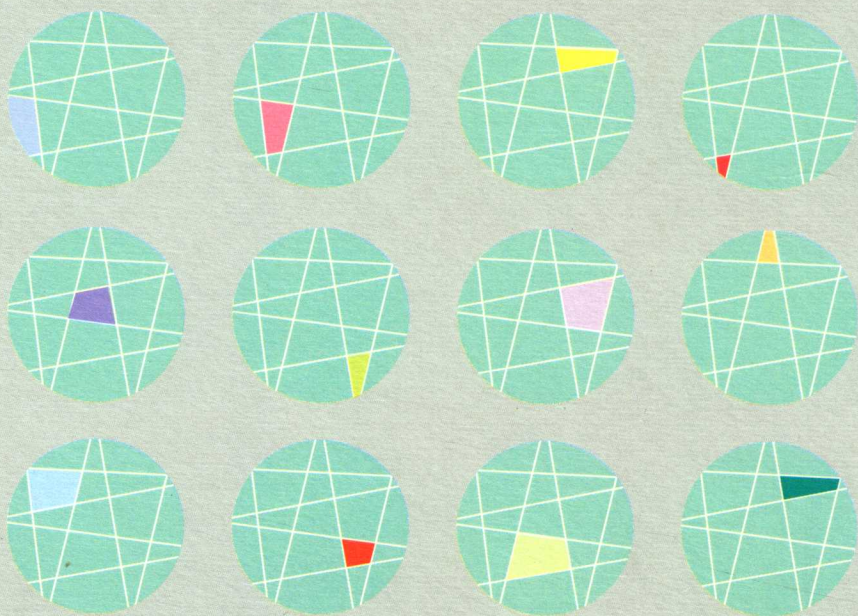


中国教育质量监测与评估丛书



教育认知诊断评估 理论与技术研究

汪文义 宋丽红 著



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

教育认知诊断评估 理论与技术研究

*Research on Theory and Technology in
Cognitive Diagnostic Assessment for Education*

汪文义 宋丽红 著



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

教育认知诊断评估理论与技术研究/汪文义,宋丽红著.
—北京:北京师范大学出版社,2015.8
(中国教育质量监测与评估丛书)
ISBN 978-7-303-19291-5

I. ①教… II. ①汪… ②宋… III. ①教育评估—
研究 IV. ①G40-058.1

中国版本图书馆 CIP 数据核字(2015)第 173328 号

营 销 中 心 电 话 010-58805072 58807651
北师大出版社学术著作与大众读物分社 <http://xueda.bnup.com>

JIAOYU RENZHI ZHENDUAN PINGGU LILUN YU
JISHU YANJIU

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:北京联兴盛业印刷股份有限公司
经 销:全国新华书店
开 本:730 mm×980 mm 1/16
印 张:17.5
字 数:256 千字
版 次:2015 年 8 月第 1 版
印 次:2015 年 8 月第 1 次印刷
定 价:68.00 元

策划编辑:陈红艳	责任编辑:齐琳 王玲玲
美术编辑:袁麟	装帧设计:锋尚制版
责任校对:陈民	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

丛书编委会

编委会主任：辛 涛

编委会成员(按姓氏笔画排序)：

丁树良	王 耘	边玉芳	任 萍
刘红云	杨 涛	李凌艳	辛 涛
张咏梅	罗 良	胡平平	

序 言

随着我国九年义务教育的全面普及，全面提升质量成为义务教育改革发展的核心任务。促进教育质量提升的重要前提是全面、准确地了解和把握教育质量状况。然而，很长一段时间以来，我们既不能客观全面评价教育质量状况，也不能有效诊断教育存在的问题及其根源。为此，《国家中长期教育改革和发展规划纲要（2010—2020年）》提出，要建立教育质量监测、评估体系，定期发布测评结果。《中共中央关于全面深化改革若干重大问题的决定》指出，要“委托社会组织开展教育评估监测”。开展教育质量监测，对教育质量进行科学、全面、有效的评价，已成为实现我国基础教育科学发展、内涵发展的重大举措和战略任务。

