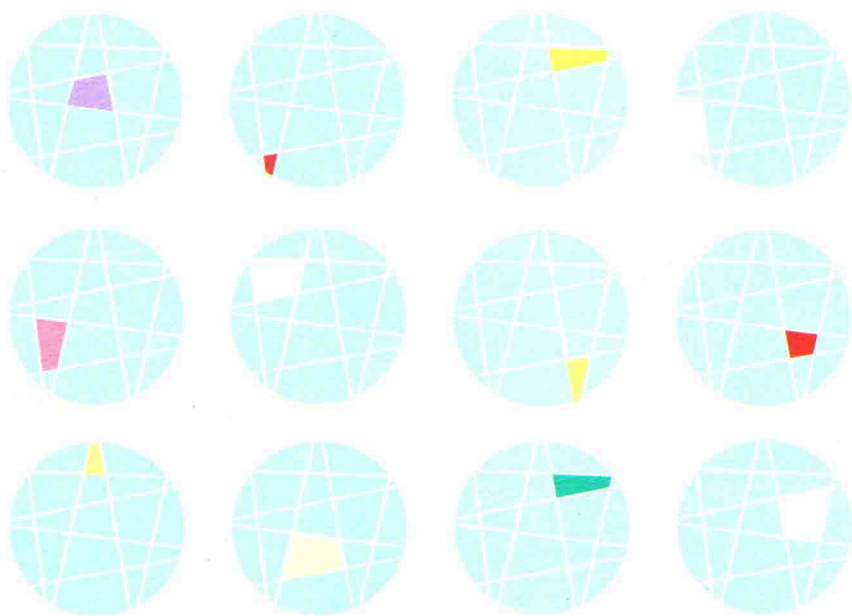


中国教育质量监测与评估丛书



基础教育质量监测 工具研发

边玉芳 梁丽婵 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

基础教育质量监测 工具研发

*Developing Tools for National Assessment
of Educational Quality*

边玉芳 梁丽婵 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

基础教育质量监测工具研发/边玉芳等著. —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 978-7-303-19294-6

I. ①基… II. ①边… III. ①基础教育—教育质量—质量管理—教材 IV. ①G632.0

中国版本图书馆 CIP 数据核字(2015)第 173395 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>

JICHU JIAOYU ZHILIANG JIANCE GONGJU YANFA

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: 13

字 数: 200 千字

版 次: 2015 年 8 月第 1 版

印 次: 2015 年 8 月第 1 次印刷

定 价: 48.00 元

策划编辑: 郭兴举 陈红艳 责任编辑: 赵雯婧 张 爽

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

目 录

绪 论	(1)
第一章 教育质量监测的内容与方法	(7)
第一节 教育质量监测的内容	(7)
第二节 教育质量监测的多元评价方法	(11)
第三节 教育质量监测的对象和评价主体	(22)
第二章 教育质量监测工具的一般研发流程和质量分析	(25)
第一节 教育质量监测工具的一般研发流程	(25)
第二节 教育质量监测工具的质量分析	(40)
第三章 学业成就测验工具的研发	(55)
第一节 学业成就测验在教育质量监测中的作用	(55)
第二节 学业成就测验工具研发的一般流程	(57)
第三节 学业成就测验的题目编写团队的构建	(59)
第四节 学业成就测验工具的初步编制	(63)
第五节 学业成就测验工具的预试与修订	(95)
第四章 问卷类工具的研发	(104)
第一节 问卷在教育质量监测中的作用	(104)
第二节 问卷研发的一般流程	(106)

第三节	问卷的初步编制	(110)
第四节	问卷的预试与修订	(137)
第五章	表现性评价工具的研发	(142)
第一节	教育质量监测中的表现性评价	(142)
第二节	表现性评价工具研发的基本流程	(154)
第三节	表现性评价工具的初步编制	(160)
第四节	表现性评价工具评分标准的制定	(165)
第五节	表现性评价工具的预试	(172)
第六章	教育质量监测工具操作手册的编制	(182)
第一节	教育质量监测工具操作手册的作用	(182)
第二节	测试操作手册的内容	(184)
第三节	测试操作手册的样例	(190)
参考文献		(196)

