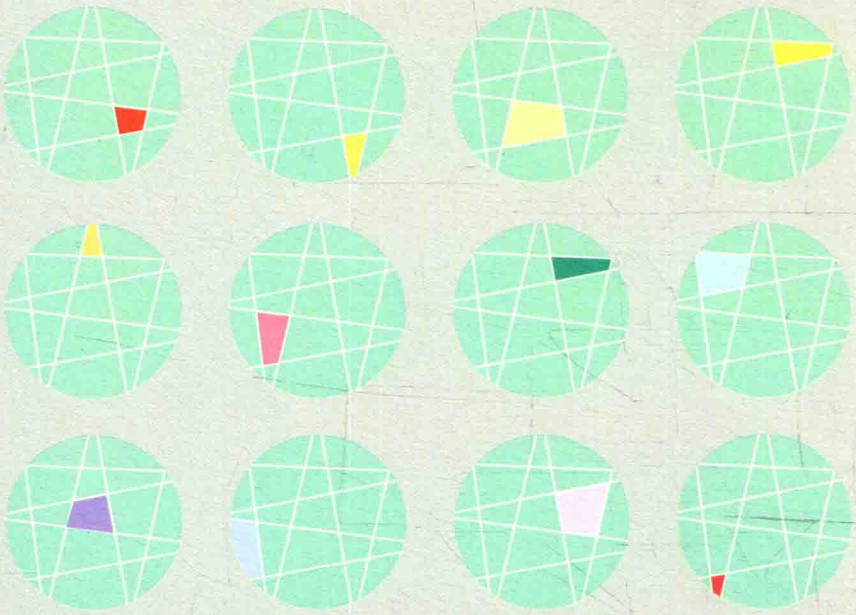
中国教育质量监测与评估丛书



大规模学业成就调查的开发：理论、方法与应用

张咏梅　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

北京市教育科学“十二五”规划2014年度重点课题
“大规模学业水平测试能力量尺的开发与应用”的研究成果
（课题编号：3002-0014）

大规模学业成就调查的开发：理论、方法与应用

The Development of Large Scale Assessment : Theory, Methodology and Application

张咏梅　著

安徽师范大学图书馆
藏书

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

大规模学业成就调查的开发：理论、方法与应用/张咏梅著. —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 978-7-303-19356-1

Ⅰ. ①大… Ⅱ. ①张… Ⅲ. ①学业评定-研究-中国
Ⅳ. ①G424.75

中国版本图书馆 CIP 数据核字(2015)第 184422 号

北京市教育科学“十二五”规划 2014 年度重点课题
“大规模学业水平测试能力量尺的开发与应用”
研究成果(课题编号：3002—0014)

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

DAGUIMO XUEYE CHENGJIU DIAOCHA DE KAIFA
出版发行：北京师范大学出版社 www.bnup.com
北京新街口外大街 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：24.75
字　　数：381 千字
版　　次：2015 年 8 月第 1 版
印　　次：2015 年 8 月第 1 次印刷
定　　价：78.00 元

策划编辑：陈红艳　　责任编辑：齐　琳　常慧青
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。

此为试读，需要完整PDF请访问：www.ertongbook.com

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、完整地开展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

序 一

我学教育学的时候，教育测量学连讲座课都不是，仅在心理学的课上极其简略地了解到一点欧美发达国家心理测量方面的概况。对于发达国家在教育研究中大量运用教学质量数据的方法，也只是将其理解为教育统计学领域的事情。80年代中后期接触国际教育成就评价协会（IEA）的“国际数学和科学教育成就趋势调查研究项目（TIMSS）”后，才模模糊糊意识到教育测量学的意义和价值及其在教育研究中实际运用的操作方式。但它那精细的测量标准拟订、信效度原则及检验方法、关照教学全过程的测量方案、严密的测量程序、复杂的因果和相关分析方法等，无不让人望而却步。

真正开始了解教育测量的机会，是由北京对义务教育学校进行教育质量监测项目提供的。该项目始于2003年，实际执行人即本书作者。这个项目进行到第五个年头时，在一次会议间隙，时任教研中心主任的王云峰教授兴奋地给我介绍了该项目的进展情况，我第一直觉是：这不就是TIMSS和PISA等项目所做的事情么！于是要来云峰当时所能提供的所有资料尝试着学习，并开始特别关注他们的进展情况。2011年，有幸与作者成为教科院组织的牛津大学培训班的“同学”，学习期间了解到她的教育背景、学术专长和研究兴趣，明白了“北京市义务教育教学质量分析与评价反馈系统”（简称“质评系统”，英文简称BAEQ）所以能做到向国际著名同类项目的快速迈进，确与她及她所领导的一支在心理与教育测量学方面训练有素的高水平专业研究团队密切相关，与这个团队严谨的科研作风、崇高的事业目标和坚忍不拔的毅力相关。

作者带领她的团队十几年如一日地坚持做下来的，是一项追赶世界

先进水平的事业。虽然他们最大限度地学习、参照了国际上所有知名的测量项目的理论、方法和经验，但对于在北京实际操作来说，仍然几乎等于从零开始。他们借助与国际同行的不断交流和自己在测量实践中的深刻反思，逐渐积累了大量丰富的经验，形成了自己的测量风格及施测程序。本书既是作者十余年来学术经历的总结，也是国内首批基于作者实践经历的大规模学业成就测量学方法论专著。

