 .....	52
六、标准差分析 .....	55
七、个体分数分析 .....	57
八、相关性测试 .....	58
九、预测结果 .....	64
确定回归线 .....	66
总结 .....	69
习题 .....	70
推荐阅读 .....	72

第三章 赋予分数意义 .....	73
一、分数的本质 .....	73
1. 参照系 .....	74
2. 标准参照测试和常模参照测试的领域 .....	76
二、标准参照评估 .....	77
三、常模参照评估 .....	80
1. 年级常模 .....	82
2. 年龄常模 .....	86
3. 百分位数常模 .....	88
4. 标准分数常模 .....	94
5. 转换的正态分布 .....	99
6. 九分评分制 .....	101
四、不同类型常模的互换性 .....	103
五、商数 .....	106
六、分数组合 .....	107
七、标准参照分析报告 .....	112
八、学校平均水平常模 .....	115
九、常模使用注意事项 .....	116
十、第三参照系:项目反应理论 .....	119
总结 .....	126
习题 .....	127
推荐阅读 .....	128
第四章 测量程序应有特性:信度 .....	130
一、引言 .....	130
二、信度和一致性 .....	131
不一致性的来源 .....	132
三、表示信度的两种方法 .....	133
1. 标准测量误差 .....	133
2. 信度系数 .....	134

四、评估信度的方法 .....	135
1. 同材料重复测试 .....	135
2. 平行测试形式 .....	137
3. 单次测量方法 .....	138
4. 方法对比 .....	144
五、信度数据分析 .....	145
1. 标准测量误差 .....	145
2. 信度系数 .....	147
六、影响信度的因素 .....	148
1. 测量对象的差异性 .....	149
2. 测量对象的特征等级 .....	149
3. 测试长度 .....	151
4. 评估信度的方法 .....	152
5. 实际信度和理论信度 .....	153
七、最低信度 .....	154
八、差异分的信度 .....	156
九、不可靠性对变量之间相关系数的影响 .....	159
十、标准参照测试的信度 .....	160
十一、计算机自适应测试的信度 .....	165
总结 .....	169
习题 .....	170
推荐阅读 .....	171
<b>第五章 测量程序应有特性:效度</b> .....	<b>172</b>
一、引言 .....	172
二、内容相关效度证据 .....	173
1. 设计命题蓝图 .....	174
2. 测量能力倾向与典型表现的内容效度 .....	181
三、标准相关效度证据 .....	182
1. 表面效度 .....	182
2. 实证效度 .....	182

四、建构相关的效度证据 .....	194
1. 相关性的预测 .....	195
2. 对群体差异的预测 .....	197
3. 对有关实验或干预反应的预测 .....	197
五、效度的统一定义 .....	198
1. 效度验证作为一项科学追求 .....	199
2. 作为整体效度的建构效度 .....	200
3. 梅西克的扩大效度理论 .....	204
4. 对梅西克思想的修正及关注点的变化 .....	208
六、效度理论和测试偏见 .....	210
七、信度和效度的重合 .....	211
八、标准参照测试的效度 .....	212
九、元分析和效度的泛化 .....	213
总结 .....	214
习题 .....	214
推荐阅读 .....	215
<b>第六章 测试的实际应用问题 .....</b>	<b>217</b>
一、例行测试的使用中与实际应用相关的一些因素 .....	217
1. 经济性 .....	217
2. 有助测试实施的因素 .....	219
3. 有助解读和应用测试分数的因素 .....	220
4. 电子化测试 .....	222
二、测试评估指南 .....	223
1. 基本信息 .....	224
2. 关于测试的信息 .....	224
3. 解析测试结果的辅助方法 .....	224
4. 效度 .....	224
5. 信度 .....	225
6. 测试的实施与评分 .....	225
7. 量表和常模 .....	226

三、了解具体的测试 .....	226
1. 现存测试有哪些? .....	227
2. X 测试到底是什么样的? .....	229
3. 评审们如何评价 X 测试? .....	230
4. 前人对 X 测试做过哪些研究? .....	232
总结 .....	234
习题 .....	234
测试信息参考资料 .....	235

