

蒋德仁 编著

Jiang Deren Bianzhu

GUOJI XUESHENG PINGJIA
(PISA) GAISHUO

国际学生评价 (PISA) 概说

蒋德仁 编著

Jiang Deren Bianzhu

GUOJI XUESHENG PINGJIA
(PISA) GAISHUO

国际学生评价 (PISA) 概说

浙江教育出版社

图书在版编目 (C I P) 数据

国际学生评价 (PISA) 概说 / 蒋德仁编著. -- 杭州:
浙江教育出版社, 2012. 11
ISBN 978-7-5536-0125-0

I. ①国… II. ①蒋… III. ①学生—教育评估 IV.
①G449. 7

中国版本图书馆CIP数据核字 (2012) 第255137号

责任编辑 邱连根 责任校对 唐弥尧
封面设计 曾国兴 责任印务 温劲风

国际学生评价 (PISA) 概说

蒋德仁 编著

出版发行 浙江教育出版社
(杭州市天目山路 40 号 邮编: 310013)

图文制作 杭州兴邦电子印务有限公司

印 刷 杭州余杭人民印刷有限公司

开 本 710×1000 1/16

印 张 16.75

字 数 282 000

版 次 2012 年 11 月第 1 版

印 次 2012 年 11 月第 1 次印刷

标准书号 ISBN 978-7-5536-0125-0

定 价 28.00 元

联系电话: 0571-85170300-80928

E-mail: zjjy@zjcb.com 网址: www.zjeph.com

序

认识作者是因为科学教育。同是科学教师,我们常常会为科学教育很愉快地讨论一些话题。因为我也非常关注 PISA,因此很多时候我们之间就有了一些有关 PISA 的共同话题。作者极其聪明,很有思想,也很能坚持。这种品质促使他只要对一件事有兴趣,就会深入其中并持续研究,最终成为这一方面出色的专家。他对 PISA 的认识与研究过程也是这样的。

PISA 可为我国的基础教育提供借鉴。目前我国教育存在的很多问题可能来自评价体系。我们往往用知识的眼光,用知识的理念来编制测量体系,并研制测量所需要的问题或试题。评价体系的偏差导致“指挥棒”的方向产生偏离,这是基础教育出现的一些我们所不愿看到的“异化”现象,这些问题可能成为我国基础教育改革的困难。

PISA 可为我们提供多重营养,改造我们的基础教育。PISA 有一套完整的系统,有学科的理论构建,其对学科目的、学科的基本特征、学科承载的对人的发展功能、学科的测量体系、学科测量方法等都有系统的阐述与架构。PISA 也非常关注学校、家长、学生、教师、社区等与教育直接关联的对象,并建立了一套相关研究的测量评价体系。同时,PISA 的研究也非常重视一个地区或一个国家的教育政策对教育过程、教育方法、教育水平等的影响,并提供国家或

地区层面的教育政策咨询服务。

PISA 可为教育工作者提供教育思想。PISA 可以让我们学习与认识一个学科的学科本质、学科体系、学科测量方式,也能为我们的学科教学提供理论上与实践上的支持。PISA 可为我们对学生的学科能力和学科素养的培养提供途径与方法。因此,学科教师、教学管理者和所有教育工作者都要关注、研究 PISA,吸取其营养并为自己所用。

目前,PISA 的研究材料很多,但多为零散,不利于初学者对 PISA 进行全面系统的了解。本书是作者长期研究 PISA,长期收集与整理,并对材料进行系统分析与整合的结果。字数不多,但看得出作者对 PISA 的总结与归纳有独特之处,也反映了作者对 PISA 的分析与研究水平很高。本书可为读者快速认识、理解 PISA 提供很好的支持。

我了解 PISA,但没有系统研究过,看了作者给我的初稿,感觉很有收获,很有启发。作者多次要我写个序,对于作者的信任我深存感激,但又不好意思以一个门外汉的身份写序。作为朋友,写上一段真心的“大实话”,可作为对朋友的交待,也权当序吧。



