# 大规模考试和学业质量评价

王蕾 著

绩效问责是教育管理的核心
学业质量评价是绩效问责的前提和基础
大规模考试可以成为学业质量评价的重要工具
发挥考试的教育评价功能是大规模考试的出路和根本目的

高等教育出版社
HIGHER EDUCATION PRESS

# 大规模考试和学业质量评价

Daguimo Kaoshi he Xueye Zhiliang Pingjia

王 蕾 著

高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

**图书在版编目(CIP)数据**

大规模考试和学业质量评价 / 王蕾著. --北京：高等教育出版社，2013.3
ISBN 978-7-04-037015-7

Ⅰ.①大… Ⅱ.①王… Ⅲ.①考试制度-教育评估-研究-中国 Ⅳ.①G424.74

中国版本图书馆 CIP 数据核字(2013)第 031223 号

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 策划编辑 | 王小钢 | 责任编辑 | 王小钢 | 封面设计 | 李卫青 | 版式设计 | 范晓红 |
| 责任校对 | 窦丽娜 | 责任印制 | 田 甜 | | | | |

| | | | |
|---|---|---|---|
| 出版发行 | 高等教育出版社 | 咨询电话 | 400-810-0598 |
| 社 址 | 北京市西城区德外大街 4 号 | 网 址 | http://www.hep.edu.cn |
| 邮政编码 | 100120 | | http://www.hep.com.cn |
| 印 刷 | 北京宏伟双华印刷有限公司 | 网上订购 | http://www.landraco.com |
| 开 本 | 787mm×1092mm 1/16 | | http://www.landraco.com.cn |
| 印 张 | 14.75 | 版 次 | 2013 年 3 月第 1 版 |
| 字 数 | 260 千字 | 印 次 | 2013 年 3 月第 1 次印刷 |
| 购书热线 | 010-58581118 | 定 价 | 35.00 元 |

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换



物 料 号 37015-00

# 前　言

管理就是决策，决策需要信息，信息来自数据，但数据本身不是信息。信息反映我们对事物固有规律的确定性。信息多，我们对规律的确定性就增加；信息少，我们对规律的确定性就减少。规律隐藏在数据背后，但数据有时含有噪声和污染，只有通过对数据加以整理和分析规律才能浮现出来。

考试作为教育评价的重要工具之一，被广泛应用于教育领域和社会领域中，其功能包括甄选、证明、诊断、反馈与矫正等。校内考试的功能一般局限在教学本身，由教师或学校根据教学和管理需要随时设计和实施，没有严重后果，形式比较自由。由社会管理机构设立、专业化考试机构举办的大规模高利害考试（如高考、公务员考试）则更加强调其社会功能，通常与个人或集体利益攸关，考试这时可以说是社会管理的工具，和教育本身的关系已经不大了。

在我国，大规模考试长期以来是教育体系和社会生活中错综复杂的“结”，聚焦了大量的教育矛盾和社会矛盾，是我国教育改革发展过程中的热点和难点之一。

学业质量评价则是近些年国际和国内教育界最为关心的焦点话题。现代教育管理的一个重要特征是强调绩效管理，绩效问责是教育现代化中科学决策的重要基础。然而，和经济领域中容易获得共识的投入产出不同，如何定义教育绩效是一个难题。大多数情况下，教育投入是容易定义的，产出究竟是什么就不那么容易定义。如果我们能够在教育的根本产品是人上取得共识，那么，受教育者所具备的知识和技能可以说是教育的根本产出。能对这一产出进行绩效管理的根本条件是他们必须是可观测的和可量化的。因此，学业质量评价即对学生知识和技能的定量的评估是教育绩效管理的核心内容，是现代化教育体系的重要组成部分，是全面提高教育质量和促进教育公平的前提和条件。

学业质量评价关心的是群体，大规模考试关心的是个体。表面上，这是两个完全不同的概念，事实上有密切的联系。他们所关注的共同目标都是学生的知识和技能，都需要使用教育测量的方法去测量和了解学生知道什么和能做什么。在传统教育观念和技术下，考试的主要功能是进行选拔和分流，当其用于评价时，就显现了技术和观念的多重缺陷，比如，简单地使用考试的原始分数对考生或群体排队存在方式单一、内容窄化、考试结果误用和滥用、对终结性评价过度

