JOURNAL OF APPLIED MEASUREMENT, *8*(3), 237-248

The Programme for International Student Assessment: An Overview

Ross Turner
Australian Council for Educational Research

Raymond J. Adams
Australian Council for Educational Research
University of Melbourne

This paper provides an overview of the Programme for International Student Assessment (PISA), an on-going international comparative survey study of educational outcomes at age 15. PISA is sponsored by the Organization for Economic Co-operation and Development (OECD) and for the period 1998-2010 has been designed and implemented by a consortium led by the Australian Council for Educational Research (ACER).

Requests for reprints should be sent to Ross Turner, Australian Council for Educational Research, 19 Prospect Hill Rd., Camberwell, VIC 3124, Australia, e-mail: turner@acer.edu.au.

What is PISA?

The Programme for International Student Assessment (PISA) was commissioned by the Organization for Economic Co-operation and Development (OECD) in the late 1990s as an ongoing, periodic international comparative study of the proficiency in Mathematics, Science and Reading of 15 year old students. PISA is managed and directed cooperatively by the OECD member countries, and in cooperation with a large and increasing number of non-member countries, referred to as *partner* economies.

PISA surveys take place every three years. The first survey took place during 2000, and the second in 2003; the results of these surveys have been published through the study's initial reports (OECD, 2001, 2003a, 2004b) and a wide range of thematic and technical reports.[1] The third survey was conducted during 2006, and the next will occur in 2009. For each assessment one of the three areas (Mathematics, Science and Reading) is chosen as the major domain and it is given greater emphasis. The remaining two areas, the minor domains, are assessed less thoroughly. In 2000 the major domain was reading, in 2003 it was mathematics and 2006 science. In 2009 reading will again be the major domain.

The design and implementation of PISA for the 2000, 2003, 2006 and 2009 data collections has been the responsibility of an international Consortium led by the Australian Council for Educational Research (ACER) with Ray Adams as international project director. The other partners in this Consortium have been the National Institute for Educational Measurement (Cito Group) in the Netherlands, the German Institute for International Educational Research (DIPF), the Service de Pédagogie Expérimentale at Université de Liège and the CAPSTAN Linguistic Quality Control Company in Belgium, Westat and the Educational Testing Service (ETS) in the United States and the National Institute for Educational Policy Research (NIER) in Japan.

[1] See www.pisa.oced.org.

PISA is an age-based survey, assessing 15-year-old students in school in grade seven or higher. These students are approaching the end of compulsory schooling in most participating countries and school enrolment at this level is close to universal in almost all OECD countries.

The PISA assessments take a *literacy* perspective that focuses on the extent to which students can use the knowledge and skills they have learned and practised at school when confronted with situations and challenges for which that knowledge may be relevant. That is, PISA assesses the extent to which students can use their reading skills to understand and interpret various kinds of written material that they are likely to meet as they negotiate their daily lives; the extent to which students can use their mathematical knowledge and skills to solve various kinds of mathematics-related challenges and problems; and the extent to which students can use their scientific knowledge and skills to understand, interpret and resolve various kinds of scientific situations and challenges.

PISA also allows for the assessment of additional cross-curricular competencies from time to time as participating countries see fit. For example, in the 2003 survey cycle, an assessment of general problem-solving competencies was included. Further, the PISA survey collects information from students on various aspects of their home, family and school background; and information from schools about various aspects of organization and educational provision in schools. This information is collected to facilitate a detailed study of factors within and between countries that are associated with varying levels of reading, mathematical and scientific literacy among the 15-year-old students of each country.

Features of PISA—Organizational

The PISA project can be thought of as operating at a number of levels. The different levels and key participants in the project are illustrated in Figure 1.

The PISA project is overseen by a Secretariat within the OECD in Paris. The project policy

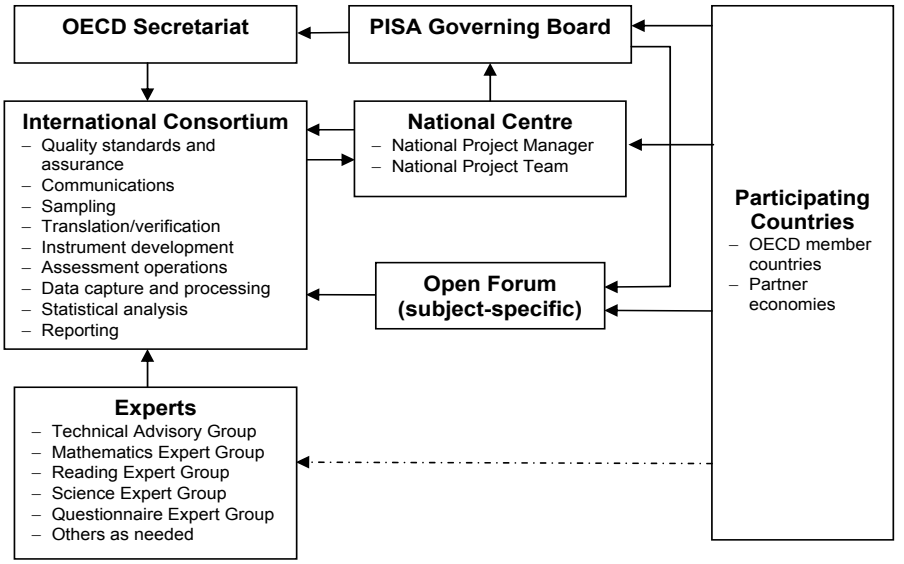

Figure 1. Key Participants in the PISA Project

parameters are set by the PISA Governing Board (PGB), which is an OECD committee comprising delegates and observers from the participating countries. It meets about twice each year. PISA therefore is directed through a collaborative process involving senior educational administrators from participating countries.

