

高级心理测量学丛书

涂冬波 蔡艳 高旭亮 汪大勋◎著

高级 认知诊断

GAOJI
RENZHI ZHENDUAN



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

高级心理测量学丛书

涂冬波 蔡艳 高旭亮 汪大勋◎著

高级 认知诊断

GAOJI
RENZHI ZHENDUAN



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

高级认知诊断/涂冬波等著. —北京:北京师范大学出版社, 2019.4

(高级心理测量学丛书)

ISBN 978-7-303-23286-4

I. ①高… II. ①涂… III. ①心理测量学—研究 IV. ①B841.7

中国版本图书馆 CIP 数据核字(2018)第 002684 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
北京市海淀区新街口外大街 19 号
邮政编码:100875

印 刷:北京京师印务有限公司
经 销:全国新华书店
开 本:787 mm×1092 mm 1/16
印 张:24.25
字 数:473 千字
版 次:2019 年 4 月第 1 版
印 次:2019 年 4 月第 1 次印刷
定 价:88.00 元

策划编辑:何琳	责任编辑:李云虎 梁丽美
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。

張華華

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)。

前言

随着现代测量理论和认知心理学的进一步发展及两者的结合,认知诊断(cognitive diagnosis)日益成为现代心理与教育测量领域重要的研究方向,并受到国内外研究者和应用者的广泛关注。

我们之前已出版过关于认知诊断的著作,本书是在此基础上撰写的。本书在笔者多年来对认知诊断的研究及实践的基础上完成,对认知诊断的相关理论、方法、技术以及应用等方面进行了全面、详细的阐述,其中有大量笔者近年来的研究成果,为我国相关学者在这一领域的发展提供借鉴和参考。本书主要有以下特点。

第一,重点介绍了我们开发的认知诊断分析平台(flexCDMs)的使用操作(第十二章)。该平台具有使用简单、操作可视化、功能强大等鲜明特点和优势,感兴趣的读者可以访问我们的平台:<http://www.psychometrics-studio.cn/>。

第二,内容更加系统、完善。本书包含了四篇:理论篇、应用篇、研究篇和程序使用操作篇,包含了当前国内外认知诊断的主要研究及应用领域。

第三,突出了认知诊断在心理障碍评估中的应用。我们在第六章重点介绍了认知诊断技术在抑郁症症状诊断评估及网络成瘾诊断评估中的应用,凸显了认知诊断技术在这一领域的潜在优势与价值。

第四,注重认知诊断当前国际最新研究前沿,如多级评分认知诊断模型、 Q 矩阵估计、属性多级化的CD-CAT以及多级评分的CD-CAT等。

第五,注重认知诊断的理论、方法与实践相结合。本书较系统地阐述了认知诊断的基本理论和方法(如第一章至第四章等),并对相关理论、方法如何在实际中应用进行了翔实介绍(如第五、第六、第七章等)。

本书写作大纲由涂冬波和蔡艳完成,参与本书编写的主要有以下几位老师,在此表示感谢。第一章,涂冬波;第二章,涂冬波、汪大勋;第三章,高旭亮;第四章,蔡艳、刘书畅;第五章,蔡艳、赵洋;第六章,涂冬波、蔡艳;第七章,谢美华、李丽;第八章,蔡艳、涂冬波;第九章,涂冬波、高旭亮;第十章,高椿雷;第十一章,喻晓锋、汪大勋、刘永;第十二章,涂冬波、赵洋、刘书畅、高旭亮。

全书最后由涂冬波统稿。本书在写作及统稿过程中得到了戴海琦教授、丁树良教授及美国伊利诺伊大学香槟校区终身教授张华华教授的帮助,还得到了江西师范大学心理学院各位领导及老师的支持,在此一并表示感谢!同时也要感谢我们指导的研究

生们对本书的校稿做出的贡献!

