



中国发展研究基金会

研究参考

第 2 号 (总 177 号)

2014 年 2 月 10 日

经合组织的“国际学生评估项目”概述

内容摘要：“国际学生评估项目”(PISA)是经济合作与发展组织(OECD)发起的大规模跨国学生测试,旨在通过测量 15 岁学生的知识和能力水平来评估各国的教育体系,以促使其不断改善。上海自 2009 年参加 PISA 测试后,先后两次夺冠,引起广泛讨论,褒贬不一。不少人认为这只是反映出中国教育的“应试性”特点而已,并不能证明中国的教育水平也雄居全球。这种说法并非全无道理,但也反映出人们并不太了解 PISA,将其与传统的国际测试等同视之。为此,本文概述了 PISA 的基本理论、设计框架、主要特点以及由此衍生的相关研究成果,并对上海参加 PISA2012 年的成绩进行了简要分析。

关键词：教育；学生评估；测试；

经合组织的“国际学生评估项目”概述

一、项目概述

作为一个全球性国际组织，经济合作与发展组织（OECD）在教育、经济、社会和公共治理等领域的研究为许多国家（特别是其成员国）的政策制订发挥了重要的参考作用。在学生测试方面，鉴于以往的国际测试都关注了学生在校习得知识的情况，具有一定局限性且不适用于跨国比较，很难对学生的实际应用能力进行可靠的测量，所以 OECD 于 1997 年至 2002 年启动了大规模跨国研究“能力的界定与遴选：理论框架与概念基础”（Definition and Selection of Competencies: Theoretical and Conceptual Foundations, 简称 DeSeCo），并以此为基石建立起“国际学生评估项目”（Programme for International Student Assessment, 即 PISA），旨在通过测试 15 岁学生的知识和能力水平来评估各国的教育体系。

PISA 的设计初衷是服务于 OECD 成员国，评估他们的青少年学习质量和国家教育体系，但随着其发展，许多非成员国和地区也不断加入项目。目前，PISA 是全球规模和影响力最大的学生评测项目之一，先后有 70 多个国家和地区参与其中。中国尚未以国家身份正式参加测评，但上海参加了 2009 年和 2012 年的 PISA 测试，均拔得头筹。

PISA 之所以将测试对象锁定为十五岁的学生，是因为在绝大多数国家，这个年龄段的学生都是即将完成义务制教育的初中毕业生。PISA 的设计者认为这些学生不可能在义务制教育阶段学到人生必需的所有知识和能力，但应该掌握能帮助他们融入社会和终身学习的知识和技能。为此，PISA 关注了学生现在和未来的生活所需的基本数学素养、阅读素养和科学素养。

自 2000 年开始，PISA 每三年测试一次，从各个参与国家（地区）随机抽取 4500 到 10000 名学生参加书面测试。自 2012 年起，OECD 开始采用电子答题系统，但由于参与国的条件所限，所以新型测试方式尚未完全取代书面测试。每次 PISA 测试选择三种素养中的一种为

考察重点，其余两种素养为辅。2000年，PISA的主测试领域为阅读素养，2003年为数学素养，2006年为科学素养，随后依次循环。有时PISA测试也会增加额外的跨课程测试，如问题解决能力、金融素养和计算机辅助阅读等。测试时长为两小时，测试题型以开放型问题 and 多选题为主。测试结束后，应试学生和学校老师、校长还要填写问卷调查来提供相关背景资料，如家庭的社会经济情况、学校的教学质量和公平性等。有时学生家长也会参与此调查。

二、PISA的组织流程和特点

PISA的决策机构是PISA理事会，由各参与国的政府代表构成。在OECD的组织下，理事会每年召开两次会议，讨论并确定教育政策研究的优先领域和PISA的发展方向。每一轮PISA的研究框架都围绕理事会最终确定的政策优先领域来展开设计。所有参与国专门成立PISA项目管理中心，指定项目管理负责人从事PISA在本国的执行和协调工作。

PISA是一个合作性流程。它汇聚了OECD顶尖专家、参与国的高层教育官员和专家以及专业试题开发公司的集体智慧，制定出在不同国家的教育和文化背景下都具有可比性且科学严谨的一套评估体系。PISA的设计框架将“能力”定义为“调动知识、技能、动机、态度、情绪和其它社会和行为能力等社会心理资源，在各种情境下成功完成复杂任务的能力”。正是这个定义使得PISA从设计理念、执行流程到结果分析、政策影响都区别于传统测试。它不像传统测试那样将测试重点放在评估学生学到多少课本知识，而是强调了他们从自己学到的知识推断出新知识，或将学到的知识应用于新情境的能力，因为这才是现代知识型社会所需要的“能力”。由于各国对教育结果的预期不同，而且很难在大规模测试中面面俱到地照顾到所有有助于成功的能力，所以PISA特别选择了阅读、数学和科学这三种最可能预测学生未来的素养，将它们作为主要的测试领域。当然，这并不意味着其它能力无足轻重。

