

编者按:国际学生评估项目(简称 PISA)测试于 2009 年 4 月 17 日在上海市 152 所样本学校同时进行(相关报告将于 2010 年底由 OECD 集中发布)。这是中国大陆地区首次参加 PISA 测评。2009 年 10 月 22-24 日,上海 PISA 项目秘书处和上海 PISA 研究中心组织召开了“2009 年国际学生评估项目(PISA 2009)工作总结会议”,就相关问题进行交流研讨。上海市教委领导、市教委上海 PISA 项目组成员、区县教育局 PISA 项目联系主任、区县招办负责人、评分专家代表、校长代表等 40 余人应邀出席会议。

会上,上海 PISA 项目秘书长、SHPISA 研究中心副主任陆璟对上海开展 PISA 测评工作的情况进行了总结,市教育考试院和市教委教研室负责人员分别对考务工作和评卷工作进行了总结,并讨论了研究报告的撰写和宣传事宜。市教委张民选副主任和基教处颜慧芬副处长在会上发言,着重阐明了上海参与 PISA 的价值和目的、参与 PISA 的感想及对 PISA 后续研究工作的期望。

本刊将刊发陆璟副主任的总结报告,同时选编会议的部分要点,以飨读者。

PISA 2009 上海实施报告

陆璟(上海 PISA 项目秘书长、SHPISA 研究中心副主任)

2009 年国际学生评估项目(简称 PISA 2009)测试于 4 月 17 日在上海市 152 所样本学校同时进行。这是上海市第一次参加 PISA 正式测试,抽样、测试实施、测试工具等都严格按照 PISA 技术标准执行。

为配合此次测评工作,上海市教委成立了 SHPISA 项目领导小组和项目组,负责全市层面的统筹协调。项目组成员单位包括市教委 5 个相关处室和 4 个直属机构。

SHPISA 秘书处和研究中心设在上海市教育科学研究院。SHPISA 秘书处负责执行项目组的决策,落实计划、组织和协调工作;SHPISA 研究中心负责试卷、问卷、考务手册、评分标准和评分管理手册的翻译和修订,测试主任和学校主考培训,学生抽样,数据管理,以及组织开展 PISA 相关课题研究工作。上海市教育考试院负责考务管理,包括试卷、问卷、宣传彩页及考务证件印刷,学生纪念笔制作,评卷管理,数据输入。市教研室负责评卷专家聘任及各学科领域的评卷培训,并在试题评审和命题研究课题中承担了主要工作。

PISA 2009 项目实施周期历时四年。2007 年开发测试工具,2008 年进行试测,2009 年正式测试,

2010 年底 OECD 将发布 PISA 2009 的国际报告。

一、PISA 测评的理念、框架和测评工具

PISA 主要测评的是,即将完成义务教育时青少年运用知识和技能迎接现实生活挑战的能力。这一取向具有前瞻性,反映了学校教育目标和课程目标本身的变化,即越来越多地关注学生能运用他们在学校里学到的内容做什么,而不仅仅看他们是否掌握了特定的课程内容。

PISA 测评的主要特点就是考查应用能力。阅读素养重在为了学习而阅读,而不是为了阅读而学习,因此并没有对学生最基本的阅读技能进行测评。换言之,阅读素养重在评价学生运用所掌握的阅读技能学习各领域的新知识及理解实际生活中可能遇到的各种文本内容的能力,而不是考查学生是否掌握基本的读写能力。数学素养要求学生从各种现实情境中识别出数学问题,用数学方法表述解题过程和结果并解释应用。科学素养要求能够理解科学概念,应用科学观点来解释现象,并且对证据进行科学思考来得出结论。

这种应用能力体现在各种实际生活情境中,因

此评价框架包括了学生实际生活中可能遇到的各类情境。通俗地说,也就是环境或背景,包括了个人、社区、学校教育、工作等各种环境。测试内容以单元来组织,每个单元前都有一段、一篇或一组引导材料,引入问题情境,接着才是若干个题目。