依赖等负面现象。

本书学习借鉴了国际上最有影响的学业质量评价项目的经验，建议采用现代教育测量理论对传统考试加以技术改造。改造主要在两个技术环节进行：第一在工具构造上，利用 Rasch 测量模型使考试结果成为客观等距量尺。在基于发展性评价的理论上，引入学科同一能力量表的技术和方法，改变传统上只关注学生一个时间点上的学习结果的片面评价理念，实现对学习过程的增值评价。第二在结果阐释上，打破传统的以分数为主要甚至是唯一依据的做法，利用问卷调查收集学生的社会经济文化等背景信息，利用结构方程模型、Rasch 模型等现代统计方法结合背景信息全面分析影响学业质量的因素。经过这样的技术改造，使传统的大规模教育考试可以成为学业质量评价体系的有机组成部分，这不只有效地利用已有资源成为学业质量评价的重要视角，而且改变了长期以来考试和教学之间互相牵制的矛盾现象，使考试为教学服务，成为教学过程中不可或缺的重要环节；使考试数据为决策服务，成为促进教育现代化，促进社会和谐公平发展的有效工具。

这种改造并非一蹴而就的，存在大量技术上的挑战。为此，决策者要尊重大规模教育考试和学业质量评价的高度专业性，加强专业化考试和评价机构的建设。考试和评价专业人员则要主动挖掘和利用大规模教育考试的海量数据开展学业质量评价，提供公共信息服务。考试评价项目要管办分离、数据公开，要利用专业力量开展再评价，确保考试评价的科学性和有效性。

本书各章的主要内容如下：

第 1 章介绍研究背景和问题，阐述研究目的和意义，梳理国内外相关研究成果，讨论研究的重点、方法、创新和不足。

第 2 章介绍相关概念和理论依据。以公共管理中公共部门绩效评估理论、教育评价中 Rasch 测量理论和发展性增值评价理论为指导，讨论基于大规模教育考试构建学业质量评价体系，实现基于同一能力量表下的增值评价的可行性、必要性和迫切性。

第 3 章精选美国、英国、荷兰、日本四国的考试评价体系，对基于考试的学业质量评价模式及实践进行比较分析。

第 4 章介绍与分析学生能力国际评价（PISA）项目分别于 2006 年和 2009 年在中国进行的两轮试测研究实践，为考试评价体系支持教育决策提供案例验证。

第 5 章具体讨论了对大规模考试进行两个环节的技术改造的理论、方法及其在我国高考中的实践探索。分别以某省高考为背景，运用 Rasch 理论构建同一能力量表从中考链接到高考的高中学业质量评价和通过 PISA 中国试测研究的问卷调查，分析阐述了依靠现代教育测量技术解决长期以来困扰学业质量评价问题的技术创新途径，指出了通过专业机构再评价提高考试和学业质量评价水平的体制和机制创新的建议。

第6章完整设计了利用现有考试开展学业质量评价的考试评价体系，论述了构建考试评价体系的各种选择和实施程序，力图从制度设计层面推动基于现有大规模教育考试开展学业质量评价的有效实施。

第7章展望整合现有大规模教育考试构建考试评价体系的前景和努力方向。

本书作者基于自己长期从事教育考试管理和研究工作的实际经验，以及组织参加国际大型教育评价项目的亲身体验，聚焦大规模考试和学业质量评价这两个热点，探讨如何采用国际最先进的教育测量和评价技术对传统的大规模教育考试进行技术改造，从技术上使考试更加科学、准确、公正，避免副作用；同时充分利用资源，使之不只用于选拔分流，还能对评价学业质量、改进教育管理、促进教学改革发挥应有的作用。本书提出的“改造利用大规模教育考试开展学业质量评价”这一议题在理念上顺应了国际上教育绩效管理的发展趋势，在技术细节上讨论了国际最领先的客观测量和数据挖掘的理论工具和模型，仔细分析了国内若干个大规模教育考试实际操作的例证，在实施上呼吁加强数据公开和共享，推进考试专业机构的再评价服务，为我国教育和考试改革提供了新的思路和视角。