PISA的设计理念和设计目的使它具备了许多不同于常规学生测试的特点，如：

1. 前瞻性。PISA 与学校课程没有直接关系，而是用真实情景来考验 15 岁青少年在阅读、数学和科学这三个领域内应该掌握的知识内容和结构以及运用这些知识的技能，从而反映出他们迎接高科技社会和知识社会的准备情况，和他们的终身学习能力。PISA 特别强调真实情景中的知识运用和思维能力，不会测试需要死记硬背的知识，所以参加 PISA 的学生只需要理解基本概念，灵活运用他们已经掌握的知识的能力，而无需进行特别准备。考试结果不用传统的及格或不及格来衡量，而是用个人/总体在一个连续量表中的位置来确定他/他们相对于其他个体/总体的水平。

2. 可比性。PISA 认为具有国际可比性的标准化测评对于教育的发展至关重要，因为它可以有效地预测一个教育体系的成败。没有比较，学生、学校及至整个教育体系就很难察觉自己的体制优势和劣势。所以国际可比性是 PISA 的设计目标之一，它主要通过两方面的工作来实现这个目标。其一，PISA 以教育、个人能力和测试等领域的先进理论模型和国际通行标准作为研究框架的设计基础，设计出普遍适用的基本测量指标和背景指标，从而确保了开发出来的 PISA 试题和测试流程不仅能够可靠有效地评估学生，而且可以得到完全可比较的测试结果和背景信息。譬如它依据《国际教育标准分类法》（International Standard Classification of Education，即 ISCED）对教育资格和学习项目进行分类，依据《国际职业标准分类法》（International Standard Classification of Occupations，即 ISCO）记录学生家长职业的社会经济情况。其次，OECD 发挥了自身作为国际组织的协调作用，组织国际一流专家以及参与国政府全程参与了 PISA 的设计和执行。参与国提出的政策优先领域和反馈可以及时得到了体现，而且各国的项目管理中心也保障了 PISA 设计团队和本国政府之间的沟通，从样本抽取、测试管理、数据收集和分析等都严格遵守规定流程。

3. 科学性。首先，PISA 制定精细的技术标准和实施手册。为了保证 PISA 结果的真实性，OECD 制定了十多个方面的技术标准，对组织机构、工作人员、项目工作会议和培训、抽样、测试语言、测试参与率、翻译、考场、保密工作、材料的印刷、评卷、数据提交、数

据管理、材料保管等工作做了严格细致的规定。PISA 国际专业协作组织开发了系列实施手册和工作表格，各参与国（地区）与 OECD 及 PISA 国际专业协作组织之间要签订多种严格的协议或计划。这些标准、手册、表格、协议和计划是保证 PISA 实施质量和数据可靠性的关键。每次 PISA 测试结束后，OECD 都会发布《技术报告》(Technical Report)，汇报测试中的所有技术细节。其次，由于 PISA 试题的研究框架和设计方向是由 OECD、参与国和专业试题开发公司共同完成且经过信度和效度的反复筛选，所以最终进入题库的试题在评估背景、知识内容、能力和难度上分布平衡，风格、质量和难度要求也保持一致。项目组尽可能用严格的管理程序和测量技术保证试题的规范性、科学性、公平性和可比性，减少命题专家个人因素的影响。

4. 政策导向性。PISA 的首要目的是为政策服务而不是筛选学生。它希望通过国际比较来展示各种教育体系的优劣，从而促进决策者改进教育政策，提高教学的质量、公平性和效率。一方面，PISA 对各国的成绩进行排名，便于各国进行横向比较，了解本国学生的学习质量相对于其它国家的位置。另一方面，由于 PISA 每三年定期举行，且已经举行了四期，所以各国可以通过考察测试成绩的纵向变化来观察本国的教育质量发展趋势。如果将测试结果和背景调查收集的信息综合考察，政府还可以分析哪些因素可能影响到学习质量，如何改进教育政策来提高现状。

从总体上来说，各国政府对 PISA 的政策反应分为两类：无动于衷和 PISA 冲击 (PISA shock)。当测试结果高于政府预期或与政府预期相符时，PISA 对政策的影响力不明显。如芬兰和新西兰在 2000 年和 2003 年的测试中遥遥领先，政府可以安心地坚持现行教育政策和理念。而美国国内有不少学生评测项目已经让公众认识到本国的教育结果并不理想，因此虽然美国的 PISA 测试结果平平，但也没有对公众和政府造成太大触动。对于测试结果低于预期的国家来说，PISA 的测试结果则“冲击”了它们的教育政策和体系。在 2000 年的测试中，德国和丹麦的测试结果不尽如人意，引发了广泛的公共讨论，促使政府开展重大教育改革，如修订国家教育标准，改进国家教育评估体系，帮助贫困儿童和移民儿童等弱势群体。日本原本在 2000 年的