2012.8

目 录

序言 / 1

第一章 PISA的基本框架

- 第一节 PISA的基本理念 / 1
- 第二节 PISA的组织方式 / 5
- 第三节 PISA的操作规程 / 7
- 第四节 PISA的评价领域 / 16

第二章 PISA对阅读素养的测评

- 第一节 阅读素养的含义 / 20
- 第二节 PISA对阅读素养的测评框架 / 22
- 第三节 PISA对阅读素养的测评标准 / 27
- 第四节 阅读素养测试样题及分析 / 30

第三章 PISA对数学素养的测评

- 第一节 数学素养的含义 / 69
- 第二节 PISA对数学素养的测评框架 / 70

第三节 PISA对数学素养的测评标准 / 75

第四节 数学素养测试样题及分析 / 77

第四章 PISA对科学素养的测评

第一节 科学素养的含义 / 92

第二节 科学素养的测评框架 / 97

第三节 PISA对科学素养的测评标准 / 102

第四节 科学素养测试样题及分析 / 106

第五章 PISA对问题解决能力的测评

第一节 问题解决的含义 / 178

第二节 问题解决的测评框架 / 179

第三节 问题解决能力的测评标准 / 182

第四节 问题解决能力测试样卷及分析 / 186

第六章 PISA测试结果与启示录

第一节 PISA 2009测试结果成绩序列 / 190

第二节 来自经合组织的测试报告 / 193

第三节 来自欧盟的报告 / 195

第四节 上海PISA 2009结果介绍 / 202

第五节 PISA 2009对浙江测试结果 / 205

第六节 来自《文汇报》的解读 / 209

第七节 对学科教学的启示 / 215

附录 PISA的调查问卷

附录一 PISA 2009学生问卷 / 222

附录二 PISA 2009学校问卷 / 237

附录三 PISA 2009家长问卷 / 249

主要参考文献 / 257

后记 / 260

第一节 PISA 的基本理念

一、时代背景

知识经济的到来,引发了世界范围的教育改革。如何实现对人才培养质量的有效监控,以引导教育的健康发展,这已成为目前世界各国教育关注的焦点问题。评价一个国家教育的总体质量,评价一所学校的教学状况,评价一名学生的基本素养,其价值导向是什么,其评价标准又是什么,以及通过什么方式、方法、手段才能科学有效地进行评价与监控,这些都是世界各国教育研究者在新世纪所面临的共同课题。国际经济合作与发展组织(Organization for Economic Cooperation and Development,简称 OECD)为满足学生学业成绩的跨国比较研究需要,满足社会发展对人力资本的质量监控要求,克服以往教育对学生评价的不足与缺陷,于 1997 年启动开发实施了“国际学生评价项目”(Programme for International Student Assessment),简称 PISA。PISA 主要对即将完成义务教育的 15 岁学生进行评估,测试学生们是否掌握了参与社会所需要的知识与技能,其评估结果对许多国家的教育改革产生了深远的影响。这一有关 15 岁在校生的国际性学生学习质量比较研究项目,从一个全新的角度展开了对人力资本的测量与评价。测量的内容主要是生活所需要的基本知识和技能:阅读素养、数学素养、科学素养以及问题解决能力,重点考查学生运用所学知识和技能解决实际问题的能力。此外,还收集学生的学习态度、学习策略、家庭背景以及学校因素等影响学生表现的背景信息。可以用于