限于时间及能力,本书仍有许多不足之处,恳请广大读者批评指正(邮箱:
tudongbo@aliyun.com)。

涂冬波

2019年2月

于江西师范大学心理统计与测量研究中心

目 录

CONTENTS

第一部分 理论篇

第一章 认知诊断概述 /2

第一节 心理与教育测验理论的发展 /2

一、经典测量理论 /3

二、项目反应理论 /6

第二节 认知诊断概述 /11

一、新一代测验理论的产生 /11

二、认知诊断简介 /12

三、 Q 矩阵理论 /14

四、认知诊断的意义 /17

第二章 认知诊断模型 /19

第一节 认知诊断模型概述 /19

一、认知诊断模型简介 /19

二、认知诊断模型分类 /19

第二节 简化的认知诊断模型 /22

一、连接型/非补偿型认知诊断模型 /22

二、离散型/补偿型认知诊断模型 /27

三、简化的认知诊断模型特点与比较 /31

第三节 饱和的认知诊断模型 /32

一、广义诊断模型 /33

二、对数线性认知诊断模型 /34

三、拓广 DINA 模型 /35

第四节 饱和的 CDMs 与简化的 CDMs 的关系及转换 /36

一、一致性连接函数 G-DINA 模型与简化的认知诊断模型的关系 /37

- 二、log 连接函数 G-DINA 模型与简化的认知诊断模型的关系 /38
- 三、logit 连接函数 G-DINA 模型与简化的认知诊断模型的关系 /40

第五节 认知诊断模型比较 /41

- 一、饱和模型与简化模型间的关系及模型选用 /41
- 二、认知诊断模型展望 /43

第三章 认知诊断模型的参数估计 /44

第一节 G-DINA 模型的参数估计 /44

- 一、G-DINA 模型 /44
- 二、G-DINA 模型的项目参数估计 /45
- 三、G-DINA 模型的被试参数条件估计 /49

第二节 简化模型的参数估计 /50

- 一、简化模型参数估计 /50
- 二、认知诊断模型参数估计小结 /51

第四章 认知诊断测验的项目与测验分析 /53

第一节 认知诊断模型—资料拟合检验 /53

- 一、项目拟合指标 /54
- 二、测验拟合检验 /57

第二节 认知诊断测验项目区分度分析 /59

- 一、区分度定义的几种方法 /60
- 二、常用认知诊断模型的题目区分度计算实例 /62
- 三、基于信息量的认知诊断区分度 /63
- 四、认知诊断区分度小结 /70

第三节 认知诊断测验项目功能差异检验 /71

- 一、项目功能差异检验的概念与意义 /71
- 二、认知诊断框架下的项目功能差异检验 /72

第四节 认知诊断测验信度、效度分析 /75

- 一、认知诊断测验的信度 /75
- 二、认知诊断测验的效度 /77

第五章 认知诊断测验分析实例 /79

第一节 测验与模型拟合度检验 /79

- 一、Tatsuoka 分数减法数据简介 /79
- 二、测验与单一模型的拟合度检验 /79

- 三、测验与混合模型的拟合度检验 /80
- 第二节 测验 Q 矩阵分析 /81
 - 一、测验 Q 矩阵对项目难度的贡献率 /81
 - 二、 Q 矩阵分析小结 /82
- 第三节 项目质量分析 /83
 - 一、项目区分度分析 /83
 - 二、项目拟合检验 /83
 - 三、项目功能差异检验 /84
 - 四、项目质量分析小结 /86
- 第四节 测验质量分析 /87
 - 一、测验信度分析 /87
 - 二、测验效度分析 /87

第二部分 应用篇

- 第六章 认知诊断技术在心理障碍评估中的应用 /90
 - 第一节 心理障碍评估中的属性与属性标定 /90
 - 一、心理障碍评估中的属性 /90
 - 二、心理障碍评估中的属性标定 /90
 - 第二节 认知诊断在抑郁症症状评估中的应用 /92
 - 一、抑郁症症状评估 /92
 - 二、基于认知诊断的抑郁症症状评估的量表开发 /93
 - 三、抑郁症症状评估结果分析 /104
 - 四、抑郁症症状评估小结 /111
 - 第三节 认知诊断在青少年网络成瘾评估中的应用 /112
 - 一、网络成瘾的概念 /112
 - 二、基于认知诊断的青少年网络成瘾评估量表的编制 /113
 - 三、G-DINA 模型在青少年网络成瘾症状中的评估诊断分析 /117
 - 四、网络成瘾诊断小结 /121
- 第七章 认知诊断技术在教育评估中的应用 /123
 - 第一节 认知诊断在汉语阅读中的应用 /123
 - 一、阅读及阅读能力的定义 /123
 - 二、初中生现代文阅读理解认知模型的建构 /123
 - 三、阅读能力认知诊断测验的编制 /124
 - 四、阅读能力认知诊断过程的研究过程 /125