测试中名列前茅，但在 2003 年的测试中成绩下滑，使民众和政府认为本国存在教育“危机”。教育部甚至为此改变了之前的“宽裕教育”（Yutori Education）政策和国家教育评估体系。

PISA 测试对各国教育政策的具体影响可归纳为以下几个方面：借鉴 PISA 排名较高的国家的教育经验以改善本国政策法规；在课程设置中更注重提高学生的实用技巧、推理能力和解决问题能力；提高本国的学生评测标准或完善测试机制。

三、PISA 的研究结果

在过去的十二年里，PISA 测试收集了大量可靠翔实的标准化数据，为研究工作提供了丰富的材料，与之相关的研究包括以下几个领域：

1. 社会公平与教育质量

各轮 PISA 研究用“PISA 经济社会文化地位指数”（PISA Index of Economic Social and Cultural Status, 即 ESCS 指数）作为家庭社会经济地位的测量指标，研究了它与测试成绩之间的关系。ESCS 指数包括三个方面：父母亲的最高教育程度（即选择父母中教育程度更高者为指标）、父母亲的最高职业地位（即选择父母中职业地位更高者为指标）、家庭拥有物。PISA 希望通过这种研究探求生源的经济背景、学校管理和教师授课以及教育政策分别在多大程度上影响学习质量。历次 PISA 测试的结果表明，社会经济背景对学生的教育结果产生重要影响。家庭社会经济地位较高的学生学习质量普遍较高，反之亦然。如 2012 年的测试表明，在 OECD 国家，社会经济条件较好的儿童比劣势儿童的成绩高出 39 分，几乎相当于多上了一年学，这充分印证了相关理论研究的结论。但与此同时，PISA 测试也表明这个结论并非绝对。2012 年的测试还显示，在 OECD 国家有 6% 的学生超越了自身经济社会条件的制约，他们的 PISA 成绩高于根据家庭社会经济地位预测出的水平。

PISA 还将排名高低与教育机会分配的公平性结合起来，把参与国分为四类：排名高且公平程度高、排名高但公平程度低、排名低且

公平程度低、排名低但公平程度高。2006 年的测试结果表明，韩国、荷兰、香港、加拿大和荷兰等国家在取得优秀成绩的同时实现了教育机会公平，为各国确立了标杆，而葡萄牙、意大利和希腊等国家尽管教育机会的公平性很高，但并没有如预期那样在 PISA 排名中居于前列。

2. 成本效益与教育质量

PISA 的研究结果发现，国家经济发展水平较高意味着它有可能为教育投入更多资金，但这并不一定能确保它的教育质量，即不一定在 PISA 测试中取得非常出色的成绩。2009 年的测试结果表明，国家经济发展水平与 PISA 排名之间并不存在线性关系。人均国民收入超过 12475 美元后，就无法再用这个指标来预测该国的 PISA 排名。而人均国民收入低于 12475 美元时，即该国属于中低收入国家时，国家经济水平对于 PISA 的成绩影响显著。

由此引出了教育的成本效益问题。PISA 对 2006 年的测试成绩和生均教育经费进行了比较，发现韩国和日本的教育投入只达到中等水平，但是最终成绩却显著高于平均水平。这暗示着单纯增加教育投入并不一定能提高教育质量，还需加强教育经费的分配和使用效益。

3. 影响学习质量的因素

PISA 从学习者、教学、学校和教育体系等层面分析了可能影响学习质量的因素。个人层面主要采用了学习参与度评估，由心理和行为两个层面的因素组成，前者与兴趣、动机、信念、情感等学习态度有关。后者指积极投入学习活动或学习实践的程度。PISA 的测试结果证明动机、信心和干劲是学生发挥学习潜力的关键。参与度差的学生无法充分利用自己得到的教育机会。

教学设置层面的因素包括教师教学策略，学生感知的课堂纪律气氛、教材使用以及班级规模、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。2012 年的测试结果也再次确认了良好的学习环境、公平的资源配置、学校自我评估、不根据成绩和能力对学生进行分类等措

施对于提高学生的教育质量至关重要。

PISA 还分析归纳了测试结果和收集的背景信息，认为成功的教育体系应具备七个特点：相信学生可以习得能力；整个教育体系拥有共同的目标且目标清晰远大；有能力提供优质教学；建立起针对学生和老师的激励体制，并开展有效的知识共享；资源配置合理；具备学习能力，能认识到未来的挑战；政策具备连贯性。

