

高级心理测量学丛书

汪文义 宋丽红 罗芬 熊建华◎著

# 智慧化测评 的理论与技术

ZHIHUIHUA CEPING  
DE LILUN YU JISHU



北京师范大学出版集团  
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP  
北京师范大学出版社

汪文义 宋丽红 罗芬 熊建华◎著

# 智慧化测评 的理论与技术

ZHIHUIHUA CEPING  
DE LILUN YU JISHU



北京师范大学出版集团  
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP  
北京师范大学出版社

---

图书在版编目(CIP)数据

智慧化测评的理论与技术/汪文义等著. —北京:北京师范大学出版社, 2019. 6

(高级心理测量学丛书)

ISBN 978-7-303-24684-7

I. ①智… II. ①汪… III. ①心理测量学 IV. ①B841.7

中国版本图书馆 CIP 数据核字(2019)第 080750 号

---

营 销 中 心 电 话 010-58805072 58807651

北师大出版社高等教育与学术著作分社 <http://xueda.bnup.com>

---

出版发行: 北京师范大学出版社 [www.bnup.com](http://www.bnup.com)

北京市海淀区新街口外大街 19 号

邮政编码: 100875

印 刷: 保定市中画美凯印刷有限公司

经 销: 全国新华书店

开 本: 787 mm×1092 mm 1/16

印 张: 17

字 数: 322 千字

版 次: 2019 年 6 月第 1 版

印 次: 2019 年 6 月第 1 次印刷

定 价: 75.00 元

---

策划编辑: 何 琳

责任编辑: 李云虎 葛子森

美术编辑: 李向昕

装帧设计: 金基渊

责任校对: 韩兆涛

责任印制: 马 洁

**版权所有 侵权必究**

反盗版、侵权举报电话: 010-58800697

北京读者服务部电话: 010-58808104

外埠邮购电话: 010-58808083

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-58805079

# 高级心理测量学丛书编委会

主 编：戴海琦 丁树良

编 委：（按音序排名）

蔡 艳 戴步云 董圣鸿 高椿雷

高旭亮 罗 芬 罗照盛 宋丽红

涂冬波 汪文义 熊建华 喻晓峰

郑蝉金

# 总序

心理与教育测量是评价个体心理特质发展水平状态的重要手段。以项目反应理论为代表的现代测量理论的发展，为指导心理与教育测量研究及实践提供了强大的理论与技术支持。在项目反应理论基础上的参数估计、等值、信息量评价、项目功能差异甄别等技术保证了测验开发更加科学。最近十几年蓬勃兴起的认知诊断评价理论，则将测量理论与技术推向了更加精细化的评价水平上。

不过，在国内的心理与教育测量实践中，大多数研究和实践仍然主要是基于经典测量理论基础上的。许多试图使用现代测量理论为指导的研究者由于担心无法很好地把握该理论的原理和方法望而却步。为了让现代测量理论的发展研究成果能够更多地用于指导研究和实践工作，测量学研究者应该做出更多的努力。

江西师范大学心理与教育测量研究中心团队在漆书青、戴海琦、丁树良等教授的带领下，从20世纪80年代初开始对现代测量理论进行深入研究，取得了许多理论和实践研究成果，研究团队也进一步发展壮大。随着研究的深入以及研究领域的进一步拓展，加之现代测量理论受到越来越多研究者的关注，江西师范大学心理与教育测量研究中心团队顺应形势和发展需要，基于自身近30年的理论研究和实践积累，出版一套关于高级心理测量的丛书，这是心理与教育研究领域的一件有益之事，也必将进一步推动心理与教育测量理论与技术在中国的发展。

2012年，该团队曾经出版了一套围绕项目反应理论研究的丛书，该丛书的出版取得了很好的反响。现在，该团队在前期研究和实践的基础上，准备再出版一套关于高级心理测量的丛书，我听后感到的非常高兴，且对中国在现代测量理论领域的发展前景充满信心和期待。

高级心理测量学丛书主要有：《计算机化自适应测验：理论与方法》《认知诊断理论》《认知诊断评价理论基础》《高级认知诊断》《Q矩阵理论及认知诊断测验的编制》以及《智慧化测评的理论与技术》等著作。这套丛书包括了当今国际上比较前沿的研究领域，涉及计算机化自适应测验、认知诊断理论和智慧化测评等，这对于推动中国的心理测量学发展及其为实践服务具有重要意义。