测量义务教育即将结束时,15 岁年轻人走向社会而准备的知识 and 能力情况。

对学生的表现进行国际比较评估的想法在 OECD 由来已久。早在 20 世纪 70 年代,OECD 的绿皮书就开始尝试建立一个教育数据的收集与比较模式。但是直到 80 年代中期,对教育成功指标的测量才变为 OECD 的重点关注领域。1988 年,OECD 在巴黎召开第 38 届年会,开始正式推动国际教育指标体系方案的建立,由 OECD 的教育研究与改革中心(CERI)负责完成。1992 年,OECD 出版了《教育概览:OECD 指标》(*Education at A Glance: OECD Indicators*)一书,提供了其成员国的教育系统在 3 个领域(教育人口、经济与社会背景;教育经费、资源和学校进程;教育成果)的 36 项核心指标上的对比状况,宣布了 OECD 教育指标体系的正式问世。此后,OECD 的《教育概览》逐步发展成年度刊物,同时教育指标体系也在不断革新和完善。

一年一期的《教育概览》大部分涉及的是教育的人力投资和财政资源,教育系统的合作,个人、经济和社会对教育投入的回报等,但一直缺乏对教育产出情况的有效考察。教育产出缺乏国家间共有和可信的指标,特别是在知识和技能方面的测评缺少可信的数据,这使得决策者、纳税人、教育者和家长都希望有一套评价教育体系的有效工具。1995 年,为了满足 OECD 成员国希望获得关于学生知识、技能及教育表现的常规、可靠的数据资料,PISA 项目被首次提出。虽然期间受到一些内部力量的阻碍,但 OECD 还是在 1997 年正式启动了 PISA 项目。从提出设想起,经过五年的研究与试点项目,OECD 最终在 PISA 调查的框架上达成了一致。2000 年,PISA 开始了第一轮测评。此后,每三年举行一轮测评。其新颖、规范、科学的设计与严格的控制标准,引起了广泛关注和强烈反响。

二、基本理念

PISA 评价的重点并不是初中学生掌握了多少学科知识,而是他们在实际生活中创造性地运用这些知识和技能的能力。它着眼于学生的终身可持续发展,为生存而学习是评价的基本理念。基于终身学习的动态模型设计测试,它评价学生现实生活与终身学习所必需的知识和技能。与学校学习科目紧密相关,它集中于学生习得技能的价值和解决问题的能力上,在一个更广阔的范围、在实际生活的情境中测试学生的实际操作能力和文化素质。它考查人们在学习化的社会里所必需的那些知识和技能,这些知识和技能需要通过终身学习才能获得。它衡量的是学生面对实际生活挑战的能力,而不仅仅局限于他们对学校课程的掌握情况。

PISA 是一种终身学习的动态模式和技术需求。学生不仅需要依靠学校的教学,而且还需要了解成人的生活。学生必须组织和规范其自身独立学习及团体学习,克服学习过程中的困难,进行能在未来生活中获得成功的学习。更进一步地说,学习和其他知识的获得,将于实际情境中持续增加。通过多领域、多维度的测试,希望了解即将完成义务教育的各国初中学生,是否具备了未来生活所需的知识与技能,基础教育阶段的学习是否为终身学习奠定了良好的基础。

三、评价目标

在 PISA 启动之前的 40 多年间,主要由国际教育成就评价协会 IEA 和美国教育考试中心 ETS 的教育发展国际评价(IAEP)所从事的研究已有许多成果。但是,这些评价研究关注的是与课程有关的成就,且仅仅评价参与国所共有的课程。

而 PSIA 的评价则是前瞻性的,它测量的是 15 岁青少年对于迎接高科技和知识社会的挑战的准备情况。因此,PISA 更加关注学生运用在校所学知识能够做些什么。PISA 的测量目标是发展常规的、可靠的、与政策相关的学生成就指标,从而达到关于国家教育体制的质量、公正性和效率的评价目标。

PISA 评价关注四个子目标的实现:学习成果的质量;学习成果的等价性和学习机会的均等性;教育过程的有效性和效率;教育对社会经济的影响。PISA 将教育系统分为教育与学习中的个人参与者、教学背景、教育服务的提供者、作为整体的教育系统四个层面进行分析。