为适应我国基础教育改革与发展的需要，北京师范大学联合国内外多家单位组建了中国基础教育质量监测协同创新中心。这是我国第一个、也是目前唯一一个通过教育部认定的教育学和心理学领域的国家协同创新中心。中心着眼于构建中国特色、国际先进的基础教育质量监测体系，围绕基础教育质量标准体系建设、监测体系构建与制度设计、大数据采集与分析平台建设、教育决策支持系统、教育质量提升工程等重要任务展开协同攻关，努力在科学研究、学科建设、人才培养、社会服务等方面取得突破与进展，为全面提升我国教育质量，促进亿万儿童与青少年的全面、个性发展提供有力支撑。

协同创新中心成立以来，集聚了国内外相关领域的优秀人才和资源，创新了合作研究工作机制，持续开展了相关领域的协同创新研究，取得了积极进展。中心先后研制了义务教育阶段语文、数学、科学、品德、体育、艺术等学科领域的监测标准和工具，并通过教育部审定。研制了

《国家义务教育质量监测方案》，经教育部审核通过并向社会正式发布。继 2007~2014 年连续承担国家义务教育质量试点监测任务后，2015 年组织开展了我国首次国家义务教育质量监测。此外，积极探索国家监测结果的运用机制，开展基于证据的教育决策咨询服务，持续多年对所有样本省、县进行了“一对一”的监测结果反馈，为国家和地方教育决策、教育教学改进提供了客观依据以及针对性建议，取得了良好的效果。

除了满足国家基础教育质量监测的需求之外，协同创新中心还致力于服务地方开展基础教育质量监测的需要。中心开展了基础教育质量监测能力建设项目，先后组织专家通过各种方式对全国 20 多个省相关人员进行多轮次基础教育质量监测通识培训和专业技术培训，从一般性理论、理念到具体的技术、方法，为参训人员呈现了基础教育质量监测的“完整图像”，对转变相关人员观念、提升专业水平起到了良好的促进作用。同时，重点对部分队伍力量相对较强的省级监测机构进行针对性培训，全程指导这些机构独立开展本地基础教育质量监测，有力地提升了这些机构的专业能力。通过培训和指导，一些省级监测机构，如重庆市，已能独立、完整地开市域内基础教育质量监测，为所在地区的教育决策、区域教育质量的提升与均衡发展提供了有力的支持。

随着各地基础教育质量监测机构的不断建立，对相关专业支持和指导的需求也不断加大，对协同创新中心也提出了新的要求。据中心初步统计，截至 2014 年底，全国共有 22 个省(区、市)成立了省级基础教育质量监测机构。这些机构大多依托省级督导评估中心、教研室、教科院、评估院、招生考试院等部门而成立，相关从业人员队伍普遍存在数量不足、年龄偏大、专业水平较低等问题，整体力量相对薄弱，难以有效开展本地区基础教育质量监测，迫切需要专业的支持和指导。在与协同创新中心交流过程中，地方监测机构有关从业人员纷纷表示，教育质量监测是一项专业性强、技术含量高的工作，希望中心能系统介绍开展监测相关的理论、技术和方法，分享自身开展这项工作的实践经验，以引领

和指导各地开展好这项工作。

为更好地满足各地对开展基础教育质量监测的不同需求,指导各地有效开展有关工作,协同创新中心围绕“什么是基础教育质量监测”“如何开展基础教育质量监测”两大主题组织编写了这套丛书。丛书包括《国际基础教育质量监测实践与经验》《基础教育质量监测工具研发》《基础教育质量监测抽样设计与数据分析》《基础教育质量监测报告撰写与结果应用》《大规模学业成就调查的开发:理论、方法与应用》《教育认知诊断评估理论与技术研究》六本。其中,《国际基础教育监测实践与经验》主要回答“什么是基础教育质量监测”的问题,《基础教育质量监测工具研发》、《基础教育质量监测抽样设计与数据分析》、《基础教育质量监测报告撰写与结果应用》则主要回答“如何开展基础教育质量监测”的问题。《大规模学业成就调查的开发:理论、方法与应用》与《教育认知诊断评估理论与技术研究》在理论与实践层面都进行了探讨。