绪 论

历史的车轮已进入 21 世纪，面对知识经济的兴起、经济全球化的挑战以及国际竞争的加剧，提高全民素质、增强国家综合国力与国际竞争力成为各个国家的共同选择。基础教育质量决定着一个国家未来的全民素质，决定着国家劳动力的水平，是影响国家能否持续发展和具有国际竞争力的核心因素。因此，世界各国莫不重视其基础教育的质量，同时也通过各种方式纷纷开展基础教育质量监测工作。

一、教育质量监测是提升教育质量的重要手段

从 20 世纪 70 年代起，基础教育质量就开始被世界各国共同重视。以美国为代表的部分国家首先建立了国家基础教育质量的监测评价系统。随后，芬兰、法国、英国、德国、加拿大、澳大利亚、新西兰、日本以及我国的香港和台湾地区也纷纷建立了基础教育质量测评体系。同时，为了准确了解教育质量的现状和发展态势，及时发现教育中存在的问题，更好地为教育决策服务，很多国际组织和国家先后出台了教育质量评估项目，如国际经济合作与发展组织(OECD)的“国际学生评估项目”(PI-SA)、国际教育成就评估协会(IEA)的“国际数学和科学学习趋势”(TIMSS)和“国际阅读素养发展研究”(PIRLS)、美国的“国家教育进展评估”(NAEP)、英国的“国家课程测验”(SATS)、澳大利亚的“国家评估计划”(NAP)、日本的“全国学力考试”都是其中典型的例子。这些监测项目为各个国家和地区了解和掌握自身教育质量现状以及与其他国家与地区

的差异做出了重要贡献。总体而言,开展基础教育质量监测和评估已成为世界教育发展的一大潮流。

国际组织和各国家(地区)的基础教育质量监测项目在评估学生学业水平现状和发展趋势外,还探索影响学生学业能力发展的真实原因,为教育系统提供改进的方法和决策依据,而且已经在各国基础教育的发展与改革中产生了巨大影响。如,由于 NAEP 监测结果显示美国学生的阅读能力在近 20 年来都处于停滞不前的状态,2005 财政年度小布什政府在学生阅读能力培养和教育方面投入了 12 亿美元的资金,比 2001 年增长了 4 倍;2005 年英国的皇家督学报告指出贫困地区的学校以及特殊教育学校缺乏优秀的教师影响了当地的教学质量,于是布莱尔政府在 2006 年出台《教育与督学法案》承诺向贫困地区的教育部门提供 3000 万英镑的资金以加强他们与先进学校的合作;巴西从教育质量评估数据的分析中发现教师的学历对学生的成绩有重要影响,拥有研究生学位的教师所教的学生成绩较为优异,巴西教育部随后就制订了鼓励教师参加培训和资格认证的计划。

目前,我国教育正从原有粗放式发展模式转向内涵式发展模式,注重教育公平、教育均衡,以学生为本、关注每一个学生的发展。教育质量的提升是当前教育改革和发展的核心任务,是当前教育工作的重点。而实施教育质量监测正是及时掌握基础教育质量现状和变化情况的重要手段,是实现教育行政部门决策和管理的科学化、民主化和专业化的有效途径,是教育质量提升、均衡发展的必然之路。因此,教育部及各级教育部门开始高度关注基础教育质量监测问题,进行了一些积极的探索。教育部先后依托上海教育科学研究院和北京师范大学成立了教育部基础教育监测中心和教育部基础教育质量监测中心,各地区教育行政部门也纷纷成立了当地的教育质量监测机构。从 2007 年 11 月开始,教育部基础教育质量监测中心,负责组织开展基础教育质量监测的日常工作,

