教育评价探新

柳斌题签

王蔷

Exploration in Educational Assessment and Evaluation

西安交通大学出版社

教育评价探新

柳斌题签

王蕾

Exploration in Educational Assessment and Evaluation

西安交通大学出版社

责任编辑：陈 丽　　等

图书在版编目（CIP）数据

教育评价探新/王蕾. -西安：西安交通大学出版社，2007.11

ISBN 978-7-5605-2610-2

Ⅰ.教… Ⅱ.王… Ⅲ.教育评估—研究Ⅳ.G449.7

中国版本图书馆CIP数据核字（2007）第176365号

出版发行	西安交通大学出版社		
社　　址	西安市兴庆南路10号	邮政编码	710049
电　　话	(029)82668316　82668315	传　　真	(029)82668280
经　　销	全国各地新华书店		
印　　刷	北京联兴盛业印刷有限公司	版　　次	2007年11月第1版
开　　本	787mm×1 092mm　1/16	印　　次	2007年11月第1次印刷
印　　张	24.5	书　　号	ISBN 978-7-5605-2610-2/G・271
字　　数	400 000	定　　价	58.00元

本书如遇缺页、倒页、脱页等质量问题，请与销售单位联系调换。

Preface

The world is rapidly becoming a different place. Our increasingly diverse and interconnected populations, rapid technological change in the workplace and in everyday life, and the instantaneous availability of vast amounts of information represent but a few of these new demands. Other demands relate to the type of world modern societies want to promote: balancing economic growth with the sustainability of natural environments, individual prosperity with social cohesion and reducing societal inequalities. The development of the knowledge, skills, and competencies of populations - through education systems and learning opportunities in the workplace and other venues through the life span - is key to meeting these demands. This necessarily sparks related questions about what are the competencies that are most important for today's and tomorrow's world, how they can be developed and fostered, and how they can be measured in order to provide useful guidance for students to learn better, teachers to teach better and schools to better meet the needs of societies.

The OECD Programme for International Student Assessment (PISA), launched by governments of the Organisation for Economic Co-operation and Development (OECD) in 1997, aims at assessing some of the key competencies that contribute to the success of individuals and societies, on a regular basis and within an international framework. PISA seeks to provide a basis for policy dialogue and for collaboration in defining and implementing educational goals, in innovative ways that reflect judgements about the skills that are relevant to adult life. The results from PISA show that across the globe - whether it is Canada in North America, Finland in Europe or Japan and Korea in Asia - some education systems demonstrate that excellence in education is an attainable goal, and often at reasonable cost. They also show that the challenge of achieving a high and socially equitable distribution of learning outcomes can be successfully addressed and that excellence can be achieved consistently throughout the education systems, with very few students and schools left behind.

This first-of-its kind trial marks the first time that the performance of Chinese students in selected cities is measured against international standards and highlights their relative strengths and weaknesses in the key competencies that were measured by PISA. The trial thus provides policy-makers and educators in China with a unique tool to bring about improvements in schooling and better preparation for young people as they enter an adult life of rapid change and deepening global interdependence.

What PISA does

PISA has become the most advanced and comprehensive international assessment to date, capturing roughly nine-tenth of the world economy. PISA defines competence as the ability to successfully meet complex demands in varied contexts through the mobilisation of psychosocial resources, including knowledge and skills, motivation, attitudes, emotions, and other social and behavioural components. PISA thus goes far beyond the mere reproduction of

subject matter knowledge. Measuring and comparing competencies across languages and cultures is a difficult challenge and is being pursued by PISA progressively. PISA focused its first three assessments on literacy skills, defined as the capacity of young adults to access, manage, integrate and evaluate information, to think imaginatively, to hypothesise and discover, and to communicate their thoughts and ideas effectively. The concept of literacy used in PISA is much broader than the historical notion of the ability to read and write. Furthermore, it is measured on a continuum, not as something that an individual either has or does not have. It may be necessary or desirable for some purposes to define a point on a literacy continuum below which levels of competence are considered inadequate, but the underlying variability is important. The acquisition of literacy is a lifelong process - taking place not just at school or through formal learning, but also through interactions with family, peers, colleagues and wider communities. Fifteen-year-olds cannot be expected to have learned everything they will need to know as adults, but they should have a solid foundation of knowledge in areas such as reading, mathematics and science. In order to continue learning in these subject areas and to apply their learning to the real world, they also need to understand fundamental processes and principles and to use these flexibly in different situations. It is for this reason that PISA measures the ability to complete tasks relating to real life, depending on a broad understanding of key concepts, rather than limiting the assessment to the understanding of subject-specific knowledge.