除了反映学生可能遇到的各种生活情境外,PISA 测评框架还要体现相关学科领域的知识体系内容,以及学生在完成该领域学习任务时所需要的各种能力。

所以,PISA 认知领域的测试框架包括情境、知识领域、能力三个方面(如表 1 所示)。

表 1 PISA 2009 认知领域测试框架概述

	阅读	数学	科学
情境	文本的使用场合: 个人的(如私人信件) 公共的(如官方文件) 工作或职业的(如报告) 教育的(如与学校相关的阅读)	数学应用的领域,关注与个人、社会和全球情境中的应用,例如: 个人的 教育或职业的 公共的 科学的	科学应用的领域,关注与个人、社会和全球情境中的应用,例如: “健康” “自然资源” “环境” “危机” “科学和技术前沿”
知识领域	阅读材料的形式: 连续文本,例如:记叙文,说明文,议论文 非连续文本,例如:图表、表格和清单	数学领域和概念群: 数量 空间和形状 变化和关系 不确定性	科学知识: “物质系统” “生命系统” “地球和宇宙系统” “技术系统” 关于科学的知识: “科学探究” “科学解释”
能力	阅读任务或过程的类型: 检索信息 解释文本 反思和评价文本	数学“能力群”: 再现(简单的数学运算) 连接(运用多个观念解决直接的问题) 反思(更广的数学思维)	科学任务或过程的类型: 识别科学议题 科学地解释现象 运用科学证据

除了认知领域的测试外,PISA 还通过问卷调查收集学生个人、家庭和学校的背景信息,分析影响教育质量的因素,为学校教育系统的改进提供政策建议。学生问卷内容涉及他们的家庭社会经济背景、阅读参与度和阅读策略等。校长问卷主要是关于教师状况、学校资源状况、学校风气和学校政策等内容(如表 2 所示)。每个学生要完成 1 套需时约 35 分钟的学生问卷,每个样本学校校长要完成需时 30 分钟的学校问卷。

表 2 PISA 2009 问卷调查框架概述

学生问卷	学校问卷
• 阅读参与度 • 阅读策略 • 家庭社会经济背景 • 学习时间 • 课堂气氛和学校风气 • 语文课的教学组织情况 • 图书馆资源利用情况	• 学校的结构和组织 • 学生和教师 • 学校资源 • 学校课程和评价 • 学校风气 • 学校政策及其实施

阅读是 PISA 2009 的主要测评领域,共有 132

个试题,总测试时间为 3.5 小时。数学和科学是次要领域,数学共 36 个试题,测试时间为 1.5 小时,科学共 53 个试题,测试时间为 1.5 小时。所以总的覆盖题量是 6.5 小时。这些试题组合成 13 套试题本,每套试题本所需测试时间均为 2 个小时。

二、抽样方法概述及样本分布

PISA 的测评对象为参加测试时年龄在 15 周岁 3 个月到 16 周岁 2 个月的 7 年级及以上在校生,简称 15 岁学生。在 PISA 最初的准备阶段就确定了上海 PISA 测评的时间是 2009 年 4 月 17 日,因此确定参加测评的对象是在上海市学校就读的、1993 年出生的所有 7 年级及以上学生。这里“上海市的学校”包括除了国际学校以外的所有在上海举办的学校,学生中包括了外来务工人员子女和外籍学生。有智力障碍和严重功能性残疾(例如盲、聋、哑)的学生以及学习汉语不足一年的学生可以在最后的抽样阶段剔除,但在上报抽样框数据时不能剔除。

PISA 采用按照概率与抽样元素的规模大小成比例的抽样法(简称 PPS 抽样)。抽样前必须要确定抽样的外显分层变量和内隐分层变量,分层的目的是为了等概率地抽取各学段各类不同教育质量的学校。外显分层变量最多可以有三个,SHPIISA 研究中心确定的变量一是学段(初中、高中、完中),变量二是学校类型(普通、职业),变量三是学生入学成绩等级或学校质量(实验性示范性、一般)。用这三个变量把学校分为六类:初中及九年一贯制学校的初中部、完全中学、示范性普通高中、一般普通高中、重点中等职业学校、一般中等职业学校。此外,还可以根据需要设定若干内隐分层变量。上海采用的变量一是公办与民办,变量二是城市与农村,变量三是考虑到有一部分中等职业学校中包含综合高中,因此区分了纯中等职业教育学校、含综合高中教育的中等职业教育学校、不适用此分类的其他学校。

抽样分为学校抽样和学生抽样两个阶段。第一阶段为学校抽样。首先由 SHPIISA 研究中心统计上报全市每一所有 15 岁在校生的学校中 15 岁在校生数,然后由 PISA 国际专业协作组织之一的 Westat 用 PPS 抽样法抽取学校;第二阶段为学生抽样,要求第一阶段抽中的学校上报所有 7 年级及以上的 15 岁学生名单,由 SHPIISA 研究中心运用 KeyQuest