值此丛书即将付梓出版之际，作为与江西师范大学心理与教育测量研究团队交流合作多年的同行，我倍感欣慰，特作此短序以示祝贺，并希望他们在今后取得更多的研究成果和更大的发展。

张华华

于美国伊利诺伊大学香槟校区  
2017年5月

## 张华华教授简介

专业领域：心理学、教育心理学、统计学。

美国伊利诺伊大学香槟校区(University of Illinois Urbana-Champaign)终身教授，世界著名心理测量学杂志《应用心理测量》(*Applied Psychological Measurement*)主编，世界心理计量学会 2012—2013 年主席(President of the International Psychometric Society for 2012—2013)，全美教育研究学会院士(AERA Fellow)。

# 自序

大数据时代，人类生活、工作和思维方式发生着重大变革。知识更新速度越来越快，知识倍增周期变得越来越短。高效的教育、评价和学习方式变得至关重要，这给测量与评价的理论和技術带来了重大挑战和发展机遇。测试能够加强学生记忆、增强其理解深度和培养其学习迁移能力。而如今许多测试只是简单的测评，并不能积极地支持个性化学习，测试效率不高，也很难对学生进行全方位的评价。我们在本书中主要介绍面向学习、自适应和大数据等智慧化测评的理论和技術，期望测评方式的进步能够引起教育、学习和决策的重大变革。

认知科学、测量技术、信息技术的发展，带给我们崭新的测评和学习方式。测评理论和信息技术与学习相互分离的现状将逐渐转变。本书采用全新的视角，结合最新的测评理论与技术、学习评价理论、云计算和大数据技术，介绍智慧化测评理论与技术，以提高学习效率，服务于个性化、自适应学习和智能决策。本书作者具有测量学、计算机科学、数学、教育学和心理学的学术背景，从不同学科视角阐述智慧化测评的理论与技术，并汇集智慧化测评的关键技术。

随着人工智能在教育中的应用，借用国际商业机器公司(IBM)原首席执行官彭明盛(Samuel J. Palmisano)于2008年11月首次提出的“智慧地球”的概念，我们把本书命名为《智慧化测评的理论与技术》，期望以此激发广大读者的智慧。本书内容分为五篇。

第一篇，测评概述与发展趋势。第一章由宋丽红撰写，简要介绍测评的概念、历史和意义。第二章由汪文义撰写，概括测评智慧化的三大关键性变化：诊断化和学习化、自适应化和网络化、大规模化和大数据化。

第二篇，诊断化和学习化测评。第三章和第四章由宋丽红撰写，分别介绍学习评价理论和认知诊断评估理论。第五章由汪文义撰写，介绍课堂评估、Q矩阵标定、自动组卷和学习进程纵向评价等内容，以及相关算法、模型和技术用以更好地促进测评诊断化和学习化。

第三篇，自适应化和网络化测评。第六章由汪文义撰写，主要介绍计算机化自适应测验和应用案例。基于云服务实施计算机化自适应测验，可更好地提高测试效率和降低测试实施成本。第七章由汪文义和罗芬撰写，介绍面向服务的测评架构和技术，为未来测评平台和测评支撑软件的开发指明方向。

第四篇，大规模化和大数据化测评。第八章由宋丽红和汪文义撰写，介绍测验偏差检验方法和决策指标等，为测评公平性检验提供统计方法，并为实现综合学业测评决策提供支持。第九章由汪文义和熊建华撰写，介绍神经网络和贝叶斯网络等在测评

中的应用。

第五篇，智慧化测评展望。第十章由汪文义撰写，主要归纳智慧化测评的应用和发展方向。

感谢宋丽红、罗芬、熊建华等作者对本书做出的贡献。感谢丁树良教授、戴海琦教授多年来的培养。感谢张华华教授对本书部分章节(如第六章和第八章)内容的指导。本书得到江西师范大学计算机信息工程学院、初等教育学院和心理学院领导和老师们的支持，在此一并对他们表示感谢。另外，感谢两位硕士研究生汪腾和高朋对本书的校对。还要感谢北京师范大学出版社何琳编辑对本书出版所做的工作。本书得到全国教育科学规划教育部重点课题“基础教育质量监测分数报告方法研究”(DHA150285)的资助。