五、阅读能力认知诊断研究结果 /126

六、阅读能力认知诊断研究小结 /139

第二节 认知诊断在中学生空间能力评估中的应用 /140

一、空间能力认知属性的确定 /141

二、空间能力认知诊断分析 /143

三、空间能力认知诊断小结 /148

第三部分 研究篇

第八章 认知诊断模型开发研究 /152

第一节 属性多级化的认知诊断模型开发及其 Q 矩阵设计 /152

一、属性多级化及其 Q 矩阵 /152

二、属性多级化认知诊断模型开发及其 Q 矩阵设计思路 /154

三、实验 1: PA-DINA 模型和 PA-RRUM 模型参数估计的实现及其精度 /160

四、实验 2: 不同样本容量下 PA-DINA 模型和 PA-RRUM 模型的性能 /163

五、实验 3: 独立型属性层级关系下属性多级化测验 Q 矩阵设计—— R_p^* 阵及对诊断结果的影响 /164

六、实验 4: 属性有层级关系下属性多级化测验 Q 矩阵设计—— R_p^* 阵及对诊断结果的影响 /168

七、属性多级化的认知诊断研究结论与讨论 /170

第二节 基于加工过程的多级评分认知诊断模型开发 /171

一、基于加工过程的认知诊断模型 /171

二、属性基于项目反应类别的多级评分认知模型开发思路 /174

三、实验 1: DM-DINA 模型合理性及科学性验证 /176

四、实验 2: DM-DINA 模型与传统的 DINA 模型的比较研究 /178

五、多级评分认知诊断模型研究结论与讨论 /181

第九章 认知诊断计算机化自适应测验研究 /184

第一节 属性多级化认知诊断计算机化自适应测验 /184

一、属性多级化认知诊断计算机化自适应测验概述 /184

二、属性多级化认知诊断计算机化自适应测验设计 /185

三、pCD-CAT 的实现及其与传统 CD-CAT 的比较 /189

四、实验 1: 定长条件下 pCD-CAT 效果 /192

五、实验 2: 不定长条件下 pCD-CAT 效果 /194

六、实验 3: pCD-CAT 与传统 CD-CAT 的比较 /195

七、pCD-CAT 研究结论与讨论 /197

第二节 多级评分的认知诊断计算机化自适应测验 /199

一、多级评分的认知诊断计算机化自适应测验概述 /199

二、多级评分的认知诊断计算机化自适应测验设计 /200

三、蒙特卡罗模拟研究 /204

四、实验 1: 定长条件下 psCD-CAT 效果 /206

五、实验 2: 不定长条件下 psCD-CAT 效果 /207

六、psCD-CAT 研究结论与讨论 /209

第三节 可修改答案的 CD-CAT /210

一、可修改答案的 CD-CAT 简介 /210

二、可修改答案的 CD-CAT 的实验研究与结果分析 /213

三、可修改答案的 CD-CAT 研究结论与讨论 /221

第十章 具有认知诊断功能的计算机化多阶段测验 /223

第一节 多阶段测验 /223

一、多阶段测验的基本结构 /223

二、MST 的特点 /225

三、MST 的影响因素 /225

第二节 具有认知诊断功能的多阶段测验及其影响因素 /226

一、认知诊断与 MST /226

二、CD-MST 简介 /228

三、实验 1: CD-MST 在线组卷方法比较 /231

四、实验 2: CD-MST 同 CD-CAT 的比较研究 /241

五、CD-MST 的研究结论与讨论 /242

第三节 CD-MST 初始阶段模块组建方法 /243

一、CD-MST 的模块组卷 /243