软件从每所样本学校中随机抽取 35 名学生,不足 35 名的则全部抽取。

为了保证各项数据可以进行有效的国际比较,PISA 规定了样本量的最低要求,每个参与国家或经济体至少要抽取 150 所学校并从这些学校中至少抽取 4500 名学生。抽样时还要按估计的应答率(response rate)适当增加样本。也就是说,如果估计应答率为 90%的话,就需要抽取 5000 名学生。上海抽取的样本为 152 所学校的 5226 名学生。实际参加测试的学生共 5115 名,各类学校参加测试的学生比例均符合全市 15 岁在校生总体分布。

表 3 上海市参加 PISA 2009 测评的学校和学生数分布

	学校数(所)	参加测试学生数 (人)	参加测试学生占比 (%)	全市 15 岁在校 生占比(%)
初中	52	1716	33.5	33.0
完中	27	923	18.0	18.2
实验性示范性高中	20	689	13.5	13.2
一般普通高中	20	695	13.6	13.2
重点中职校	29	975	19.1	19.5
一般中职校	4	117	2.3	3.0
合计	152	5115	100.0	100.0

不仅学校类型分布符合全市总体比例,公办民办学校分布比例、城市农村分布比例也与全市 15 岁学生总体比例一致。参加这次 PISA 测试的学生中,公办学校学生占 89.9%,民办学校学生占 10.1%;9 个中心城区学生占 39.5%,10 个郊区(县)学生占 60.5%。

学生年级分布比例如下表所示:

表 4 上海市参加 PISA 2009 测评的学生年级分布

	参加测试的学生数(人)	参加测试的学生数占比(%)
7 年级	52	1.0
8 年级	208	4.1
9 年级	1924	37.6
10 年级	2908	56.9
11 年级	22	0.4
12 年级	1	0.0
总计	5115	100.0

此外,参加测评的学生中男生占 49.4%,女生占 50.6%。所有学校都超过国际组织要求的 80%最低应答率,无需组织补考。

三、相关启示

1. 关于评价理念

PISA 是第一个在终身学习能力评价方面有重大突破的项目。它创造性地提出了“素养”的概念,重视评价各种情境中的实际应用能力,即有关学生在主要学科领域应用知识和技能的能力,以及在不同的情境中提出、解决和解释问题时分析、推理和有效

交流的能力;它强调对终身学习的相关性(relevance,即有用性),不仅评估学生的课程及跨课程能力,同时也要求学生报告自己的学习动机、自我信念和学习策略。

我们在实施素质教育的过程中,学生创新能力和实践能力的评价是难点,为此,可以借鉴 PISA 对情境的分析以及结合各种情境评价学生应用能力的做法,教师和教育评价工作者要加强对知识应用情境的分析和研究。在新课程三维目标的评价中,“过程与方法”、“情感态度与价值观”的评价是难点,对此,也可以借鉴 PISA 对学习策略和学习参与度的研究方法。

2. 关于评价目的

PISA 是一项以证据为本的政策研究,以改进教育政策为评价目的,而我们现有的教育评价多数是以评价学校教育质量和学生学习质量为目的。PISA 将学生学习结果的数据与学生个人特征数据以及学校内外影响他们学习的关键因素联系起来,目的是使人们注意到成绩类型的差异,并识别出那些成绩达到高标准的学校和教育体系的特点。而且每三年进行一次,数据之间可以跨年度比较,这种长期性和定期性使国家(经济体)能够监测它们在达到关键学习目标上的进步。

运用 PISA 研究结果可以使在教育政策制定过程中,用经过科学研究得到的数据说话,并且能综合考虑多种因素,减少决策过程的主观性和片面性。

3. 关于理论研究和研究方法

PISA 以课程实施理论、学校学习和学生学习理论、终身学习能力研究、学生态度和参与度研究、学生认知策略和元认知测量研究等多项理论研究为基础,并结合 OECD 多年来积累的多水平、多层次教育指标分析,在教育理论的宏观、中观、微观层面应用上为我们提供了一个范例。