四、上海的 PISA2012 测试成绩与特点

PISA2012 的主测试领域是数学，辅助测试领域是阅读、数学和问题解决。全球有 65 个国家/地区的 51000 名学生参与了测试，代表了这些国家的 2800 万 15 岁在校学生。OECD 在上海随机抽取了 155 所中学的 6374 名学生，代表上海全市 9 万左右的 15 岁在校生。在这次测试中，上海继 2009 年之后再次夺冠，而且在各个测试领域中都遥遥领先，表现相当抢眼。

在数学素养的测试中，上海的平均成绩为 613 分，比排名第二的新加坡高出 40 分。另外，它比 OECD 国家的平均分数高出 119 分，相当于约三年的学校教育结果。在阅读素养的测试中，上海的平均成绩为 570 分，比排名第二的香港高出 25 分，比 OECD 国家的平均分数高出 74 分。在科学素养测试中，上海的平均成绩为 580 分，比排名第二的香港高出 25 分，比 OECD 国家的平均成绩高出 79 分。从总体来看，上海的测试成绩相当突出。

此次测试反映出，上海的高水平学生比例很高。PISA 测试将测试成绩达到精熟度 5 级或 6 级的学生定义为高水平学生。在数学和阅读素养测试中，上海的高水平学生比例分别达到 55% 和 25%，是所有参与测试的国家中最高的。相比之下，OECD 国家的平均水平分别为 13% 和 8%。

性别对于上海学生的数学素养总成绩没有产生显著性差异，但男生的数学表述和数学阐释成绩显著高于女生，在变化和关系、数量这两个测试分领域上也有显著优势。在数学成绩相同的学生里，男生和女生在使用信息、快速理解事物和解决问题的能力上存在显著性差

异。此外，上海学生虽然在解决常见数学问题上的自信度高于其它参与测试的国家（地区），但不少学生对数学学习感到焦虑。女生的坚持性较差，容易焦虑，对自己的数学能力评价不高。男生解决复杂问题的意愿、学习兴趣和参与方面都高于女生。

OECD 从教育公平性、学习者特征和学校环境等因素出发分析了 PISA2012 的测试成绩，从中可以观察到上海测试成绩的部分特点：

1.教育公平性。如前文所述，2012 年的测试成绩证明了，社会经济条件较好的学生成绩高于社会经济条件较差的学生，但也有 6% 的学生摆脱了这种家庭背景的约束。在上海，这种情况更加突出。在参与 PISA2012 且排名居前 25% 的学生中，有 13% 的上海学生属于此类学生。从总体来说，上海的教育机会公平性低于 OECD 的平均水平，但成绩远高于教育机会公平程度高的国家（地区），背后的原因非常值得探讨。

2.学习者特征。上海学生在学习的参与程度、动力和自信等许多方面的表现都优于其它国家。如 OECD 国家平均有 25% 的学生出现过迟到或旷课的情况，导致其平均分下降 37 分。而上海学生的迟到率或旷课率只有 4%，平均分因此下降 33 分。有 82% 左右的上海学生称自己喜欢在校的学习生活，有较强的归属感，高于 OECD 的平均水平（80%）。此外，学生的学习动机也相当强，在所有国家中排名第三。上海学生在学习者特征方面的优势或许是其成绩突出的原因之一。

3.学校环境。上海学生的数学成绩差异主要是由学生个人能力和家庭背景方面造成的（53%），学校因素造成的差异占 47%。具体说来有以下几个方面的特点：1. 上海学生接触正式教学内容的机会是所有国家（地区）中最高的，而且课业负担相对不轻。上海学生报告的校内上课时间为平均每周 28.2 小时，在 65 个国家（地区）中位于第 9 位，作业时间为平均每周 13.8 小时，列第 1 位。但作业量与成绩之间并非线性关系。上海的数据结果表明，上海 15 岁学生每周的最佳作业时间在 11 小时左右，超过 11 小时以后成绩随之提高程度很小。2. 师生关系融洽。90% 以上的学生同意或非常同意“学生和大部

分老师相处融洽”。但只有 31.3%的学生认为“老师布置需要我们花费很长时间思考的作业”，可见上海的课后作业不太重视思考。

3.OECD 国家的测试结果反映出，生均教育经费超过一定水平后，其效果就由经费数额和资源分配水平共同决定，而不一定是投入的越多，教育产出就越高。资源分配的公平性与教育公平以及学校的整体表现都息息相关。而将上海与其它参与测试的国家比较后发现，上海的教师工资水平和教育资源的分配公平性居于中游，并未表现出明显优势。

（中国发展研究基金会 刁琳琳 整理）

地 址： 北京市东城区安定门外大街 136 号皇城国际中心 A 座 15 层
电 话： (010) 64255855 邮 编： 100011
传 真： (010) 84080850 网 址： www.cdrf.org.cn
责任编辑： 任晶晶 电子邮箱： renjj@cdrf.org.cn