The reasoning behind shifting the emphasis from assessing whether students can reproduce what they have learned towards whether they can extrapolate from what they have learned and apply their competencies in novel situations, derives from the nature of knowledge and skills required in modern life. For example, the tasks that can solved through simple memorisation or with pre-set algorithms are those that are also easiest to digitise, automatise and offshore, and will thus be less relevant in a modern knowledge society.

- 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. Contrary to traditional school mathematics, that is often taught in an abstract mathematical world, in ways that are removed from authentic contexts, PISA tried to give the usefulness of mathematics in the real world attention and to succeed in PISA, students had to be able to draw connections between the real world and the mathematical world, often in complex open-ended tasks. Many of the PISA tasks therefore confronted students with real-life problems in open-ended format. As a first step, students had to translate the situation or problem they faced into a form that exposed the relevance of mathematics. They then had to make the

problems amenable to mathematical treatment, using the relevant mathematical knowledge to solve problems, and finally evaluate the solution in the original problem context (for examples of PISA tasks see OECD, 2006).

- PISA continued along this line of reasoning with an assessment of science in 2006 that focussed 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

Box 1: The assessment areas in PISA 2006

	Science	Reading	Mathematics
Definition and its distinctive features	The extent to which an individual: -Possesses scientific knowledge and uses that knowledge to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues. -Understands the characteristic features of science as a form of human knowledge and enquiry. -Shows awareness of how science and technology shape our material, intellectual and cultural environments. -Engages in science-related issues and with the ideas of science, as a reflective citizen. Scientific literacy requires an understanding of scientific concepts, as well as the ability to apply a scientific perspective and to think scientifically about evidence.	The capacity of an individual to understand, use and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society. In addition to decoding and literal comprehension, reading literacy also involves interpretation and reflection, and the ability to use reading to fulfil one's goals in life. The focus of PISA is on reading to learn rather than learning to read, and hence students are not assessed on the most basic reading skills.	The capacity of an individual to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. Mathematical literacy is related to wider, functional use of mathematics; engagement includes the ability to recognise and formulate mathematical problems in various situations.
Knowledge domain	Knowledge of science, such as: * "Physical systems" * "Living systems" * "Earth and space systems" * "Technology systems" Knowledge about science, such as: * "Scientific enquiry" * "Scientific explanations"	The form of reading materials: * Continuous texts: including different kinds of prose such as narration, exposition, argumentation * Non-continuous texts: including graphs, forms and lists	Clusters of relevant mathematical areas and concepts: * Quantity * Space and shape * Change and relationships * Uncertainty
Competencies involved	Type of scientific task or process * Identifying scientific issues * Explaining scientific phenomena * Using scientific evidence	Type of reading task or process: * Retrieving information * Interpreting texts * Reflection and evaluation of texts	Competency clusters: * Reproduction (simple mathematical operations) * Connections (bringing together ideas to solve straightforward problems) * Reflection (wider mathematical thinking)
Context and situation	The area of application of science, focusing on uses in relation to personal, social and global settings such as: * "Health" * "Natural resources" * "Environment" * "Hazard" * "Frontiers of science and technology"	The use for which the text is constructed: * Private (e.g. a personal letter) * Public (e.g. an official document) * Occupational (e.g. a report) * Educational (e.g. school-related reading)	The area of application of mathematics, focusing on uses in relation to personal, social and global settings such as: * Personal * Educational and occupational * Public * Scientific

of science, as a reflective citizen. PISA has also begun to look at students' dispositions to learning, their approaches to learning, their self-concept and their engagement with school more generally.

Why PISA competencies matter for individuals

The question remains whether those competencies that PISA assesses are predictive for the future success of students. This can be examined by analysing how the competencies assessed by PISA relate to subsequent educational, social and economic outcomes.

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, provide one way to assess this. In 2000, 29 330 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 shows that the performance on the PISA test was a better predictor for subsequent student success than the school marks which students had received.

A similar study undertaken in Denmark led to similar results, in that the percentage of youth who had completed post-compulsory, general or vocational upper-secondary education by 19 increased significantly with their reading ability assessed at age 15 (see http://www.sfi.dk/sw19649.asp).

How PISA can contribute to educational policy and practice

PISA was initiated to provide schools, local communities and countries with an opportunity to identify their own strengths and weaknesses, by analysing them in

the light of the performance of other systems. The idea was that comparisons with others would allow governments and practitioners to contemplate how they might further optimise their existing education system, and perhaps beyond that, also reflect on the transformation of some of the paradigm and beliefs underlying it.