二、CD-MST 初始阶段模块组建方法 /244

三、CD-MST 除初始阶段以外的模块的组建方法 /248

四、CD-MST 组卷方法的实验研究设计 /248

五、CD-MST 组卷方法实验结果 /249

六、CD-MST 组卷方法的实验结论与讨论 /254

第十一章 测验 Q 矩阵估计方法研究 /257

第一节 测验 Q 矩阵估计方法综述 /257

- 一、 Q 矩阵估计方法 /257
- 二、认知诊断测验 Q 矩阵估计方法及其思路 /258
- 三、测验 Q 矩阵估计方法的比较 /269
- 四、测验 Q 矩阵估计方法未来研究方向 /271

第二节 基于似然比 D^2 统计量的测验 Q 矩阵估计 /273

- 一、 Q 矩阵的构建与估计 /273
- 二、IRT 下的模型拟合度评价方法 /275
- 三、使用 D^2 统计量估计测验 Q 矩阵 /275
- 四、 Q 矩阵估计的模拟研究 /278
- 五、 Q 矩阵估计研究结论与讨论 /285

第三节 一种简单有效的 Q 矩阵估计方法开发：海明距离法 /286

- 一、 Q 矩阵的估计 /286
- 二、基于海明距离的 Q 矩阵估计方法开发 /287
- 三、基于海明距离的 Q 矩阵估计方法示例 /289
- 四、实验 1：基于海明距离的 Q 矩阵估计方法效果验证 /291
- 五、实验 2：基于海明距离的 Q 矩阵估计方法在实测数据中的研究 /294
- 六、 Q 矩阵估计研究结论与讨论 /296

第四节 一种非参数化的 Q 矩阵估计方法：ICC-IR 方法开发 /298

- 一、基于理想得分的 ICC 指标(ICC-IR)法开发 /298
- 二、ICC 指标计算及理想得分矩阵构建示例 /300
- 三、实验 1：ICC-IR 法估计 Q 矩阵效果验证 /302
- 四、实验 2：ICC-IR 法与 D^2 统计量的比较 /306
- 五、实验 3：ICC-IR 法和 D^2 统计量方法在实测数据中的比较研究 /308
- 六、ICC-IR 方法研究结论与讨论 /309

第四部分 程序使用操作篇

第十二章 认知诊断分析平台 /312

第一节 认知诊断分析平台概述 /312

- 一、认知诊断分析平台简介 /312
- 二、平台主要功能 /313
- 三、平台主界面 /314
- 四、平台网址与浏览器要求 /315

第二节	认知诊断分析平台使用操作	/315
一、	Q 矩阵理论模块使用操作	/315
二、	参数估计模块使用操作	/317
三、	模型拟合模块使用操作	/325
四、	测验质量分析模块使用操作	/327
五、	项目功能差异(DIF)模块使用操作	/329
六、	Q 矩阵修正与估计模块使用操作	/332
七、	图形输出模块使用操作	/339
八、	数据服务模块使用操作	/342
第三节	认知诊断分析平台的优势	/345
一、	G-DINA 包	/345
二、	CDM 包	/345
三、	认知诊断分析平台的优势	/345
第四节	认知诊断分析平台参数估计精度验证	/346
一、	认知诊断分析平台的认知诊断模型及验证方法	/346
二、	蒙特卡罗模拟实验	/346
三、	参数估计精度评价指标	/349
四、	flexCDMs 参数估计精度实验结果	/350
五、	flexCDMs 参数估计精度结论	/351
参考文献		/352

第一章 认知诊断概述

本章回顾了心理与教育测验理论的发展，重点介绍了经典测量理论和项目反应理论，并在此基础上，阐述了新一代测验理论产生的背景与意义，最后重点介绍了认知诊断的产生、相关概念及理论等。