PISA 的测评模型就是围绕教育系统的上述四个层面展开的。这四个层面分别从教育成果、政策杠杆、前提条件三个领域入手,不仅关注国家的经济、人口等因素,关注学校的因素,也关注学生个人的家庭、移民、兴趣、爱好等因素。

表 1-1 PISA 评价目标分析表

	教育成果	政策杠杆	前提条件
层面 A: 个体学习	知识、技能的质量与分布	个体态度、投入程度与行为	学习者的社会经济背景
层面 B: 教学背景	教学实施的质量	教学、学习实践与课堂气氛	学生学习、教师工作条件
层面 C: 教育机构	机构的表现与产出	学校的学习环境	社区与学校的特色
层面 D: 教育体制	教育的社会经济成果	结构、资源分配与政策	国际的教育、社会和经济环境

PISA 从社会、文化、经济以及教育因素等方面考查学生和学校的特征。个体学习者层面的因素包括性别、社会背景,以及有关自我的认知、动机因素以及教育期望等。教学设置层面的因素包括教师教学策略、学生知觉到的课堂气氛、教师支持、教材使用以及班级大小、学校组织和结构等因素。学校层面的因素包括学校类型、学校结构、学校资源、学校风气、学校管理以及课堂实践等。

使用来自学生问卷和学校问卷的数据,将背景信息与学生的成就链接起来,可以比较研究下列问题:

- ①各国在学生水平因素与成就之间关系上的差异;
- ②各国在学校水平因素与成就之间关系上的差异;
- ③学校之间的差异,以及各国之间学校差异的差异;
- ④各国在学校提高学生成就效应上的差异;
- ⑤与学生成就差异有关的教育体制和国家背景的差异;
- ⑥通过与 PISA 2000 建立链接,所有这些关系之间的变化情况。

PISA 为各国提供了一个良好的教育质量检测平台,为各国及时调整自己的教育政策提供了非常有用的信息。对于政策制定者而言,通过对比自己国家和其他国家教育系统的成就表现,总结已有政策的经验,改善教育体制,并基于 PISA 提供的指标更好地评价和监控教育体制的效力与发展。

四、基本特性

1. 政策的导向性。PISA 项目的设计和报告方法主要为政府政策决策的需要而设计,以便政府汲取政策上的成功经验。

2. 定义的独创性。譬如关于素养(literacy)概念:关注学生应用知识、技能的潜力,以及面对实际情境时,理解问题、解决问题,分析、推理和有效地交流的潜力。

3. 学习的可持续性。不局限于评价与课程相关的能力,还要求学生报告学习动机。不仅要评价学生的知识和技能,也要让学生报告他们的学习动机、学习策略、自身的信念、对自己的了解和对不同学习环境的态度,以便综合评价学生。

4. 操作的规范性。PISA 的检测工具开发、检测标准设定、检测数据处理和检测结果判断等各环节的操作严密可行、规范有效。PISA 提供的结果使得各国可以监控其教育质量。

5. 内容的全面性。主要是生活所需要的基本知识和技能:阅读素养、数学素养、科学素养以及解决问题的能力,重点考查学生运用所学知识和技能解决日常

生活问题的能力。此外,还收集学生的学习态度、学习策略、家庭背景以及学校因素等影响学生表现的背景信息。

6. 主体的多元性。PISA 突破了传统的单一学校评价方式,形成了多元化的评价主体。PISA 的评价主体包括学生、家长、校长和教师四个方面,从各个方面对学生作了全面的评价。

7. 参与的广泛性。参与 PISA 项目的国家与地区的国民生产总值之和占全世界国民生产总值(GDP)的 90%左右。

8. 取样的参照性。PISA 选取义务教育阶段结束时的青少年作为测试对象,便于各国评价其教育系统的质量。

第二节 PISA 的组织方式

一、指导监管

PISA 管理委员会(PISA Governing Board)是由各 OECD 成员国教育部任命的一位该国代表和 PISA 的各参与国(地区)教育主管部门任命的一位观察员所组成。管理委员会主席由委员会商议决定。管理委员会在 OECD 教育目标的指导下,确定每次 PISA 测评的重点,制定测评指导原则,以保证测评顺利实施。