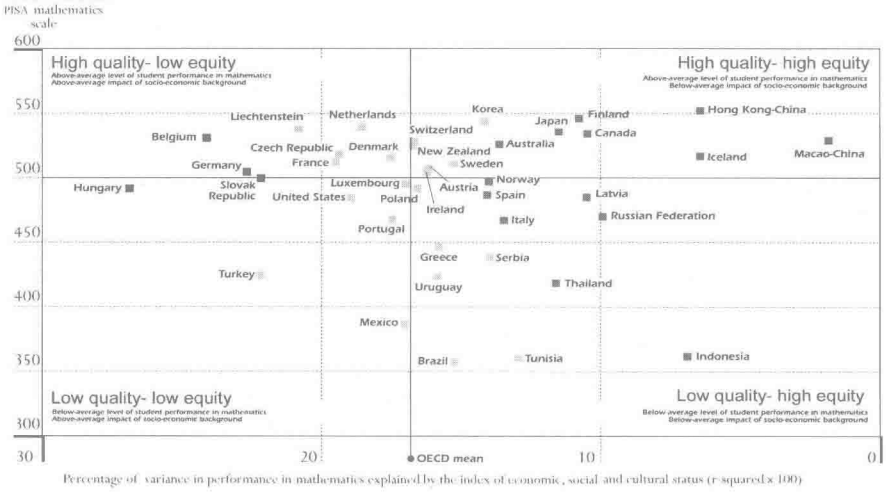
Much of the initial debate on PISA results was fuelled by a comparison of the average performance of education systems. The need for an international perspective in such comparisons seems apparent: If the percentage of students achieving high grades in national examinations increases, some will claim that the education system has improved. Others will claim that the requirements for attaining the grade must have been lowered. Behind the suspicion that better results reflect lowered criteria there is often a belief that overall performance in education cannot be raised. However, PISA has shown that some countries do much better on those competencies that were measured than others and provides compelling evidence that excellence in education and, indeed, improvement is possible.

Assessing equity in the distribution of learning opportunities

PISA has been able to shed light on the question to what extent education systems provide an equitable distribution of learning opportunities, by measuring the extent to which learning outcomes of students and school depend on their social background. Some of the most widely debated and instructive findings of PISA are that countries such as Canada, Finland or Japan show that poor performance in school does not automatically follow from a disadvantaged socio-economic background of students and that education systems can succeed in combining high performance levels with a socially equitable distribution of learning opportunities (OECD, 2004). In contrast, other countries that for decades had assumed that learning opportunities were equitably distributed, simply because schooling inputs were equally distributed, were stunned by the large inequalities that PISA revealed (See Figure).

Quality and equity in educational performance

Average performance of countries on the PISA mathematics scale and the relationship between performance and the index of economic, social and cultural status



PISA has provided policy-makers and practitioners also with important insights on how to address such inequalities. It has done so by separating equity-related issues between those that relate to the socio-economic heterogeneity within schools and those that relate to socio-economic segregation through the school system. This allows examination of the extent to which education systems moderate or reinforce socio-economic background factors. In some of the countries in which a considerable proportion of the total variance is between schools, this appears to be a consequence of education policy: In Germany, for example, students are sorted into schools of different types from age 10 on the basis of achievement at that stage and a judgement of whether a more academic or vocational school would be most appropriate for the next step. The deliberate intention of the policy is to reduce variation within schools, by bringing relatively similar students together, and to increase variation between schools that will then be reflected in differences between the schools in curricula. In many countries the consequence is that students from more privileged social backgrounds are directed into the more prestigious academic schools which yield superior educational outcomes (as indicated by their higher performance on the PISA measures) and students from less privileged social backgrounds are directed into less prestigious vocational schools which yield poorer educational outcomes (as indicated by lower performance on the PISA measures). PISA has therefore suggested that the school organisation, therefore, both reflects and reproduces existing social divisions.

Assessing coherence in school performance

PISA also partitioned variation in student performance into within and between school components. Such comparisons can help to assess to what extent the quality of learning outcomes varies across schools and thus is a predictable and consistent outcome of the education system. These comparisons show that there are major differences among countries in the extent to which student performance varies among schools, with performance variation between schools in some of the best performing countries amounting to less than 10 per cent of students overall performance variation - so that parents in these countries can rely on high and consistent performance standards across the entire education system - whereas in other countries more than half of the OECD average performance variation originates at school and/or programme levels, often combined with only moderate overall performance (OECD, 2004). In Finland, the country with the strongest overall results in PISA, the performance variation between schools amounts to just 4 per cent of students overall performance variation such that parents can rely on high and consistent performance standards across the entire school system.

PISA has revealed wide differences in the extent to which countries succeed in fostering knowledge and skills in key subject areas as well as effective learning strategies. For some countries, the results were disappointing, showing that their 15-year-olds' performance lagged considerably behind that of other countries, sometimes by the equivalent of several years of schooling and sometimes despite

high investments in education. PISA also highlighted significant variation in the performance of schools and raised strong concerns about equity in the distribution of learning opportunities. However, the performance of countries such as Canada, Finland, Japan or Korea in PISA also revealed that excellence in education is an attainable goal. It also showed that the challenge of achieving a high and socially equitable distribution of learning outcomes can be addressed.