实际从事一项大规模学业成就调查，远不像教育测量学教科书所讲的那么简单。本书14章中的中间12章可视为一项大规模学业成就调查所必备的基本步骤。本书的价值在于通过理论研究和实践反思，向读者提供了一份完整的开展大规模学业成就调查的基本框架，同行大可参照本书框架研制自己的学业成就调查方案。当然，从另一角度说，书中所提供的每一步骤大都可视为一个独立的理论和方法体系，都值得全面展开、深入探讨。诸如“命题”“评价与标准的一致性”“等值”“题库”“结果解读与应用”等，尽管本书已经尽可能做出了比较全面的阐述，但每一个问题也许都值得从理论到方法，从经验到问题做出更加深入和详尽的论述，甚至都有必要独立成为一部专著。

使教育测量学服务于教育评价和教学改进，是国际上教育研究界的共识，也是我国教育评价和教学改进的未来前景。愚以为，本书算得上是我国推进此项事业的奠基著作之一。

耿申

2015年8月

序　二

本书作者邀请我为新书写序，我很珍惜这个机会。因为觉得此事重要，迟迟未有动笔。十二年的积淀，或薄或厚的一本书，除了书中的内容，背后定有讲不完的故事。作为作者多年的好友，我能体会她十几年如一日的坚持不懈、执着追求。

至今清楚地记得“北京市义务教育教学质量分析与评价反馈系统”（中文简称“质评系统”，英文简称 BAEQ）项目初期的艰难。说实话，起初我觉得这事情太难、太专业、太辛苦，结合国内教育测评专业发展的现状和测评技术在实际中的诸多问题，以及没有专业测评团队的实际，我甚至劝她放弃。而她却执着地认为，不管怎样，这事情总要有人做。接下来的几年，她如饥似渴地参加各种教育测量专业的培训，自学国际测评项目的相关资料，培养和组建专业化团队。这些几乎用掉了她所有的假期和周末，多少个晚上我们在电话中切磋交流、相互学习，为了一个技术问题讨论、争执。正是她的这份执着和不放弃，才有了她和她所领导的团队专业上的成长和发展。我很欣慰也很敬佩这支团队这么多年取得的成就。

几个月前，当她把书稿交给我的时候，问我：“你认为在教育评价领域这些年最大的收获是什么？”我想，一件事情坚持做了十几年，收获的肯定不单单是面前这一摞书稿吧？十二年的探索与实践，也许最大的收获是培养了一支专业高效的测评团队，也许最大的收获是对于专业知识更深的理解与更灵活的应用，也许只是收获了从青春年少的激情与执着走入不惑之年的豁达与睿智。无论怎样，这本书的出版，可以说是对十二年工作的回顾和总结，意味着作者对这一领域专业化工作的执着追求

此为试读，需要完整PDF请访问：www.ertongbook.com

和无私奉献。

这本书全面地介绍了大规模学业成就调查的理论、方法和应用，内容非常好。最主要的特点之一就是，着重笔墨于系统阐述特定的教育评价应用领域。要想做到这一点很不容易，它不仅要求作者对这一领域的专业知识有系统深入的了解，而且要求作者对这一领域应用研究中遇到的诸多问题有独立见解和方法。读者可能很容易找到一本介绍教育评价中某个技术环节的书，如项目反应理论、等值、标准划定，但是如果想要找到一本系统完整将这些技术综合应用到实践研究领域的著作，据我所知目前本书还是第一本。因此，我认为本书对于想要从事教育测评研究和实践工作的研究者是很好的参考资料。本书的另一突出特点在于，所有研究方法都从中国的实际情况出发，借鉴而非照搬国外大型测评项目，因而，产生了一批具有原创性的科研成果。在中国做教育评价会遇到很多西方国家难以想象的困难，作者通过多年的探索，广泛吸纳了国际和国内诸多最前沿专家的智慧和经验，在设计上充分考虑了我国教育发展的特点。NAEQ项目包括各个学科的评价指标体系的建立，科学系统的测验开发流程、数据管理和分析系统的建立，相应的统计分析软件的开发和自动化报告生成系统的建立，等等，这些都将成为国家和地方基础教育未来发展的可贵资料。

作者结合自己十几年的工作实践和体会，历时几载，文字几经修改，终于成书。宁静质朴中，彰显了作者对专业化工作孜孜不倦的追求。可以肯定地说，本书的出版必定对教育评价工作起到积极的推动作用。时间将会证明，这是此领域一本不可多得的开创之作。

刘红云

2015年8月于京师园

目 录

第一章　全球大规模学业成就调查项目概述

随着全球社会与经济发展水平的不断提升，教育质量的内涵及其评价体系开始引起世界各国的广泛关注。学业质量作为评价教育质量的核心指标，早在20世纪60年代就在由国际教育成就评价协会(IEA)发起的国际数学和科学教育成就趋势调查研究项目(TIMSS)、由美国教育部发起的国家教育进步计划(NAEP)中开始了系统性调查、研究与实施。21世纪初，随着各国教育改革政策的不断颁布与实施，国际经济合作与发展组织(OECD)又开发了国际学生评价项目(PISA)，旨在通过开展学生能力国际性比较研究来帮助各国改善教育政策、提高教育质量。在此背景下，各国也纷纷开展了基于本国课程体系、指向教育质量评价需要的学业成就调查项目，教育发达地区的省级层面也启动了相关的研究。本部分不仅对全国范围内国际层面、国家层面、国家省级层面的学业成就调查项目进行阐述，还对当前我国的学业成就调查项目进行说明。并在此基础上，就大规模学业与成就调查项目的特征进行分析。

第一节　国际级学业成就调查项目

一、关于PISA

(一)什么是PISA

PISA是国际学生评价项目(Programme for International Student Assessment)的简称，这是一项由国际经济合作与发展组织(Organization for Economic Cooperation and Development，OECD)开展的学生能力国际性比较研究。PISA通过对全球接近完成基础教育的15岁学生，即未来的社会公民，在个人、工作和社会生活中运用已学知识和已掌握技能解决

此为试读，需要完整PDF请访问：www.ertongbook.com