二、日常管理

OECD 秘书处(OECD Secretariat)负责 PISA 的日常管理,包括监督调查实施的情况、为 PISA 管理委员会管理行政事务、通过跨国交流得出一致意见等。它架起了 PISA 管理委员会和 PISA 国际联合处之间沟通协作的桥梁。OECD 秘书处的现任主席是 Andreas Schleicher,他同时也是 OECD 教育理事会下属的指标与分析部(Indicators and Analysis Divison)负责人。

三、设计实施

PISA 的具体设计和实施由多国机构组成的 PISA 国际联合处(PISA Consortium,亦称国际承包商 international contractor)来负责。该联合处由一些国际大型的测评机构组成,它以澳大利亚教育研究委员会(Australian Council for Education-

al Research, ACER)为首,其他联合机构还包括荷兰的全国教育测量研究所(CITO)、美国的 Westat 公司和教育考试服务中心(ETS),以及日本的全国教育政策研究所(NIER)。

四、发展规模

自 2000 年以来,已经完整地实施了四轮 PISA 测评,包括 PISA 2000、PISA 2003、PISA 2006 和 PISA 2009。PISA 在 2000 年开始第一轮测评,最初有 32 个国家参加,包括 28 个 OECD 国家和 4 个非 OECD 国家。随后,又有 11 个国家和地区加入了 PISA 的测试,共有 43 个国家和地区参与。2003 年进行第二轮测评,除了 30 个 OECD 国家参加外,还有 11 个非 OECD 国家(地区)参与。2006 年进行第三轮测评,有 57 个国家和地区参与;2009 年进行第四轮测评,共有 74 个国家和地区参与,非 OECD 国家(地区)的数量已经超过了 30 个。不过,在 PISA 2009 报告中,只有 65 个国家和地区的结果,因为有 9 个国家稍后才参加 PISA。

越来越多的非 OECD 国家(地区)参与了 PISA 的测试,一方面是因为有些国家(地区)希望从国际的视角了解本国或本地区教育发展的水平;另一方面,也有一些国家(地区)希望通过参与 PISA 学习 OECD 的学生学业测试的组织管理模式,以便为本国(地区)建立长期有效的质量监测体系提供参照。2009 年对 9 岁和 15 岁学生分别评价并将评价结果建立链接,从而检查各国学生的学习进程。

我国教育部考试中心 2006 年引进并启动了 PISA 2006 中国试测研究项目,2006 年 10 月正式启动。PISA 2006 中国试测研究并不代表中国大陆正式参与 PISA。天津、北京和潍坊三个城市一起参加了 2006 年 PISA 在中国大陆地区的试点性测试研究。其中,天津市共有 50 所学校的 1700 多名学生参加了最终测试。

2009 年 4 月 17 日,PISA 正式在上海开始测试。5000 名在 1993 年 1 月 1 日至 1993 年 12 月 31 日期间出生的学生参与测评,分布在 150 余所中学。PISA 要求各国(地区)的样本范围在 4500~10000 名学生之间,学生来源于 150 所中学以上。我国的学生和学校的样本数与取样要求完全符合 OECD 的要求。

PISA 很重视试题所具有的跨文化性和跨国跨地区性。在 PISA 2006 中,国际专业协作组织委托五个国际知名的研究机构设立国际命题专家组负责 PISA 科学的命题,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)、OSLO 大学(ILS,挪威)、KIEL 大学(IPN,德国)、日本国立教育研究所(NIER)。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,也对各参与国家和地区提交的试题进行评审和修订,保证 PISA 试题的质量。其中,澳大利亚教育研究理事会根据 PISA 专业协作组织的管理和科学素养框架的要求,从整体上协调并监督各命题专家组的工作。