具体而言,《国际基础教育质量监测实践与经验》主要包括基础教育质量监测概述、世界各国与国际组织开展的基础教育质量监测实践、中国开展基础教育质量监测的探索等内容。《基础教育质量监测工具研发》以监测工具开发的程序与规范为主线,详细介绍了学业成就测试工具、相关因素调查问卷开发的科学流程与具体要求。《基础教育质量监测抽样设计与数据分析》对抽样的基本概念、抽样方法、数据处理与分析、标准划定、测试题目质量检验分析方法等内容进行了介绍和说明。《基础教育质量监测报告撰写与结果使用》介绍了监测结果报告的种类、推动监测结果发挥最大使用效益的策略和方法等内容。《大规模学业成就调查的开发:理论、方法与应用》从介绍全球范围内大规模学业成就调查项目入手,详细阐述了当前大规模学业成就调查开发的理论基础、常用方法与结果应用。《教育认知诊断评估理论与技术研究》主要探讨了认知诊断测验编制原理、项目属性标定新方法、认知诊断新模型、认知诊断测验的信度和效度,以及计算机化自适应诊断测验等重要的前沿研究热点。

本套丛书既可作为相关专业与方向培训的教材，也可供有关从业人员自学之用。丛书从酝酿选题、内容编排到成书出版，经历了三四年的时间。期间，协同中心组织编著人员围绕有关内容进行了反复多次讨论，并在重庆、江西等省市围绕相关内容开展了系统的培训和试点，请参训人员提出了意见和建议，并进行了不断修改完善。有别于“急就章”式的著作，本套丛书凝聚了协同创新中心多年来的探索实践经验，当中的很多观点和内容均为中心这些年来不断思考、积淀的结果，是中心为广大正在从事或有志于从事基础教育质量监测工作的读者奉献的诚意之作。

本套丛书的出版得到了教育部基础二司和世界银行的有关项目资助，世界银行还为这套丛书的编写无偿提供了相关的材料，北京师范大学出版社，特别是策划编辑陈红艳女士为此倾注了大量心血，在此一并表示衷心的感谢。丛书虽经反复修改、不断完善，但由于教育质量监测相关的测量理论、技术、方法日新月异，疏漏和错误在所难免，恳请广大专家和读者批评指正。

中国基础教育质量监测协同创新中心
2015年8月

目 录

第一篇 教育认知诊断评估背景

第一章 认知诊断评估概述	(3)
第一节 什么是认知诊断评估	(3)
第二节 为什么要进行认知诊断评估	(4)
第三节 认知诊断评估发展历程	(6)
第四节 如何进行认知诊断评估	(10)

第二篇 教育认知诊断评估属性界定

第二章 认知诊断评估中属性界定方法	(25)
第一节 Q 矩阵	(25)
第二节 Q 矩阵指定、选择和验证方法	(27)
第三节 小结与讨论	(38)
第三章 数据驱动的属性界定方法	(42)
第一节 为什么要提出数据驱动方法	(42)
第二节 形式概念分析简介	(46)
第三节 基于形式概念分析的属性辅助标定方法	(51)
第四节 形式概念分析方法属性标定效果评价	(57)

第五节	小结与讨论	(66)
第四章	结合理论和数据的属性界定方法	(68)
第一节	三种属性标定方法	(68)
第二节	训练集充分条件下各种方法的准确性评价	(71)
第三节	训练集不充分条件下各种方法的准确性评价	(83)
第四节	小结与讨论	(90)
第五章	属性界定方法的应用	(92)
第一节	各属性辅助标定方法的适用条件	(92)
第二节	修正 Q 矩阵理论及其应用	(93)
第三节	基于形式概念分析的标定方法的实证研究	(101)
第四节	在线标定方法在纸笔测验中的应用	(102)
第五节	小结与讨论	(108)

第三篇 教育认知诊断评估计量模型

第六章	心理计量模型介绍	(111)
第一节	理论基础或认知假设	(111)
第二节	常见的认知诊断模型	(119)
第三节	模型—资料拟合检验	(135)
第四节	常用软件及使用	(148)
第七章	改进的确定性输入噪音与门模型提出与参数估计	(154)
第一节	模型的提出	(154)
第二节	模型参数估计	(157)

第三节	模型参数估计准确性检验	(165)
第四节	小结与讨论	(173)
第八章	认知诊断模型—资料拟合检验指标及其应用	(175)
第一节	认知诊断模型—资料拟合检验指标	(175)
第二节	在三个认知诊断模型比较中的应用	(178)
第三节	在 Tatsuoka 分数减法数据分析中的应用	(186)
第四节	小结与讨论	(195)
第九章	混合模型开发	(197)
第一节	可逆跳转马尔可夫链蒙特卡洛 (RJMCMC) 方法 ..	(197)
第二节	混合模型的 RJMCMC 算法构建	(200)
第三节	RJMCMC 算法检验	(201)
第四节	混合模型的应用	(204)
第五节	小结与讨论	(205)