# 序

技能比以往任何时候都更能改变生活和促进经济，它已经成为21世纪经济生活中的货币。道理很简单：如果没有合适的技能，人们将会被社会边缘化，技术进步无法转化为经济增长，国家也无法在今天的经济中竞争。

然而，更多的教育并不会自动转化为更好的经济和社会产出。现在有一种病态的社会状况：一方面很多大学生毕业即失业，一方面很多雇主又说找不到具有合适技能的员工。要想成功地将更好的技能转化为更好的工作和生活，关键是要理解促使成功和产出的技能是什么。

在上一代人的时候，老师可以期望他们所教授的内容将会伴随学生的一生。因此，教授具体的知识内容是教育的核心。今天，学校需要帮助学生适应比以往任何时候都更迅速的社会变革，以胜任目前还未出现的工作，使用目前还未出现的技术，解决未知的问题。教育在我们的社会中促进可持续的价值观方面也具有关键作用。

今天，我们可以访问互联网上的内容，常规的认知技巧被数字化或外包，工作岗位不断地迅速变化。这种背景下，知识的积累变得不那么重要，成功越来越多地受思维方式影响，包括创造性、批判性思维，问题解决和判断；也受工作方式影响，包括协作和团队合作；还受能使我们与世界互动的社会文化工具影响。而所有这一切都是基于良好的公民意识、人格、社会责任感以及对建立在不同文化、价值观和信仰上的个人能力的崇尚。

毫无疑问，一个学科内最先进的技能将永远是重要的。创新型人才通常在某一知识领域或实践中具有特殊技能。“学会学习”的能力也是很重要的，我们总是从学习中学习。然而，教育的成功主要不再是复述学科知识，而是在新的情境中去推广和运用我们所得到的知识。

例如，我们现在面临的挑战是要确保数学对学生而言不只是充斥着大量的公式和定理的世界，而是能让他们描述、构建、理解和预测这种世界的一种语言。传统的数学教学往往是在一个抽象的数学世界中进行的，与真实的环境大不相同——例如，学生学习算术技巧时，会被要求完成大量的算术运算；或者老师演示如何解答特定类型的方程后，会给学生很多类似的方程来解答。但是，学生是否能够理解数学的基本概念，是否可以将他们面对的新情况或问题转变与为数

学相关的表达形式？他们是否可以对这些问题进行数学处理和识别，并使用相关的数学知识来解决？他们是否能根据原始问题的脉络来评价解决方案？这些才是核心的问题。

同样，现在我们可以搜索和访问越来越多的知识内容，那么对这些内容作出合理解释的能力，对大家公认的知识提出质疑或寻求改进的能力，以及合理使用时间的能力变得越来越重要。过去，我们可以告诉学生在他们需要信息时去查看百科全书，我们可以告诉他们可以信赖他们找到的内容是正确的。过去，“有文化”主要指能阅读，用来处理一套完整的编码知识体系的技术技能每个人往往只需要学习一次就可以运用终身。今天，“有文化”意味着会学习，更多地是指在不断变化的环境中与各种各样的情况下能够识别、理解、阐释、创造以及交换知识的能力和动力，包括处理歧义、解释和解决不同信息间的冲突。

通常，我们教学时处理问题的方法是将问题分解成易于操作的片段，然后教学生解决这些片段问题的技巧。但今天价值往往创造自将各离散的片段整合起来，这与好奇心、思想开放、将看似无关的想法联系起来等有关，这就要求个人熟悉并易于接受其他领域的知识。如果我们把一生都用在一个单一的学科，我们就无法获得富有想象力的技能来连接可能孕育下一个发明的片段。