As a cross-sectional survey, PISA cannot show which policies or practices cause success, but it does reveal some common characteristics of students, schools and education systems that do well, which the presentation will examine. These include the strive to engage constructively with the diversity of student interests, capacities and socio-economic contexts; the shared commitment to professionalised teaching, in ways that imply that teachers are on a par with other professions in terms of diagnosis, the application of evidence-based practices and professional pride; as well as the move beyond systems of external accountability towards building capacity and confidence for professional accountability in ways that emphasize the importance of formative assessment and the pivotal role of school self-evaluation.

In a modern world, comparative assessments like PISA are an essential tool for educational improvement and research shows that the existence of standardised assessments and examinations is one of the most powerful predictors for the success of an education system. That is not hard to understand, because in the darkness, all students, schools and education systems look the same, it is impossible for teachers and school administrators to detect institutional and systemic strengths and weaknesses, and to support and intervene where expectations are not met. Comparative analyses like PISA will become ever more important, as the best performing education systems, not simply improvement by national standards, will increasingly become the yardstick for success. Countries will not simply need to match the performance of these countries but do better if their citizens want to justify higher wages. The world is indifferent to tradition and past reputations, unforgiving of frailty and ignorant of custom or practice. Success will go to those individuals and countries which are swift to adapt, slow to complain and open to change. The OECD will continue to join China in the search for effective strategies for addressing this challenge.

Andreas Schleicher

Head, Indicators and Analysis Division, Directorate for Education
The Organization for Economic Co-operation and Development

序 Preface

整个世界正在发生着日新月异的变化，联系日益密切的多元化人群，工作和生活中发展迅速的科学技术以及瞬息万变的海量信息对我们提出了新的要求。现代社会希望满足的其它要求还包括：平衡经济发展和自然资源可持续利用之间的关系，调和个人发展与社会凝聚力的关系以及降低社会不公平。人们知识、技巧和能力的发展对达到这些要求来说至关重要，而知识、技巧和能力的发展则需要通过教育体制，工作期间的不断学习和生活中的其它一些渠道获得。由此引出相关的问题：对现在和未来的世界来说什么能力才是最重要的呢？如何才能开发并培养这些能力呢？如何才能评价这些能力，从而为学生更好地学习、教师更好地教授以及学校更好地满足社会的需求提供有效的指导呢？

1997年，经济合作与发展组织（OECD）的成员国发起了学生能力国际评价（PISA），目的是在适当的基础上，在国际框架下评价对个人和社会的成功具有重要意义的关键能力。PISA不仅希望提供政策对话平台，而且希望能够为建设性地确定和执行教育目标提供合作基础，并使其反映与成人生活相关的技能评价。PISA评价显示出，在全球范围内，不论是在北美的加拿大、欧洲的芬兰，还是在亚洲的日本和韩国，一些教育体制证明卓越教育是一个能够达到的目标，而且成本通常都比较合理。此外，PISA评价还显示出，在整个社会公平分配学习资源的挑战能够成功实现，在落后学生和学校最少化的基础上能够在整个教育体制内获得优异成绩。

PISA中国试测研究在中国试测城市评价出的教育成效首次依据国际标准测量了学生关键能力的相对优势和不足，为中国的决策者和教育工作者提供了改进学校教育体制，为年轻人进入日新月异、全球依赖性逐步加深的成人生活做好准备的独特工具。

学生能力国际评价PISA

目前，PISA已成为全球最先进、最全面的国际评价，评价了占据全世界近90%的经济体。PISA将能力定义为在各种背景下利用社会心理资源成功达到复杂要求的才能，包括知识技能、动机、态度、情绪以及其他一些社会和行为因素。因此，PISA不仅仅是知识的简单重复应用。在不同的语言和文化中进行能力测评和比较是一件困难的事情，PISA一直在追求这个目标。PISA前三轮评价侧重于素养技能，也就是年轻人接触、处理、整合和评价信息的能力，想象性思考的能力，假设和发现能力以及有效传达思想和主见的能力。PISA评价的素养远比传统概念中读和写的技能广泛。此外，PISA界定的素养是作为一个统一体来评价的，而不仅仅说明某个人拥有或不拥有这些素养。从某种意义上来说，设定连续统一体素养的水平分级标准是必须的，也是众望所归的，但是不可忽视潜在的可变因素。素养的获得是件终生的事情，不仅在学校或通过正式学习能够获得，通过与家庭、伙伴、同事和更广泛的社区互动交往也能够获得。不能指望15岁的年轻人已经学到了成人所需要了解的一切东西，但他们却需要在阅读、数学和科学领域打下一个坚实的基础。为了在这些学科领域中继续学习并把所学的知识运用到现实世界，他们还需要知道在不同的情境下灵活运用相关知识的基本程序和原理。正是基于此，PISA在对主要概念广泛认知的基础上评价年轻人完成和现实与生活相关的任务的能力，而不是将目光仅仅局限在了解特定学科知识上。