至今已有8年。2015年国务院教育督导办印发《国家义务教育质量监测方案》，决定从2015年起在全国开展义务教育质量监测工作。由此可见，基础教育质量监测在我国已经被提高到了前所未有的高度。此外，我国各地区也相继开展基础教育质量监测工作。从2006年开始，北京市探索制定“学科学业质量标准”，已在此标准体系的指导下形成了在三、五、八年级实行年度抽样监测，并每年形成质量报告，向社会公布研发的机制。2008年至2011年，江苏、上海、重庆、北京等地区成立了省(市)层面的基础教育质量监测机构；浙江省杭州市下城区、江苏省如皋市等区县层面的基础教育质量监测机构也在不断地建立中。而其他省市均在积极筹备本区域的基础教育质量监测机构，全国范围的区域基础教育质量监测与评估系统正在逐步形成。

基础教育质量监测在我国之所以重要，是因为它能达成以下目标：了解全国各地基础教育阶段贯彻党和国家教育方针，实施素质教育，促进学生全面发展的落实情况；分析我国基础教育均衡发展在教育公平的状况；把握我国农村地区、西部地区基础教育，特别是中小學生的发展状况；了解新课程改革的进展状况、具体成效以及存在的问题；发现影响基础教育质量的相关因素，并提出相应的政策建议。

二、监测工具的质量是有效开展教育质量监测的关键因素

工具研发是基础教育质量监测的核心程序，也是基础教育质量监测最关键的环节之一。基础教育质量监测的科学性、有效性，在很大程度上取决于工具研发。因为监测工具研发的质量，决定了监测数据的有效性、准确性，决定了数据分析结果的科学性。基础教育质量监测目的的实现与监测工具的质量也有密切关系。一套有效的教育质量监测工具需要经历科学严谨的研发流程，需要不同领域的专家共同合作。无论是各国际大型教育评估项目，还是各国基础教育质量监测工作，都将工具研

发视为教育质量监测工作的核心内容,有的国家或国际组织更是专门斥资组建负责监测工具研发与结果数据分析的专业单位。著名国际学生评估项目 PISA 中,以 PISA 2006 为例,设立由五个国际知名的研究机构组成的国际命题专家组,它们分别是:澳大利亚教育研究理事会(ACER)、荷兰教育考试院(CITO)、奥斯陆大学、基尔大学、日本国立教育研究所。另外,还有 4 个专家组(科学专家组、阅读专家组、数学专家组、技术专家组)配合命题的工作。这些命题专家组不仅负责一部分的命题工作,也对各参与国提交的试题进行评审和修订,保证 PISA 试题的质量。从命题的时间上来看,主要经过最初的设计框架研究(2002.12 ~ 2003.06)、确定测试框架(2003.11 ~ 2004.08)、命题(2003.07 ~ 2004.10)、各参与国提交试卷和参与试题评审(2004.02 ~ 2004.06)等程序。时间顺序显示,PISA 的命题时间在三年左右。著名的美国教育进展评估(National Assessment of Educational Progress, NAEP),就是由美国国会授权教育部的国家数据中心(NCES)、美国教育考试服务中心(ETS)专门研发组织实施的基础教育质量监测活动。由专业人员负责整个评估中监测工具的研发、数据收集与结果分析。目前该评估是美国唯一的全国性的、持续性的教育质量监测项目,也是当今国际上最为著名的国家级教育质量监测项目之一。英国也在教育部下成立了资格与课程局(QCA),负责管理与调试基础教育的评估体系,以适应学习者与社会的要求。该组织负责研发高质量的基础教育监测工具,评价学生的学业成就发展状况,检查学校的教学质量以及国家课程的实施情况。我国教育部基础教育质量监测中心也设置了标准与工具研发部,专门负责组织专家制定基础教育质量监测所涉及领域的监测框架和指标体系,研发标准化测试工具。由此,不难看出,大型国际教育质量监测项目和国家(地区)教育质量监测相关部门和机构对工具研发的重视程度真是非同一般。