世界也已不再分为专家和通才。这里所说的专家一般都在较狭窄的专业范围内有很深奥的技能，他们的专业知识深受同行认可但不被其他领域重视。而通才一般具备较肤浅的技能但涉及更广阔的专业范围。今天社会更需要的是全才，他们必须具有多种能力，能够将技能逐步深入地运用到更大的范围和情况中去，从而获得新的能力，建立关系，并承担新的角色。他们不仅能够不断适应新环境，还能不断地学习和成长，在一个快速变化的世界中不断重新定位自己。

最后，值得一提的是过去学生往往在学校里独自学习，在学年结束的时候，学校要证明他们的个人成绩。但在世界变得越来越相互依存的今天，我们更需要伟大的合作者和组织者。今天的创新很少是独立工作的个人的产品，而更多是我们如何调动、分享和联系知识的结果。当今世界，今天我们专有的一切知识明天将成为可供他人使用的商品。换句话说，我们正目睹着世界从一个大货仓——装满着大量堆叠起来的迅速贬值的知识——转变成为一个沟通日益丰富、协作不断增加的世界。

PISA 于 1997 年推出，开始挑战测量学生学习成果的固有模式，基于一个国际公认的共同框架定期对学生成就进行考量。现在有 74 个国家正在参加 PISA，约占世界经济总量的 87%。PISA 旨在为政策对话和制定、实施教育目标的合作提供一个基础，以创新的方式反映与成人生活有关的技能的判断。

由于有时间限制的评价（例如 PISA）在一个时间点只能选择捕捉一种能力，而且由于目前各种方法的限制，使能力的本质并不能完全在大规模评价项目中揭示出来，PISA 正努力在教育系统的比较层面上逐步拓宽能力的评价范围。

• PISA2000 开始重点测评阅读素养，测查了学生使用、解释和反思书面材料的能力。

• PISA2003 的重点测评领域是数学素养，侧重于学生以多样的、反思的、基于洞察力的方式把数学知识实际运用在多种不同情况中的能力。

• PISA2006 对学生的关注集中在：(1) 科学知识和使用这些知识来识别问题、获取新知识、解释科学现象，并基于证据得出与科学相关问题的结论；(2) 理解科学作为一种人类知识和探索的特征；(3) 明白科学技术如何塑造我们的物质、知识和文化环境；(4) 以及他们作为一个公民，以科学思想参与与科学相关问题的愿望。PISA 也开始更普遍地关注学生的学习特点、学习方法、自我意识和他们对学校事件的参与度。

• PISA2009 重点测评的领域回到阅读素养，但自 2000 年起，对阅读能力的需求已经从根本上改变了。在许多国家，阅读不再仅仅是应对传统的印刷形式的文本，而是越来越需要个人建构新的知识。在这个建构的过程中，他们可能只在网页上看到一小部分信息，然后利用这些信息建立复杂的数字信息结构的心理表征，灵活地处理相互冲突的信息来源。为抓住这些新的技能维度，PISA 引入了一种新的以电子形式交付的数码素养的评价。

• PISA2012 的重点测评领域回到数学素养，但是新增加了一个灵活解决问题技巧的评价。

• PISA2015 重点测评的领域再次返回到科学素养，但也会引入一种新的具有革新意义的协作解决问题技巧的评价，以便开始进行人际交往技能维度的评价。

那么，我们如何能得知 PISA 测评的技能就是学生未来成功所需的素养呢？加拿大青年成长调查(YITS)——一个纵向调查评价青年人生活模式及对其影响重大的教育、培训和工作转换的调查项目从一方面帮助回答了这一问题。2000 年，加拿大有 29 330 名 15 岁的学生既参与 YITS 调查又参加 PISA 测试，4 年后，考查这些已经 19 岁的学生教育成果(反应率为 90.5%)和他们 15 岁时 PISA 阅读素养成绩的相关性。研究结果显示，那些在 15 岁时 PISA 测试阅读能力达到 1 级的学生在 19 岁时选择接受高等教育的人数比例为同龄人的 2 倍，即使考虑了学校参与、性别、母语、居住地、父母、教育和家庭收入方面的差别后也是如此；PISA 测试阅读能力为 4 级的学生将可能性增加到 8 倍；而 PISA 测试阅读能力为 5 级的学生则将可能性增加到 16 倍。研究结果还表明，19 岁前读完高中的学生比例随着 PISA 测试的阅读能力水平的增加而增加。整体来说 87% 的学生在 19 岁时读完高中，而 PISA 测试阅读能力为 1 级及以下的学生比例仅为 62%；测试阅读能力为 2 级的比例则为 77%。该结果考虑了性别、母语和教学语言、父母教育、家庭收入、学术和社会背景等因素的影响。纵观 PISA 的 5 级阅读能力水平的研究结果，显现出了门槛效应：PISA 阅读能力为 1 级及以下和 2