PISA评价重点从学生是否能再现他们所学的知识转移到了学生能否从所学知识中向外推断并在异常情况下运用这种能力。此种做法是由现代生活所需要的知识和技能的本质造就的。那些能通过简单的记忆或预设法则解决的任务最容易被数字化和自动化，因此与现代知识社会距离较远。

- 2000年，PISA侧重评价阅读素养，测评学生运用、解释和思考书面材料的能力。
- 2003年，PISA侧重评价数学素养，测评学生运用数学知识的能力，利用各种各样的深度模型考察学生在不同情境下运用数学功能的能力。在学校里，数学是在抽象世界里传授的，通常剥离了真实的背景，PISA注重数学在真实世界中的功用，并获得成功。学生必须能在现实世界和数学世界之间建立联系，从而回答一些复杂的开放性问题。

PISA2006评价领域

	科 学	阅 读	数 学
界定与特色	学生程度： -掌握科学知识并利用相关知识识别问题、获取新知识、解释科学现象、得出与科学相关问题的结论。 -理解科学作为人类知识和探究的一种形式的典型特征。 -了解科学和技术是如何造就我们物质、智力和文化环境的。 -自愿解决与科学相关的问题，形成科学思想，做一个有思想的公民。 **科学素养**要求了解科学概念，并利用科学视角科学地思考相关现象。	学生为完成目标而理解、运用和思考书面信息的能力、开发个人知识和潜力的能力和参与社会活动的能力。 除了解读和字面理解，阅读素养还包括阐释、思考以及利用阅读实现生活目标的能力。 PISA对阅读的关注是将阅读作为一种学习方式，而不是单纯地学习阅读，因此并不会评价学生最基本的阅读技巧。	学生确认和理解数学在世界中所扮演的角色，做出有理有据的判断，以满足个人生活需要的方式利用和处理数学，做一个有建设性和思想性的公民。 **数学素养**涉及到对数学的广泛和功能性运用：包括在各种情况下认知和系统表达数学问题的能力。
知识领域	**科学知识**，例如： * 物理系统 * 生命系统 * 地球和空间系统 * 技术系统 **与科学相关的知识**，例如： * 科学探究 * 科学解释	阅读材料形式包括： * **连续性文本**：包括各种散文，记叙文、说明文和议论文 * **非连续性文本**：包括图形、表格和目录	相关的数学领域和概念包括： * **数量** * **空间与形状** * **变化与关系** * **不确定性**
涉及的能力	科学任务或程序类型： * **识别科学问题** * **解释科学现象** * **使用科学证据**	阅读任务或程序类型： * 提取信息 * 解释文本 * 反思并评价	能力群： * **复用**（简单的数学操作） * **联想**（组合概念直接解决问题） * **反思**（较广泛的数学思考）
背景与境况	科学应用领域，集中于个人、社会和全球背景下的运用，比如： * 健康 * 自然资源 * 环境 * 危机 * 科学技术前沿	文章的构架用途： * **个人**（例如：一封私人信件） * **公众**（例如：一份正式文件） * **职业**（例如：报告） * **教育**(例如：与学校相关的阅读)	数学应用领域，主要集中在与个人、社会和全球背景下的运用： * **个人** * **教育和职业** * **公众** * **科学**

因此，PISA的很多任务都是让学生在开放模型中面对现实问题。首先，学生必须将他们所面对的情况或问题转化为相关的数学形式；然后，他们必须对这些问题进行数学处理，并利用相关的数学知识解决问题；最后，学生需要在最初的问题情境中评价解决方案。

- 2006年，PISA侧重评价科学素养，测评关注领域：1）科学知识和利用科学知识去识别问题、获取新知识、解释科学现象以及推断出相关问题的可靠解决方案；2）理解科学作为人类知识和探究的一种形式的典型特征；3）了解科学和技术是如何造就我们物质、智力和文化环境的；4）自愿解决与科学相关的问题，形成科学思想，做一个有思想的公民。PISA更关注学生的学习倾向、学习方法、自我观念以及与学校的相处活动。