本书可供心理统计、心理测量、教育统计测量评价、教育技术学和计算机科学等专业教学使用，也可供从事测评的学习者、工作者和研究者参考。

限于时间和能力，本书仍有许多不足之处，恳请广大读者批评指正。

汪文义  
于江西师范大学瑶湖校区  
2019年3月

# 目 录

---

## CONTENTS

### 第一篇 测评概述与发展趋势

#### 第一章 测量与评价 /3

##### 第一节 测量与评价的概述 /4

一、测量与评价的基本概念 /4

二、测量与评价的基本性质 /4

##### 第二节 测量与评价的发展 /6

一、测评起源阶段 /6

二、测评量化阶段 /7

三、测评兴起阶段 /9

四、测评大规模阶段 /11

五、测评学习化阶段 /14

##### 第三节 测量与评价的意义 /17

一、测评的分类 /17

二、测评的教育功能 /17

三、测评的其他功能 /18

#### 第二章 测量与评价智慧化 /20

##### 第一节 测评诊断化和学习化 /21

一、测评诊断化 /21

二、测评学习化 /22

##### 第二节 测评自适应化和网络化 /24

一、测评自适应化 /24

二、测评网络化 /25

第三节 测评大规模化和大数据化 /26

一、测评大规模化 /26

二、测评大数据化 /27

## 第二篇 诊断化和学习化测评

### 第三章 学习评价理论 /31

#### 第一节 学习段评价 /32

一、学习段评价的缘起 /32

二、学习段评价的概念 /32

三、学习段评价的内涵 /33

#### 第二节 学习性评价 /34

一、学习性评价的缘起 /34

二、学习性评价的概念 /34

三、学习性评价的内涵 /35

#### 第三节 学习化评价 /37

一、学习化评价的缘起 /37

二、学习化评价的概念 /38

三、学习化评价的内涵 /38

#### 第四节 学习评价的启示 /39

一、学习评价观念的转变 /39

二、学习评价的发展 /40

### 第四章 认知诊断评估概述 /42

#### 第一节 认知诊断评估的缘起 /43

一、认知取向的心理测量学的发展需要 /43

二、个性化认知信息的社会需求 /44

#### 第二节 认知诊断评估的概念 /45

一、认知诊断评估的属性 /45

二、认知诊断评估的概念 /45

#### 第三节 认知诊断评估的内涵 /46

一、认知诊断评估的理论 /46

二、认知诊断评估的发展 /47

- 第四节 认知诊断评估的框架 /50
  - 一、测量目的描述 /50
  - 二、认知模型构建 /51
  - 三、测验  $Q$  矩阵设计 /54
  - 四、任务开发 /56
  - 五、模型评价 /57
  - 六、数据分析和结果报告 /58

## 第五章 认知诊断评估核心技术 /60

- 第一节 课堂评估 /61
  - 一、课堂评估的缘起 /61
  - 二、课堂评估的概念 /61
  - 三、课堂评估的内涵 /62
  - 四、课堂评估的发展 /64
- 第二节 认知诊断中智能化  $Q$  矩阵标定 /65
  - 一、EM 算法 /65
  - 二、 $Q$  矩阵探索性方法 /67
  - 三、 $Q$  矩阵修正性方法 /69
  - 四、 $Q$  矩阵质量评价 /74
- 第三节 认知诊断中智能化组卷算法 /75
  - 一、认知诊断中项目质量评价 /75
  - 二、认知诊断中测验质量评价 /79
  - 三、认知诊断中组卷算法 /80
- 第四节 认知诊断中学习进程评价 /84
  - 一、学习进程 /84
  - 二、学习进程评价模型 /86