级的学生完成高中教育的可能性显著低于阅读能力为3级及以上的学生。高中教育是争取进一步学习和培训机会的基础，也是从学习到全职工作成功转型的基础，这些数据给出建议：15岁时阅读能力为2级及以下的学生有潜在面临困难的危险。值得注意的是，研究分析还表明PISA测试能很好地预测反映传统学校标志的未来成功。澳大利亚、丹麦、乌拉圭也进行了类似的研究，并得到了非常相似的结果。

PISA体现了参与国家的政府对测量相关教育成果的努力，提供了国际合作的基础。来自各参与国的专家也在分析研究结果和发现方面紧密合作。他们的工作已经证明了国家间及国家内教育成就的巨大差异。通过PISA证明可以实现卓越和公平的学习成果的教育体系向其他国家表明了什么样的目标是可以到达的。

当然，人均国内生产总值(GDP)影响着教育的成功，但是它只能解释6%的学生的平均表现差异，其余的94%则反映了公共政策影响差异的潜力，PISA2009的评价结果证明了这个结论。

中国上海学生以明显优势独占鳌头的惊人成功证明了在适度的经济资源和多样化的社会环境中所能取得的成功。在数学素养方面，超过四分之一的上海15岁的学生可以根据他们自己的调查和对复杂问题的建构来形成概念，概括并创造性地使用信息。他们可以在新的环境中，深入理解、发现和应用新方法，而在OECD成员国中，仅有3%的学生可以达到这个水平。同时，PISA建议上海也应从世界其他国家中吸取成功经验，因为上海学生的信息理解策略排名仅位于11位，信息整合策略位于15位，而且学生的自主学习能力也低于工业化国家。

更好的教育产出是对经济增长的有力预测，而财富和教育支出不足以保障更好的教育产出。总体而言，PISA显示出，一个将有良好教育的富裕国家与不良教育的贫穷国家划分的世界图景已经过时了。这个发现既是一种警示，也是一种机会。它向发达的经济体警示，他们不能想当然地以为将永远占据优于世界其他部分的“人力资本”，在全球竞争日益激烈的时代，他们需要更加努力地掌握基础知识和技能，跟上不断变化的时代需求。

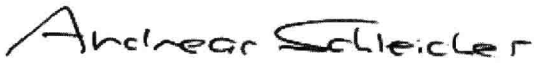
尤其值得注意的是，PISA提醒许多发达国家要应对教育表现不佳的现象，以保证许多未来的劳动力可以尽可能地掌握参与社会和经济发展的基本能力。同时，研究结果表明，即使是在收入较低的国家，贫乏的技能素养也是可以避免的——但重要的是这些国家需要做到事半功倍。

最后，同样重要的是，起点不同的许多国家都表现出了教育质量持续提高的潜力。韩国学生在2000年中PISA的平均成绩已经很高，但韩国的政策制定者仍在担忧只有极少数的精英学生取得了卓越成绩，在不到10年的时间里，韩国卓越学生的比例在阅读素养上已经翻番。波兰学校系统的重大改革大大降低了学校间学生成绩的差异，降低了低等级水平学生的比例，学生成绩总体上提高了

相当于半个学年的水平。2000 年 PISA 的测评结果震惊了德国:低于平均水平的成绩和大范围的教育失衡,被迫采取的改进措施已经在提高成绩和促进公平方面取得了进展。以色列、意大利和葡萄牙已经更接近 OECD 的平均水平;与当初的低成绩相比,巴西、智利、墨西哥和土耳其等国也取得了令人印象深刻的进步。但 PISA 最大的价值在于激发各国努力帮助学生学得更好、教师教得更好、学校系统表现得更加有效。