为什么说PISA评价的能力对个人发展来说至关重要

这个问题在于PISA所评价的关键能力是否对学生的未来成功起到预测作用。这可以通过检验PISA评价的能力如何与后续的教育、社会和经济成功相关联的分析来评价。

加拿大“年轻人转换调查”（Youth in Transition Survey，YITS）提供了一个检验这个问题的方法。这个纵向调查研究了年轻人生活中的主要教育、培训和工作转换的方式和影响。2000年，加拿大有29,330名15岁的学生参加了“YITS”和“PISA”。4年后，同一批学生在19岁的时候再次接受了教育成效评价（应答率：90.5%）。研究人员对后者的结果和学生15岁时在PISA中所获得的阅读成绩进行了对比研究，结果发现，15岁时在PISA阅读能力中获得1级水平的学生，在19岁时参加高等教育（postsecondary education）的比例高出了两倍，即使将学校教育、性别、母语、居住地、父母受教育水平和家庭收入因素考虑在内也是如此。而在PISA中获得4级阅读能力水平的学生受高等教育的几率增长到8倍，取得5级的学生几率增长到16倍。结果还显示，19岁前完成高中学业的年轻人比例随着他们在PISA中所获得的能力水平的提高而增长。总共有87%的学生在19岁之前完成了高中学业，而这个比例在阅读能力水平为1级或以下的群体中最低为62%，能力水平为2级的学生为77%。在将性别、母语、教学语言、父母受教育水平、家庭收入、学校和社会活动等因素考虑在内后，结果仍然保持不变。依次验证PISA的5个阅读能力水平的结果后，门槛效应变得非常明显，相对于在PISA阅读能力为3级或更高水平的学生来说，阅读能力为2级、1级或更低水平的学生完成高中学业的比例非常低。由于高中教育是继续学习和培训的基础，也是实现从学校到全职工作岗位成功转换的基础，这些数据由此表明15岁时阅读能力为2级或以下水平的学生潜在地面对着一些困难。通过分析，值得注目的结论是，PISA测试的能力水平对学生未来成功的预见性比学生在学校中获得的成绩更有效。

丹麦进行的类似研究也得出了相同的结论，在19岁前完成高中教育或普通职业教育的学生比例随着他们15岁时阅读能力水平的提高而大幅增长（参见网站：http://www.sfi.dk/sw19649.asp）。

PISA如何对教育政策和实践做出贡献

PISA旨在为学校、地方政府和国家提供认识自身优势和不足的机会，并通过与其他体制的比较对其进行分析。设计思想是通过与他者的比较，

使政府和实践者考虑如何才能进一步优化他们现存的教育体制，甚至是考虑改变现存教育体制下面隐藏着的一些陈规和信念。

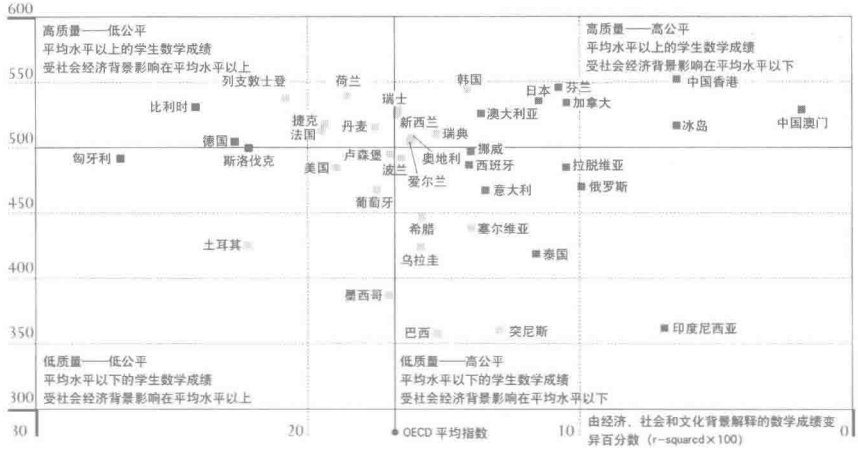
对PISA最初的争议主要集中在各种教育体制平均成绩的比较上。比较需要国际视角达成了一致共识，如果按国内标准测验出学生高分成绩比例较大，一些人便会认为国内教育体制已经得到改进，另一些人则会认为达到相关等级的国内标准要求太低。在这种较好的成绩是由较低的标准造成的质疑背后，通常隐藏着教育的总体成绩能否随时代要求提高的疑虑。PISA国际比较发现，一些国家在某些能力培养方面做得比另一些国家更出色，由此可以证实存在提高教育水平的可能性。