第四篇 教育认知诊断评估的信度和效度

第十章	认知诊断测验的信度和效度研究	(209)
第一节	涉及的基本概念	(211)
第二节	认知诊断测验的信度研究	(213)
第三节	认知诊断测验的效度研究	(217)
第四节	小结与讨论	(219)
第十一章	认知诊断测验的信度和效度评估方法	(221)
第一节	分类一致性和分类准确性指标	(222)

第二节	分类一致性和分类准确性指标的评价	(225)
第三节	信度和效度指标在一大规模考试中的应用	(228)
第四节	小结与讨论	(228)

第五篇 教育认知诊断评估自适应测验

第十二章	计算机化自适应诊断测验研究	(233)
第一节	认知诊断测验题库	(233)
第二节	初始题选择	(234)
第三节	选题策略	(234)
第四节	知识状态估计方法	(241)
第五节	终止规则	(242)
第六节	原始题参数在线估计和属性在线标定	(243)
第七节	小结与讨论	(244)
第十三章	计算机化自适应诊断测验新选题策略及应用	(246)
第一节	兼顾测验效率和题库使用率的新选题策略	(247)
第二节	各选题策略比较与评价	(249)
第三节	小结与讨论	(253)
第十四章	教育认知诊断评估展望	(255)
参考文献		(258)
后 记		(265)

第一篇 教育认知诊断评估背景

教育认知诊断评估旨在测量学生特定的知识和加工技能、报告学生掌握或尚未掌握哪些知识或技能。教育认知诊断评估的终极目标是成为“促进学习的评价”。要诊断学生掌握了哪些知识或技能，需要一系列的过程来完成。例如描述测验目的、构建认知模型、设计测验蓝图、开发测验项目、评价认知模型和选择心理计量模型、估计模型参数及报告诊断结果等。随着认知心理学的发展和一批著名测量学家的积极投入，作为新一代测验理论的代表，认知诊断评估在结合现代统计学和现代心理学基础上取得了迅速地发展。20 世纪末，Tatsuoka 等人开创性地提出并发展了认知诊断评估理论，20 世纪末和 21 世纪前 10 年，认知诊断模型取得了长足的进步，约在 2010 年之后，认知诊断评估更多地关注于测验 Q 矩阵设计、测验 Q 矩阵标定、信度和效度评估方法以及自适应测验等方面的研究。

第一章 认知诊断评估概述

第一节 什么是认知诊断评估

认知诊断评估(cognitive diagnostic assessment, CDA)是 20 世纪中后期兴起的一种新的测验形式。它以认知心理学为理论基础,结合现代测量学,旨在探索人类在特定领域的潜在的认知过程和认知结构,试图分析和解释人类的个体差异现象。统计测量理论的发展分为两个阶段,一个是以项目反应理论(Item Response Theory, IRT)为代表的标准测量理论(standard test theory)阶段,另一个是新一代测量理论阶段(new generation of test),认知诊断评估便是现代测量理论的核心内容(Mislevy, 1993)。由于认知心理学在认知诊断评估中起着重要的理论指导作用,这一新的测量形式又被称为认知取向的测量学范式。

广义的认知诊断是指建立起观察分数和被试内部认知特征之间的关系。狭义的认识诊断则是指在教育教学领域中,按是否掌握了测验所测的技能或特质来对被试加以分类。认知诊断评估的目的是测量/评价个体特定的知识结构(knowledge structure)和加工技能(processing skills)(Gierl, Leighton, & Hunka, 2000)。认知诊断评估通过测验获得被试在测验上可观察的反应模式(observed response pattern, ORP)而推知该被试不可观察的知识状态(knowledge state, KS)。认知诊断评估中被试的知识状态是指被试对于评估任务中的认知属性(attribute)或技能(skill)的掌握情况。其中,属性或技能是指被试要成功完成某项任务所需具备的认知加工技能、知识或策略,是进行某一具体领域工作所需的程序性和陈述性知识(本书中,技能和属性两个术语将交替使用)。属性间可以是没