随着 PISA 被广泛认同,来自世界各地的专家们也在密切合作来改进这项工作,以克服现有测评工具的重大限制。现代评价工具的要求很高,不仅需要公平、技术可行和有效,而且需要建立一个更通用的方法来确保充分的预设建构和不同措施的区分性,从而为不同层次的教育系统的决策服务。他们需要提供对具体的教育层次更详尽、更有效的反馈,为教育系统在各个层次上的改进和问责制提供着力点。教师要求能更好地明晰评价结果所代表的学生思维,学校管理者、政策制定者和教师要求能够更好地利用评价信息来确定如何为学生学习创建更好的机会。未来还将要求评价要认识到成功学习的过程和事实、数据的重要性,并要求评价可以作为一个窗口去认识学生在动态任务环境中运用理解力和概念策略来解决问题的能力。

本书的成就令人印象深刻,书中阐述了通过独特的工具测评中国试点地区的学生素养,研究结果提供给中国的政策制定者和相关人员,以期改进学校教育系统,并让年轻人更好地适应即将踏入的变化纷呈的成人生活和日益加深的全球化生存环境。



Andreas Schleicher

OECD 秘书长、教育政策特别顾问、教育司副司长

# Preface

More than ever, skills transform lives and drive economies, skills have become the currency of 21st century economies. The bottom line is simple: Without the right skills, people are kept on the margins of society, technological progress doesn't translate into economic growth, and countries can't compete in today's economies.

But more education does not automatically translate into better economic and social outcomes. There is this toxic co-existence of unemployed graduates on the streets, while many employers say they cannot find the people with the skills they need. To succeed with converting better skills into better jobs and lives it is essential to understand what those skills are that drive success and outcomes.

A generation ago, teachers could expect that what they taught would last for the lifetime of their students. So teaching specific educational content was at the centre of education. Today, schools need to prepare students for more rapid social change than ever before, for jobs that have not yet been created, to use technologies that have not yet been invented, and to solve problems that we don't yet know will arise. Education also has a key role to play to foster sustainable values in our societies.

Today, where we can access content on the Internet, where routine cognitive skills are being digitized or outsourced, and where jobs are changing rapidly, accumulating knowledge matters less, and success becomes increasingly also about ways of thinking, including creativity, critical thinking, problem-solving and judgment; about ways of working, including collaboration and teamwork; and about the socio-cultural tools that enable us to interact with the world. And all of that is centered around good citizenship and personal and social responsibility and capacity of individuals to appreciate and build on different values, beliefs and cultures.

There is no question that state-of-the-art skills in a discipline will always remain important. Innovative or creative people generally have specialized skills in a field of

knowledge or a practice. And as much as "learning to learn" skills are important, we always learn by learning something. However, educational success is no longer mainly about reproducing content knowledge, but about extrapolating from what we know and applying that knowledge in novel situations.

For example, the challenge is to ensure that mathematics doesn't just remain a world of equations and theorems for students, but a language that allows them to describe, structure, understand and predict the world. Traditionally mathematics is often taught in an abstract mathematical world, in ways that are removed from authentic contexts—for example, students are taught the techniques of arithmetic, then given lots of arithmetic computations to complete; or they are shown how to solve particular types of equations, then given lots of similar equations to solve. But do students have an understanding of the fundamental concepts of mathematics; can they translate a new situation or problem they face into a form that exposes the relevance of mathematics? Can they make these problems amenable to mathematical treatment, identifying and using the relevant mathematical knowledge to solve the problem, and can they evaluate the solution in the original problem context? These are central questions.