以学习机会的均等来评价公平性

PISA在考虑社会背景的基础上评价学生和学校的学习成效，从而能够阐释教育体制提供的教育机会分配的公平程度。PISA一些备受争议的有益发现是，在加拿大、芬兰或日本等国家学校成绩很差并不意味着学生就来自于不利的社会经济背景，而是教育体制能够成功地将较高的成绩水平和学习机会的均等在整个社会的公平分布有机结合（OECD，2004）。相反，一些国家数十年来仅仅依据学校平等投资一直认为学习机会的分配比较公平，然而，当PISA把其中的不公平现象揭示出来后，他们目瞪口呆。（见下图）

教育成效的质量和公平

在 PISA 数学量表上的国家平均成绩和成绩与经济、社会及文化地位之间的关系



PISA向政策制定者和实践者还提供了解决这些不公平现象的重要见解。有关公平的议题将通过学校教育体制剥离出学校内部社会经济不均匀性和社会经济分隔现象。这可以使人们评价教育体制是减缓还是加剧社会经济分隔程度的背景因素影响。在一些国家中，这种不公平很大一部分表现在学校层面，是教育政策所造成的后果。比如，在德国，从10岁起学生就会根据成绩表现进入不同类型的学校，并将决定是需要接受进一步的学业教育，还是职业教育。一方面，这种政策将相对类似的学生聚集到一起减少了学校内部的差异；另一方面，由于不同的学校课程设置不相同，从而增

加了各个学校之间的差异。因此，在一些国家中，来自较为优越的社会背景的学生被送进了声望较高的学院式学校，从而创造了较好的教育成效（在PISA中获得的成绩较高），而来自相对不利的社会背景的学生被送进了声望不是很高的职业学校，从而产生了较差的教育成效（在PISA中的成绩也较低）。由此，PISA评价显示德国学校组织体制反映和再现了德国当前的社会等级分割。

评价学校成绩的一致性

PISA在学校内部和学校之间区分学生的成绩差异。这种比较可以帮助人们评价学校对学习质量的影响程度，因此是教育体制的一个可预测的、一致性指标。通过比较发现，各个国家之间学校的不同对学生成绩的影响程度存在较大的差异。在一些成绩最好的国家，各个学校之间的成绩差异不到学生总体成绩差异的10%，因此这些国家的父母可以信赖整个教育体制成绩标准的一致性。但在另一些国家，一半以上的平均成绩差异是由学校或课程的差异造成的，而且学生的总体成绩也较一般（OECD，2004）。芬兰在PISA中获得的总体成绩最高。在芬兰，学生归属不同学校造成的成绩差异仅是学生总体成绩差异的4%，因此父母可以信赖整个学校教育体制成绩标准的一致性。

PISA评价显示，各个国家在主要学科领域的知识技能掌握以及推行有效的学习策略方面存在着较大差异。一些国家的结果让人沮丧，他们15岁学生的成绩远远低于其他国家，有时候是差几年的学业，尽管某些国家的教育投资较高。PISA还突出了学校之间的成绩差异，并对学习机会分配的公平性表现出很大程度的关注。加拿大、芬兰、日本和韩国等国在PISA评价中的表现还显示出，卓越教育是一个能够实现的目标，在全社会范围内获取优异教育成效的困难也是能够解决的。

作为一种横向评价，PISA没有指出哪种政策或实践能够导致成功，但确实揭示出成绩较好的学生、学校和教育体制的一些共同特征。这些特征包括尽量适应学生兴趣、能力和社会经济背景的多样化需求；对专业教师职业的忠诚，这意味着教师和其他专业人员是平等的，应用基于证据的实践，培养职业自豪感；此外，注重形成性评价的重要性和学校自我评价所扮演的重要角色，超越外部问责制度，培养专业责任能力和信心。

在现代社会中，类似PISA的比较评价是发展教育的必要工具。研究发现，标准化评价和测试的存在是一种教育体制成功与否最重要的指示器之一。这并不难理解，因为在黑暗中，所有学生、学校和教育体制看起来都是相同的，教师和学校的管理者也不能够发现制度和组织上的优缺点，没达到预期目标时也不能够干涉。因此类似PISA的评价比较就变得格外重要，因为成绩最好的教育体制不仅意味着国家标准的提高，而且将成为成功的准绳。一些国家不仅需要优异的成绩，而且在他们的公民有实力时需要做得更好。这个世界对传统和过去的名誉漠不关心，也不会原谅脆弱和无视国情或实践的做法。成功终将属于那些适时调整、疏于抱怨和乐于改变的个人和国家。经合组织将继续和中国一道寻找迎接挑战的有效策略。

Andreas Schleicher

经合组织教育委员会教育指标与分析部主任

前言 FOREWORD

随着教育的普及和发展，教育质量日益受到关注，教育评价也得到了越来越多的重视和运用。我国教育评价的改革与创新如何适应深化教育体制改革、全面实施素质教育、建设学习型社会的新形势，如何使教育评价成为促进教育发展和提高教育质量的有效途径，目前还缺乏系统的理论研究和成功的科学实践。