2. 命题程序。

PISA 从命题到最后确定试题主要有两个阶段。第一阶段是 PISA 国际命题专家对各参与国家和地区提交试题的审查,第二阶段是对所有提交试题的国际审查(如图 1-1 所示)。

一般而言,由各参与国家和地区提交的试题都必须经过至少一个国际命题专家组的审核和修订,并由此专家组将经过筛选或修订后的试题提交至 ACER,进行进一步的审核、筛选和修订(过程如图 1-1 所示)。各国际命题专家组提交的试题也必须至少由另一个国际命题专家组进行审核,再与各参与国家和地区提交并经过审核的试题一起,共同汇集到 ACER,开展国际性的审核和修订,并筛选出试测所用的新试题。最后,筛选出来的新试题与确定的连接试题一起作为试测试题,在所有的参与国家和地区中试测,并利用试测结果的数据分析,以及各参与国家和地区对试测结果的反馈,再次筛选或修订试题,确定最终正式测试所用的试题。

3. 命题特点。

(1) 多方命题,专家反复审核修改。无论是 PISA 国际命题专家组提交的试题单元,还是各参与国家或地区提交的试题单元,都要经过本地的审查和试题的实验性认知活动(cognitive laboratory activities)。每经过一个阶段,就要对试题进行筛选、剔除,或作进一步修改和完善,这样的审核修改至少有 4 次,最多可达 7 次,保证每一道试题都能经过不同专家的反复审核和修改。

(2) 进行试测,并对测试数据进行统计分析。无论是本地的实验性测试、由 ACER 开展的国际实验性测试,还是所有参与国家和地区都参加的试测,都需要将测试结果录入,进行统计分析,得出试题的功能性差异等指标,并以此为基础对试题进行删减或进一步修改完善,确保正式测试所用试题的质量,尽可能保证试