Similarly, the more content knowledge we can search and access, the more important becomes the capacity to make sense out of this content, the capacity of individuals to question or seek to improve the accepted knowledge and practices of their time. In the past, we could tell students to look into an encyclopedia when they needed some information, and we could tell them that they could generally rely on what they found to be true. In the past, literacy was mainly about *learning to read*, a set of technical skills that individuals would acquire once for a lifetime in order to process an established body of coded knowledge. Today, literacy is about *reading for learning*, the capacity and motivation to identify, understand, interpret and create and communicate knowledge, using written materials associated with varying situations in continuously changing contexts. It also involves dealing with ambiguity, interpreting and resolving conflicting pieces of information.

Conventionally our approach to problems in schooling was breaking them down into manageable bits and pieces, and then to teach students the techniques to solve them. But today we create value by synthesizing the disparate bits. This is about curiosity, open-mindedness, making connections between ideas that previously seemed unrelated, which requires being familiar with and receptive to knowledge in other fields than

our own. If we spend our whole life in a silo of a single discipline, we will not gain the imaginative skills to connect the dots where the next invention will come from.

The world is also no longer divided into specialists and generalist. Specialists generally have deep skills and narrow scope, giving them expertise that is recognized by peers but not valued outside their domain. Generalists have broad scope but shallow skills. But what counts more and more are the versatilities that are able to apply depth of skill to a progressively widening scope of situations and experiences, gaining new competencies, building relationships, and assuming new roles. They are capable not only of constantly adapting but also of constantly learning and growing, of positioning themselves and repositioning themselves in a fast changing world.

Last but not least, in schools, students typically learn individually and at the end of the school year, we certify their individual achievements. But the more interdependent the world becomes, the more we need great collaborators and orchestrators. Innovation today is rarely the product of individuals working in isolation but an outcome of how we mobilize share and link knowledge. In today's world, everything that is our own proprietary knowledge today will be a commodity available to everyone else tomorrow. Expressed differently, we are seeing a shift from a world of stocks—with knowledge that is stacked up somewhere depreciating rapidly in value—to a world in which the enriching power of communication and collaborative flows is increasing.

The Programme for International Student Assessment (PISA), launched in 1997, took up the challenge of measuring learning outcomes against these paradigms, on a regular basis and within an internationally negotiated framework. The 74 countries currently participating in PISA make up 87% of the world economy. PISA seeks to provide a basis for policy dialogue and for collaboration in defining and implementing educational goals, in innovative ways that reflect judgments about the skills that are relevant to adult life.

Since a time-bound assessment such as PISA can only capture a selection of competencies at any single point in time, and since various methodological constraints limit the nature of competencies that are currently amenable to large scale assessment, PISA seeks to progressively work towards widening the range of competencies on which education systems are compared.

- In 2000, the PISA began with a focus on reading literacy, where it examined the

capacity of students to use, interpret and reflect on written material.

- In 2003, PISA focussed on the capacity of students to put mathematical knowledge into functional use in a multitude of different situations in varied, reflective and insight-based ways.

- In 2006 PISA focused on students: i) scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to draw evidence-based conclusions about science-related issues; ii) understanding of the characteristic features of science as a form of human knowledge and enquiry; iii) awareness of how science and technology shape our material, intellectual and cultural environments; and their iv) willingness to engage with science-related issues, and with the ideas of science, as a reflective citizen. PISA also began to look at students' dispositions to learning, their approaches to learning, their self-concept and their engagement with school more generally.

- By 2009 PISA returned to reading literacy, but since 2000, the demands on reading skills had fundamentally changed. In many countries, reading was no longer only about managing classical texts in printed form, but increasingly required individuals to construct new knowledge, to build mental representations of complex digital information structures of which they may only see a small part on the web, and to deal with conflicting sources of information dynamically. To capture these new dimensions, PISA introduced a new assessment of digital literacy that was delivered in electronic form.

- In 2012, PISA returned to mathematics, but this time includes a new assessment of dynamic problem-solving skills.

- In 2015, PISA will return again to science, but also introduce a new and innovative assessment of collaborative problem-solving skills in order to begin assessing interpersonal skill dimensions.