关注国外大规模教育评价项目的评价理念、技术、手段、结果及其发展动向，并通过实践，深入学习和熟知国外大规模教育评价项目的流程体系，掌握国际教育评价的设计理念和操作方法，为我所用，无疑将大大推动我国在教育评价领域的研究和发展。

学生能力国际评价（Programme for International Student Assessment, PISA）是经济合作与发展组织（The Organization for Economic Co-operation and Development，OECD）发起并组织实施的为各国协作监控教育成效的评价项目。PISA应用现代教育测量理论测试发达国家和地区义务教育结束阶段15岁学生在阅读、数学、科学领域的发展水平，配套调查问卷进而评价各参与国家与地区的教育成效，进行国际比较，超越了传统考试手段的局限性，是世界上颇具影响的国际教育评价项目之一。PISA在2000年首次开始评价，每三年一次，以评价年命名。PISA2000的评价有32个国家参与，PISA2003有41个国家和地区参与，PISA2006有56个国家和地区参与，PISA2009将有68个国家和地区参与。为了保证评价的效度和信度，由来自各参与国家和地区的教育政策制定者和相关领域的专家共同决定评价的范围、本质、学生背景信息收集等，评价材料也考虑到不同的文化和语言，其翻译、取样和资料收集过程都采取了严格的质量监控机制，并通过实施大规模实地预试等各种手段，将测试在各个国家正式实施中可能存在的误差降到最小。

OECD/PISA研究的目的是形成一套指示或报告，用来说明，为了把15岁的学生培养成为积极的、善于思考的、有智慧的公民，从他们运用科学、数学、阅读技能的角度看，各个国家的教育成效如何。为了达到这个目的，PISA创立了科学、数学、阅读评价量尺，评价的焦点是确定学生运用所学知识的能力水平。PISA的评价是前瞻性的，测量的是15岁青少年对于迎接现今高科技和知识社会的挑战的准备情况。PISA关注年轻人运用知识技能处理现实生活的挑战能力，而不仅仅是考查他们对学校特定课程的掌握程度。PISA的测量目标是常规的、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以及教育对社会经济的影响。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。

国家教育事业发展“十一五”规划纲要指出要建立素质教育评价检查体系，逐级考核各级政府落实素质教育工作的情况。各级教育行政部门正积极筹备建立相应的教育评价体系和评价标准，教育评价研究与实践也初现繁荣局面。科学有效的教育评价可以实现跨省、跨地区的横向比较，总结出各地教育的成功与不足；也可以提供纵向跨年度的比较，使各级教育行

政部门了解本地区教育教学发展的状况，从而制定出相应的措施，达到预期的教育目标。

教育部考试中心戴家干主任指出："当前，以考试代替评价是素质教育难以落实，教育改革难以突破的关键问题。利用评价的观念和方法改造我们的考试，利用考试的数据开展评价工作，是实现考试改革和推动考试评价的最佳结合点。在传统考试的基础上构建一个科学的教育考试与评价体系，实现从单一考试到多元评价的跃升，既是考试改革的核心问题，也是教育改革的关键所在。"

为此，教育部考试中心2006年引进并启动了PISA2006中国试测研究项目，作者正是此项目负责人。写本书的目的是通过介绍PISA的背景、理论基础和先进的手段和技术，以及作者在实践过程中的一点体会，希望给对教育评价改革有兴趣的教育工作者提供一点参考与借鉴。

本书相关内容得到了主要引用文献版权拥有者的准许，适合于相关教育行政工作者，考试与评价机构的相关人员，也适合于学校、教研机构、考试机构命题工作者。

本书能顺利出版首先应当感谢OECD/PISA国际合作和共享的理念。PISA2006中国试测研究的成功离不开OECD教育委员会教育指标与分析部主任Andreas Schleicher先生，美国加州大学伯克利分校教育研究院Mark Wilson教授，中国香港PISA中心主管何瑞珠女士、邝伟良先生，台湾中正大学王文中教授，澳大利亚教育研究院Ray Adam先生、 Margaret Wu女士、Bibby Yan女士和中国澳门PISA中心原主管卢林发先生的大力支持，更离不开教育部、教育部考试中心及试点地区领导和同仁、学科专家、教育测量和数理统计师生的全力协助。国内参与PISA2006中国试测研究人员因人员保密规定不能一一列出，在此一并衷心致谢！最后谨将本书献给爱我、包容我的家人，他们给了我无微不至的生活照顾和莫大的精神支持。

由于时间仓促、作者自身条件局限，书中肯定有很多不足之处，敬请读者指正。

王蕾

2007年8月于北京