## 第二部分 测试的应用

第七章 教育决策与评估 .....	241
一、引言 .....	241
二、价值观和决策 .....	241
三、《有教无类法案》 .....	242
1. 《有教无类法案》总览 .....	242
2. 标准和评估 .....	244
3. 问责制 .....	245
四、人员安置决策 .....	246
1. 残障学生的主流化问题 .....	247
2. 人员安排决定是如何做出的 .....	248
五、课堂教学决策 .....	249
1. 使用教学目标 .....	250
2. 评估方法的类型 .....	250
六、日常教学决策 .....	253
七、汇报学业进展 .....	254
1. 相对于完美水平的表现 .....	255
2. 相对于同等水平的表现 .....	255
3. 相对于潜能水平的表现 .....	256
4. 评定成绩 .....	257
5. 成绩的重要性 .....	257

八、影响未来教育的决策 .....	258
1. 选拔性决策 .....	259
2. 高风险决策 .....	260
九、其他教育决策 .....	262
1. 有关课程设计的决策 .....	262
2. 公共决策和政治决策 .....	263
总结 .....	264
习题 .....	264
推荐阅读 .....	265
<b>第八章 评估特殊人群:心理测量、法律及道德问题 .....</b>	<b>267</b>
一、引言 .....	267
二、重大立法与诉讼概览 .....	268
1. 影响重大的立法 .....	268
2. 影响重大的诉讼 .....	269
三、特殊教育评估过程 .....	270
项目实施和评估程序的移交 .....	270
1. 识别和移交 .....	271
2. 资格认定 .....	271
3. 项目计划、实施和评估 .....	272
四、特殊教育评估涉及的主要领域 .....	273
1. 智力和认知能力 .....	273
2. 适应性行为和自理能力 .....	274
3. 行为和社会-情感能力 .....	275
4. 神经心理能力 .....	276
五、评估母语为非英语者 .....	277
1. 简介 .....	277
2. 语言水平评估 .....	278
3. 对母语为非英语者进行学业能力评估 .....	279
4. 对母语为非英语者进行特殊教育评估 .....	280
六、传统学业能力 .....	280

1. 阅读、数学和书面语评估 .....	280
2. 课程评估 .....	282
3. 生态评估 .....	283
七、专业标准和道德规范 .....	284
简介 .....	284
八、专业训练和专业能力 .....	285
1. 专业训练 .....	285
2. 专业能力 .....	286
九、专业责任和科学责任 .....	286
教育测试与心理测试的标准 .....	286
1. 测试结构、评估和记录 .....	287
2. 测试和公平 .....	287
3. 测试应用 .....	288
十、尊重他人的权利和尊严 .....	288
隐私和保密 .....	288
1. 谁会从收集到的信息中获益? .....	288
2. 怎么使用这些信息? .....	289
十一、社会责任 .....	289
1. 公平分配 .....	290
2. 测试的社会效益 .....	290
3. 积极影响的最大化 .....	291
总结 .....	293
习题 .....	294
推荐阅读 .....	294
<b>第九章 测试开发原则</b> .....	<b>296</b>
一、引言 .....	296
二、编写客观题的一些建议 .....	296
1. 客观题的一般原则 .....	296
2. 编写判断正误题 .....	300
3. 编写多项选择题 .....	305



4. 编写匹配题 .....	318
三、编制可用的客观测试 .....	320
四、客观测试的评分 .....	323
1. 猜题校正 .....	323
五、利用试题分析改进客观测试 .....	325
1. 简化的试题分析程序 .....	326
2. 更多正式试题分析程序 .....	328
六、编写论述题 .....	331
1. 编写论述问题 .....	332
2. 设计论述题测试 .....	334
3. 论述题的评分 .....	335
总结 .....	336
习题 .....	337
推荐阅读 .....	338
<b>第十章 表现评估和作品评估 .....</b>	<b>340</b>
一、引言 .....	340
二、传统认知测试的人为因素 .....	340
三、作品评估 .....	341
四、在认知任务中运用表现评估和作品评估 .....	342
表现评估的评分 .....	343
五、过程评估 .....	345
1. 使用检查表 .....	345
2. 使用定级量表 .....	346
六、评估作品及表现 .....	347
1. 多位观测者的优点 .....	348
2. 多位观测者的可靠性或一致性 .....	348
七、系统观测 .....	350
1. 进行系统观测 .....	351
2. 系统观测的优缺点 .....	354
总结 .....	357