## 第三篇 自适应化和网络化测评

## 第六章 计算机化自适应化测验 /91

- 第一节 多维项目反应理论 /92
  - 一、项目反应理论特点 /92
  - 二、项目反应理论模型 /92

第二节	计算机化自适应测验	/93
一、	CAT 题库	/93
二、	CAT 选题算法	/93
三、	CAT 能力估计	/95
四、	CAT 曝光控制	/97
五、	CAT 终止策略	/100
六、	CAT 题库扩充	/102
七、	CAT 应用案例	/103
第三节	计算机化多阶段自适应测验	/105
一、	MST 核心思想	/105
二、	MST 关键设计	/107
三、	MST 应用案例	/109
第四节	计算机化自适应诊断测验	/112
一、	CD-CAT 题库	/113
二、	CD-CAT 初始题	/113
三、	CD-CAT 选题算法	/113
四、	CD-CAT 分类方法	/120
五、	CD-CAT 终止规则	/121
六、	CD-CAT 题库扩充	/122
七、	CD-CAT 应用案例	/122
第七章	面向服务的测评架构和技术	/128
第一节	面向服务的测评架构	/129
一、	测评服务化的必然性	/129
二、	测评服务化的愿景	/130
三、	面向服务化的测评架构	/132
第二节	云测评平台和技术	/136
一、	云测评平台的发展	/136
二、	云测评平台的特征	/137
三、	云测评平台的技术	/138
第三节	云题库的发展和技木	/142
一、	云题库的发展	/142
二、	云题库的特征	/143
三、	云题库的技术	/144

#### 第四节 教育大数据分析和技术 /146

- 一、教育大数据 /146
- 二、教育大数据的特征 /147
- 三、教育大数据分析技术 /147

## 第四篇 大规模化和大数据化测评

### 第八章 大规模测评的公平性技术 /153

#### 第一节 测评公平性 /154

- 一、教育公平性 /154
- 二、测评公平性 /154

#### 第二节 测评公平性评价发展 /155

- 一、国外公平性评价 /155
- 二、国内公平性评价 /157

#### 第三节 测评公平性检验方法 /157

- 一、基本概念 /157
- 二、方法分类 /158
- 三、检验假设 /158
- 四、检验方法 /160

#### 第四节 测评公平性决策指标 /167

- 一、决策指标的作用 /167
- 二、决策规则 /172
- 三、决策指标 /173

#### 第五节 测评公平性的启示 /178

- 一、科学命题和科学施测 /178
- 二、科学评分和科学分析 /178
- 三、科学决策和科学研究 /180

### 第九章 测评大数据分析技术 /183

#### 第一节 深度学习在测评中的应用 /184

- 一、深度学习 /184
- 二、神经网络在测评中的应用 /185
- 三、深度学习在测评中的应用展望 /189

- 第二节 知识计算在测评中的应用 /189
  - 一、知识计算 /189
  - 二、贝叶斯网在测评中的应用 /191
  - 三、知识计算在测评中的应用展望 /202
- 第三节 社会计算在测评中的应用 /203
  - 一、社会计算 /203
  - 二、社会计算在测评中的应用展望 /204
- 第四节 可视化技术在测评中的应用 /205
  - 一、可视化技术 /205
  - 二、可视化技术在测评中的应用展望 /205

## 第五篇 智慧化测评展望

- 第十章 智慧化测评的发展方向 /209
  - 第一节 智慧化测评的应用研究 /210
    - 一、大规模测评中智慧化测评的应用 /210
    - 二、课堂评估中智慧化测评的应用 /210
    - 三、在线教育中智慧化测评的应用 /210
  - 第二节 智慧化测评的理论创新 /211
    - 一、大规模智慧化测评的理论创新 /211
    - 二、智慧课堂测评的理论创新 /212
    - 三、智慧在线教育测评的理论创新 /214
- 参考文献 /216

## 第一篇 测评概述与发展趋势



## 第一章 测量与评价

本章主要介绍测量与评价的基本概念和性质，测量与评价的五个发展阶段，以及测量与评价的分类、功能和意义，让读者了解潜在特质为何可测，如何才能保证测量的客观性、测量与评价的演变过程等。20世纪初期，测评处于研究智力、人格测验等心理测量范式的阶段。20世纪中期，测评注重经典测量理论和教学目标测试等。20世纪中后期，客观化测量理论和统计模型催生了大规模测验和计算机化自适应测验。20世纪末至今，测评主要关注学习的促进，从大规模测验中挖掘诊断信息，注重形成性评价和课堂评价。