监测工具的研发需要一个严谨、科学的过程。目前,基础教育质量

监测在我国还刚刚起步,我们发现,在当前我国基础教育质量监测工作中,监测工具的研发存在不少问题。如果我们不重视这些问题,将会影响监测工作的顺利开展,我们必须高度重视。

目前在监测工具的研发中存在的问题主要有:

1. 将教育质量监测等同于统考统测,表现在工具的研发上就是将日常的考试试卷等同于监测工具。教育质量监测的目的不同于传统的统考统测,二者在内容上更是不可同日而语的。教育质量监测的主要目的之一是在整体上了解一个国家与地区在贯彻党和国家教育方针、实施素质教育、促进学生全面发展的落实情况,内容涵盖教育质量涉及的方方面面,方法上是多元的,体现在工具研发上就是有适合于教育质量监测的多样的评价工具。

2. 缺乏对工具研发的整体设计。一个好的工具必须首先明确监测目的、监测内容、监测对象、监测工具类型、结果使用方式等,但目前有些监测工具的研发过程常常“倒置”:常常在缺乏理论的思考基础上先研发工具,等有了工具甚至收集好了数据再来考虑能做出什么报告、可以得出什么结论的问题。这导致监测工具常常发生“无意中”脱离监测目的、不适合监测对象作答的情况。

3. 研发随意性大。有人曾说,不就是一份考试卷和一份问卷,有什么不会的,于是随意编制几个题目构成一份学业水平测试工具或一份问卷,并未严格按照科学原则和流程设计工具。有的工具依赖“便利性”就地取“材”,拼凑而成;有的从别人研发的原有工具中“不管三七二十一”拿来就用,合适不合适就不考虑了,这将严重影响监测工具的科学性和有效性。

4. 监测工具没有经过科学的预试。在正式测试前没有对工具的质量进行系统的分析和验证,工具的有效性和科学性值得怀疑。我们发现,在一些地区,有一些负责教育质量监测工具研发人员对什么是信度、什

么是效度等概念都不了解，更不用谈工具质量的分析和验证了。

5. 工具缺乏标准化。没有对工具操作、实施过程、评分过程、结果解释等环节进行标准化，使得工具存在误用、错用的情况，不能保证结果的有效性和可靠性。

以上的问题使得监测工具的有效性和可靠性大打折扣，所收集数据的真实性和准确性降低，导致监测结果不科学、不严谨。鉴于以上现象，我们撰写了此书。本书的主要内容是介绍如何科学地研发基础教育质量监测工具，希望通过本书，能帮助教育质量监测一线的研究人员了解监测工具研发的科学流程、编制技巧和相关注意事项，为我国基础教育质量监测工作的有序开展提供帮助。

第一章 教育质量监测的内容与方法

监测工具的研发与教育质量监测的内容、方法有密切关系，它们决定了工具研发的目的、内容和形式。工具研发各环节的工作都必须围绕教育质量监测的内容和方法展开。是否准确、客观、全面地反映教育质量监测的内容，是否严格遵循所选取的监测方法的实施原则，是工具质量研发的一个最基本、最重要的衡量标准。

第一节 教育质量监测的内容

教育质量监测的内容与监测目的密切相关。制定教育质量监测的内容是教育质量监测实施的关键环节，对整个教育质量监测工作都有重要的指导作用。

一、教育质量监测的主要内容

教育质量监测的目的是准确地了解基础教育质量的现状和发展趋势，为教育决策提供信息、依据和建议，以便教育决策部门调整教育政策，改进教育行为，从而提高国家教育质量。因此，教育质量监测的内容应该包括教育对象自身、教育者自身、教育过程中的重要因素进行监测。一般来说，教育质量监测的主要内容有：学生核心学科素养和跨学科综合能力、非学业部分的素养和能力、影响教育质量的重要因素。