第一节 心理与教育测验理论的发展

自比内和桑代克时期以来，心理与教育测验理论已经过百年的发展，先后发展出经典测量理论(classical test theory, CTT)、概化理论(generalizability theory, GT)、项目反应理论(item response theory, IRT)和认知诊断(cognitive diagnosis, CD)(见图 1-1-1)，这些测验理论的产生及发展对解决许多心理、教育甚至其他学科中的实际问题起了重要作用。

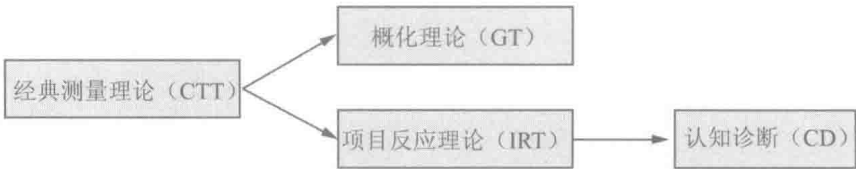


图 1-1-1 心理与教育测验理论发展

经典测量理论是最先发展起来的一种测验理论，也是迄今为止应用最为广泛的一种测量理论，其核心内容包括真分数、信度、效度等概念，至今它仍在指导测验开发、分析和应用方面发挥着重要作用。研究者在经典测量理论的基础上发展出了概化理论和项目反应理论。

概化理论主要是在经典测量理论的基础上，应用实验设计及其分析、方差分析模型等统计工具对 CTT 的信度进行深化与推广，即结合测量的情境关系对 CTT 给出的笼统的误差进行探查和分解，探明误差的不同来源，并在一定范围内改变测量情境，以考察这种改变引起的误差的相对变化，从而对误差进行控制，以提高测验的信度。概化理论使人们能站到新的理论高度，更加有力且有预控性地去改善和提高测量精度与准确性。

项目反应理论是为了克服经典测量理论的一些不足而发展起来的一种新兴测量理论,相比于经典测量理论,项目反应理论不论是对被试的分析还是对测验项目的分析均更为细致与深入。更为重要的是,在项目反应理论指导下发展起来的计算机化自适应测验(computerized adaptive testing, CAT)实现了“因人施测”的自适应测量思想,是测量方式上的重大变革。

认知诊断则是在项目反应理论的基础上发展起来的,其最为突出的贡献是将传统测验的功能——“评价与筛选的功能”上升为“促进发展的功能”,这是心理与教育测验理论发展中一次质的飞跃。

一、经典测量理论

(一)经典测量理论的数学模型

$$X = T + E \quad (\text{公式 } 1.1.1)$$

公式 1.1.1 为 CTT 的数学模型。该模型假设被试的观察分数(X)由两个因素决定,一是被试的真实能力分数即真分数(T),另一个是测量误差分数(E),而且假设观察分数就是真分数与误差分数的简单线性相加。

在公式 1.1.1 的基础上,CTT 假设被试真分数与误差分数相互独立,即被试真分数越高,其误差分数不一定也越高(或越低),即两者在数量上的变化方向无规律,也即不相关。在这一独立假设之下,根据公式 1.1.1 可得:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (\text{公式 } 1.1.2)$$

公式 1.1.2 表明,被试的观察分数的变异或方差(σ_X^2)也由两部分构成:真分数的变异(σ_T^2)和误差分数的变异(σ_E^2)。心理与教育测量中,人们希望观察分数的变异绝大部分是由被试真分数的变异构成的,极少部分来自误差分数的变异,这种情况一般表明我们测量到的观察分数能较好地反映被试的真实水平,也即测量结果可信,这也是 CTT 对信度的统计定义,参见公式 1.1.3:

$$r = \frac{\sigma_T^2}{\sigma_X^2} \quad (\text{公式 } 1.1.3)$$

r 为 CTT 的统计信度。

然而在实际测量中,被试的真分数(T)是未知的,即公式 1.1.3 中的分子部分无法求取,也即 CTT 定义的统计信度无法计算,因此信度的统计定义只有理论上的意义而无实际意义。

(二)经典测量理论的信度

由于经典测量理论无法计算其统计信度,因而需开发可实际计算的信度。我们知道,用一个工具对同一个物体的长度测量多次,如果多次测量的结果都不同