PISA 也是综合应用统计测量技术的一个范例。PISA 抽样应用了 PPS 抽样方法,在试题组卷中采用了矩阵抽样法,因此只要少量的学生、每个学生做一部分试题就可以达到测试的目的,这样可以大大减轻学生的负担。PISA 在试题分析和能力量表制定中应用了项目反应理论,在结果数据分析中采用了多层线性模型分析等国际上前较前沿的数据分析方法,能够更加科学地分析数据。

4. 关于系统设计和过程管理

PISA 是一项大型系统工程,首先是设计测评框架;接着根据框架设计测评工具,包括试题本和问卷,以及包括项目管理、翻译修订、抽样、数据管理、考务、评分等各方面的实施手册,用精细化管理确保测试达到 PISA 技术标准的要求;之后组织试测,对试测结果进行技术分析,用于筛选和修订试题;其后再组织正式测试;最后得出研究结果和政策启示;根据研究结果再改进下一轮的研究。

上述过程中的各个环节是环环相扣的整体设计,一开始就考虑到后面的需要。例如设计认知测试的时候不仅要全面均衡地反映测评框架各个维度的要求,还要考虑下一个环节的试题分析技术方法,问卷的变量设计也要充分考虑多层统计的需要。又如 2007 年试测准备时不仅要确定 2008 年试测的时间,也要同时确定 2009 年正式测试的时间,以避免试测学生和正式测试学生年龄交叉重复,等等。

先进的理念、科学的方法、系统的设计都需要严格的过程管理来落实。PISA 2009 是有 60 多个国家和地区参与的大型研究,如果没有严格的程序规范,实施过程就会产生多方面的不一致性,影响数据的可比性。因此,PISA 不仅制订了严格的技术标准,还编制了一系列操作手册(有 40 本左右),并定期组织培训和工作会议,定期要求上报在线表格,以规范实施过程中的具体做法。

总之,PISA 是一项具有前瞻性的评价项目,我们希望通过参与 PISA,从国际的视角来审视上海学生是否为终身学习做好了准备;PISA 是一项在理论、技术、管理上具有先进性的系统研究,我们希望能借鉴 PISA 的理论、技术和管理方法,结合上海的实际情况,建立科学的质量监测体系;PISA 是一项以改进教育政策为目的的实证研究,我们希望借鉴这一视角,加强教育决策的科学性。

我们从 PISA 测评中学到了什么

张民选(上海市教委副主任、上海 PISA 项目负责人、上海师范大学教授)

2009 年国际学生评估项目(PISA 2009)测试,有很多单位参与了相关工作,包括教委的 5 个处室、4 个直属单位,19 个区县的教育局招办,以及 150 多所学校,还有很多评卷专家和质量监测专家,其中测试当天参加考务工作的领导和管理人员就达 375 人。为了一个科研项目,投入如此多的人力,在上海历史上可以说是很少见的。PISA 是一个要坚持很多年的项目,从 2006 年底上海市教委决定要参与 PISA,到 2007 年、2008 年做准备,整理和翻译大量资料,进行试测,2009 年正式测试,2010 年要出报告,之后还要为 PISA2012 的数学(主项)测试做准备,到 2015 年进行科学(主项)的测试。用这么长的时间、投入这么多人力去参与一项测试,是一件很不容易的事。

或许有人会心存疑问:用这么多的心血、投入这么多钱、花这么长时间开展 PISA 测评,究竟值不值得?关于这个问题,我想说的是:中国大陆,只有上海

正式参与了 PISA。PISA2009 共有 68 个国家和经济体参加,这些国家和经济体代表了世界 GDP 总量的 90%。虽然全世界大大小小的国家有近 200 个,联合国成员有 192 个,但参与 PISA 测评的国家和地区数量却很有限,几乎所有大国都参与其中。这里要思考的是,我们究竟可以从 PISA 中学到什么?我觉得从开始参与 PISA 到现在,我们确实学到了很多。

首先,一项测评可以完成多个目标。这在以前是没有的。之前很多测评都是所有人共同参加的,只为完成同一个目标,就是看学生学得好不好、这所学校所谓的质量高不高。一项测评解决一个问题,大家都很辛苦,而这次 PISA 测评同时要完成多项任务,包括评估学生的质量、教育的公平、学生的参与、政府的投入等。所以,如果一项测评组织得好,就应该可以同时实现多项目标。我们之前的想法是,如果是一个高相关的选拔性测试,就一定不能完成水平性的测试;而一个水平性的测试,就一定不能完成选拔