有关联彼此独立的，也可以是存在相关关系，甚至彼此间具有一定的层级关系，即属性掌握的先后存在着一定的逻辑或心理顺序或先决关系。有研究者指出，在教育评估中，可以获得学生所说的、所做的或特定情境下所做的数据，然后试着推断学生所知的、能做的或可以完成的信息(Mislevy, Steinberg, & Almond, 2003)。认知诊断评估就是要通过测验设计和分析来获取关于被试的所知、所能的属性侧面(attribute profile)信息。

第二节 为什么要进行认知诊断评估

认知诊断评估的兴起有其必然性。首先，认知诊断评估的兴起缘于心理测量学本身发展的需要。认知诊断评估是建立在对传统心理与教育测量学的问题与危机的反省基础之上的。顾名思义，心理测量学(psychometrics)是心理学(psychology)与测量学(metrics)的结合，是心理学与测量学的交叉学科。然而在经历了短暂的与实验心理学紧密结合的早期阶段之后，心理测量学逐渐演变成去理论化的、功能性和实用性取向的、带有浓重数理统计色彩的应用学科，与主流心理学的距离越来越远(杨向东, 2010)。一直以来，很多著名学者既是心理测量学家也是心理学家(Anastasi, 1967)，如高尔顿、比纳、斯皮尔曼、克龙巴赫等，他们都对人的心理加工过程十分感兴趣。然而，这些专注于心理测量领域的心理学家越来越追求先进的统计技术，而忽略了心理研究本身。Mislevy也曾说，20世纪80年代以前，主导教育领域的测量，不夸张地说，仅仅是将20世纪的统计技术加之于19世纪的心理学(Leighton & Gierl, 2007)。

无论是经典测量理论(Classical Test Theory, CTT)、概化理论(Generalizability Theory, GT)还是项目反应理论，其测量的主要目的都是将被试在某一特质(如学术成就)量尺上进行定位，并据此进行选拔和安置。它们未很好地使用心理学理论解释学生在项目上的作答，没有对心理变量如何影响被试项目反应做出假设，对测验所测结构的心理过程未做出明确描述，对于学生如何推理和解决问题的分析主要是基于研究

者的期望而非学生的实际情况。比如说,对于学生在一个项目中作答错误这一现象的解释,传统测验只能说学生没有掌握这个项目,而无法诊断学生到底是在哪个知识点没有掌握或心理过程上哪里出现障碍。

测验开发者呼吁变革测验功能,心理或教育测量应将重心由选拔转移到诊断。1989年,Messick指出了测验分数中提供用于探查学生心理过程的相关信息的重要性。他指出,多数教育工作者感兴趣的是学生的心理过程,而教育测量学家应该提供此类信息作为测验的目标。也就是说教育测验要能够提供关于学生在所测领域的认知过程和知识结构中的强项和弱项的反馈信息(Leighton & Gierl, 2007)。所以,心理和教育测量应该解释学生在推理和解决问题过程中所用到的知识和技能,并提供关于学生在知识技能上的掌握情况的诊断信息。认知诊断评估就是一种旨在探查心理过程的一种实质性的评估方法。不同于传统测验侧重评估学习结果,认知诊断评估侧重于评价学生的学习过程。

认知诊断评估的兴起缘于社会的需求。现在,社会各方面对教育测量提出了迫切的要求。例如,我国新一轮基础教育课程改革提出“万校一书”的课程体系和“万人一面”的培养方式已经不能适应社会发展的需求。课程评价方面,以前的评价过于强调甄别与选拔,现在强调评价是为了改进教学、促进发展。再如,美国前总统布什2001年提出“不让一个孩子掉队(No Child Left Behind)”法案,美国总统奥巴马2009年提出“力争上游(Race to the Top)”教改计划(美国复苏法案中的一部分),均要求大规模的测验除必须给学生提供关于学习结果的信息外,还要给教师提供诊断性的信息,以指导教学。课程改革要体现在考试内容上,通过考试使得教师明白测验项目是根据什么教学目标设计的,从而影响教学。每次考试,找出几所学校,派专家回到这些学校告诉他们:你的学生考试成绩如何,哪部分是薄弱环节,如何去补救。

随着社会、经济、文化和教育等的发展,学生、家长和教师越来越不满足于传统测量所提供的个体的宏观层次的评价。人们不仅要了解个体总的来说怎么样,还需要了解个体具体来说怎么样。传统测量主要是提供一个测验总分,关注的焦点是考生分数的结果。这个分数过于粗糙,