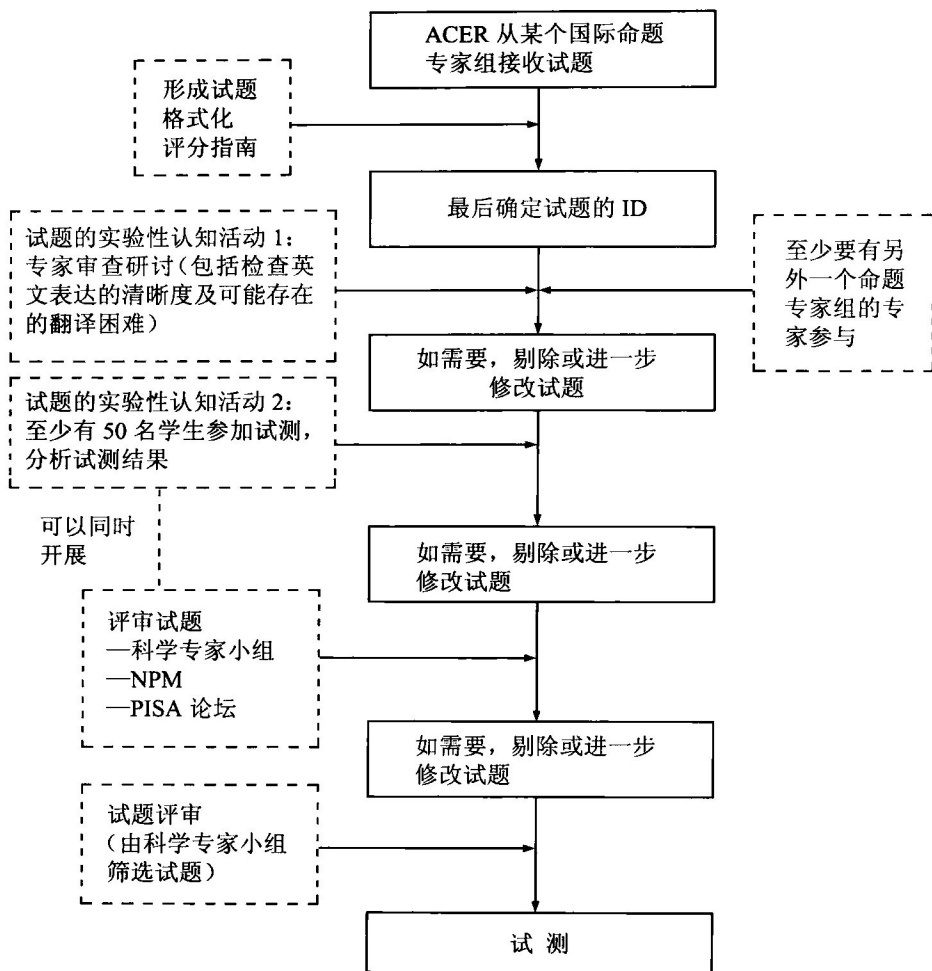


图 1-1 PISA 命题程序

题对所有参与国(地区)学生的公平性。

4. 评分标准的制定。

PISA 试题命题的过程,也是评分指南逐步完善的过程。在题目的最初设计中,PISA 就要求将每道试题的评分指南草稿列入其中;随后,在实验性认知活动中,通过与学生座谈等,收集并整理世界各国学生对试题的回答;最后,将这些回答与测试框架相结合,确定试题的评分要点描述和相应的学生回答案例。

PISA 测试共有五种题型,包括选择题、封闭式问答题、简答题、复合选择题以

(3) 家长问卷。

家长问卷内容一般包含基于父母的特点、孩子过去的阅读参与度、父母自己的阅读参与度、家庭阅读资源和支持、父母的背景、父母参与学校的观念、父母选择学校的范围等。

PISA 2009 的学生问卷、学校问卷和家长问卷请参见附录。

三、学生抽样

在抽样方法上,PISA 的目标总体采用了基于年龄的界定。PISA 选取的样本是测试时年龄在 15 岁 3 个月到 16 岁 2 个月之间的在校学生,不管学生在哪个年级或哪种教育机构就读,也不考虑他们接受的是否是全日制教育。但是,不包括未在教育机构就读或在国外上学的 15 岁学生。每个国家(地区)所制定的特定样本设计和样本大小都是为了学生层面估计值的抽样效率最大化。

PISA 评价项目所采用的取样设计为两阶段分层取样,即首先随机抽出样本学校,继而在样本学校中随机抽出样本学生。

为了保证取样的精确性,要求必须从至少 150 所学校中抽取 4500 个以上的样本容量。第一阶段的取样单元为含有 15 岁学生的个体学校,以随机抽样的方式在每层抽取学校,各层的抽取率与各层学校多少及学校内的学生数量成比例。第二阶段的取样单元为样本学校中的学生,从每一所样本学校中等概率地抽取 35 名学生(底线为 20 名)。并且,在第一阶段选取的学校中,学校参与率要求达到 85%,所选取考生的作答率要求达到 80%(底线为 50%)。如果初始选取学校的作答率低于 85%,那么,必须通过从候补学校中取样来达到要求。为了保证总体覆盖率,在学校水平和学生水平的总的排除率必须低于 5%。

四、具体施测

在 PISA 的具体测试中,各参与国(地区)政府要任命一个国家项目经理(PISA National Project Manager)来负责监督每个参与国(地区)的评价实施情况。PISA 的测评形式为纸笔测试,测试均在学生所就读学校内进行。

测评包括两部分:素养测试和个人背景与学习情况问卷。每个学生的素养测试时间为 2 小时,但由于有不同的题本组合,因此总测评时间加起来一共约 390 分钟。在每位学生素养测试的 2 小时内,每轮测评重点领域(如 2009 年为阅读素养)的测试时间占总测试时间的 2/3。此外,PISA 2009 年还加入了额外的大多数国