How can we know whether the skills measured by PISA are indeed those that matter for the future success of students? The Canadian Youth in Transition Survey (YITS), a longitudinal survey which investigates patterns of and influences on major educational, training and work transitions in young people's lives, has provided one way to

assess this. In 2000, 29330 fifteen-year-old students in Canada participated both in YITS and PISA. Four years later, the educational outcomes of the same students, then aged 19, were assessed (response rate: 90.5 %) and the association of these outcomes with PISA reading performance at age 15 was investigated. The results show that students who had mastered PISA performance Level 1 in reading at age 15 where twice as likely to participate in postsecondary education at age 19, even after accounting for school engagement, gender, mother tongue, place of residence, parental, education and family income. The odds increased to eightfold for those students who had mastered PISA Level 4 and to sixteen-fold for those who had mastered PISA Level 5. The results also show that the percentage of youth who had completed high school by age 19 increased each increase in a PISA reading proficiency level. While 87% of the students overall had completed high school at 19, the percentage was significantly lower at proficiency level 1 or below (62%), and proficiency level 2 (77%). This pattern of results held up after taking into account the effects of gender, mother tongue and language of instruction, parental education, family income, and academic and social engagement. In looking at the results for each of the five PISA reading proficiency levels in turn, a threshold effect became apparent: students whose performance was at PISA Level 1 or below, and at Level 2, had significantly lower odds of completing high school relative to students with reading abilities at Level 3 or above. As high school education serves as a foundation for further learning and training opportunities, and for a successful transition from schooling to full-time employment, these data suggest that students with a reading performance at Level 2 or below at age 15 are potentially at risk of facing difficulties. It is noteworthy that the analysis also showed that the performance on the PISA test was a far better predictor for future success that reflected in traditional school marks. Similar studies undertaken in Australia, Denmark and Uruguay led to very similar results.

PISA represents a commitment by the governments of the participating countries to measure relevant outcomes of education and provides a basis for international collaboration. Experts from the participating countries also collaborated closely to analyze the findings. Their work has revealed wide differences in educational outcomes, both within and across countries. The education systems that have been able to demonstrate strong and equitable learning outcomes on PISA, and to mobilise rapid improvements, show others what is possible to achieve. Naturally, GDP per capita influences educational success, but this only explains 6% of the differences in average student performance on the PISA 2009 assessment. The other 94% reflect the poten-

tial for public policy to make a difference. The stunning success of Shanghai in China, which tops every league table in this assessment by a clear margin, shows what can be achieved with moderate economic resources and in a diverse social context. In mathematics, more than a quarter of Shanghai-China's 15-year-olds can conceptualize, generalize, and creatively use information based on their own investigations and modeling of complex problem situations. They can apply insight and understanding and develop new approaches and strategies when addressing novel situations. In the OECD area, just 3% of students reach that level of performance. At the same time, PISA suggests that also Shanghai can learn important lessons from the world. It ranks only 11th when it comes to students' strategies to understand information, 15th when it comes to strategies for synthesizing information and below the industrialized world when it comes to strategies for self-directed learning.

While better educational outcomes are a strong predictor of economic growth, wealth and spending on education alone are no guarantee for better educational outcomes. Overall, PISA shows that an image of a world divided neatly into rich and well-educated countries and poor and badly-educated countries is out of date. This finding represents both a warning and an opportunity. It is a warning to advanced economies that they cannot take for granted that they will forever have "human capital" superior to that in other parts of the world. At a time of intensified global competition, they will need to work hard to maintain a knowledge and skill base that keeps up with changing demands. PISA underlines, in particular, the need for many advanced countries to tackle educational underperformance so that as many members of their future workforces as possible are equipped with at least the baseline competencies that enable them to participate in social and economic development. At the same time, the findings show that poor skills are not an inevitable consequence of low national income—an important outcome for countries that need to achieve more with less.

Last but not least, countries from a variety of starting points have shown the potential to raise the quality of educational outcomes substantially. Korea's average performance was already high in 2000, but Korean policy makers were concerned that only narrow elite achieved levels of excellence in PISA. Within less than a decade, Korea was able to double the share of students demonstrating excellence in reading literacy. A major overhaul of Poland's school system helped to dramatically reduce performance variability among schools, reduce the share of poorly performing students and raise overall performance by the equivalent of more than half a school year. Germany