Within each participating country, a national centre is established. A National Project Manager (NPM) is appointed to coordinate all activities at the national level. Typically, the NPM works closely with the country's PGB member to establish a national perspective on policy matters, on matters related to project implementation, and on the analysis and reporting of outcomes that may be of particular relevance to the country. Generally the national centre has a small team working on project development, implementation and reporting at the national level.

The project is implemented internationally by a contractor appointed by the OECD to carry out this work. A contractor is selected for each assessment cycle following an open competitive tendering process. An international consortium led by the Australian Council for Educational Research (ACER) with Ray Adams as international project director has been the contractor for each of the first four PISA survey cycles. The contractor is responsible for implementing all aspects of the assessment, including:

- developing quality standards in relation to all aspects of the project;
- developing procedures to ensure that those quality standards are properly met by all participants, and mechanisms for monitoring the quality of project implementation;
- maintaining open and effective communications among all project participants;
- working closely with National Centre personnel to gather national input to matters related to project development and implementation;
- developing the assessment frameworks;
- developing all assessment and survey instruments;
- developing and implementing sampling plans;
- developing operational procedures for test administration and all related documentation;

- training key national centre staff in the requirements for implementing the study;
- developing all data capture procedures;
- capturing and processing data from the assessments;
- analyzing the results and preparing material to assist the OECD in producing the reports they require; and,
- assisting the Secretariat with review of procedures and planning improvements for future survey cycles.

To carry out all of these tasks, the contractor is required firstly to put together a large team of people with expertise collectively in a variety of areas. Some of this expertise resides in the staff available through its consortium partner organizations. In addition, the consortium works with several key groups and individuals: a number of expert groups comprising internationally recognized experts in areas such as the particular cognitive assessment domains, translation experts, technical specialists in sampling, statistical analysis and development of questionnaires. The contractor engages other expertise when required.

Most importantly, the contractor works very closely with the personnel engaged at the national level. The connection between NPMs and the international consortium is critical to the success of the project. Each depends on the other to ensure successful implementation of the project. National centres provide the consortium with information about conditions and constraints operating in the country, with feedback regarding the various project elements that are being developed, and with important advice on how the project can best be implemented in that country. The contractor provides national centres with information about project requirements, draft materials for national consideration and feedback, training and materials to facilitate project implementation.

Other consultative mechanisms have been used from time to time, such as the Science Forum for PISA 2006. This was an open forum that provided the opportunity for participating countries to nominate national experts who could directly represent the interests and views of the country in considering certain detailed technical aspects of the project. It allowed for a wider base of expert input than is possible through the international contractor's expert groups. In the case of the Science Forum, this group considered priorities and issues at the time the science framework was being conceptualized, and it provided important input in the development of survey material related to the assessment of science and the assessment of student attitudes to science. Delegates to the forum were nominated by each country's PGB member. A similar Mathematics Forum operated in PISA 2003, and a Questionnaire Forum has been convened from time to time, and a Reading Forum currently operates for PISA 2009.

Who participates?

With the completion of PISA 2009, 65 countries will have participated in one or more PISA surveys. They are listed in Table 1.

Features of PISA—Technical

The technical characteristics of the PISA survey involve a number of different elements. The design of the test, and the features incorporated into the test developed for PISA are critical features. The sampling design, including both the school sampling and the student sampling requirements and procedures are a second critical area. Features related to the multi-lingual nature of the test are a further area raising a number of important technical issues. This involves the rules and procedures designed to guarantee the equivalence of the different language versions used within and between participating countries while taking into account the diverse cultural contexts of those countries. Various operational procedures, including test administration arrangements, data capture and processing and quality assurance mechanisms designed to ensure the generation of comparable data from all countries form another area of technical focus for the project. Then of course the technicalities related to scaling and

Table 1

Countries participating in PISA

National Centre Name	2000[1]	2003	2006	2009[2]
Albania	✓			✓
Argentina	✓		✓	✓
Australia	✓	✓	✓	✓
Austria	✓	✓	✓	✓
Azerbaijan			✓	✓
Belgium	✓	✓	✓	✓
Brazil	✓	✓	✓	✓
Bulgaria	✓		✓	✓
Canada	✓	✓	✓	✓
Chile	✓		✓	✓
China (Shanghai)				✓
Hong Kong - China	✓	✓	✓	✓
Macao - China			✓	✓
Chinese Taipei			✓	✓
Colombia			✓	✓
Croatia			✓	✓
Czech Republic	✓	✓	✓	✓
Denmark	✓	✓	✓	✓
Dominican Republic				✓
Estonia			✓	✓
Finland	✓	✓	✓	✓
France	✓	✓	✓	✓
Germany	✓	✓	✓	✓
Greece	✓	✓	✓	✓
Hungary	✓	✓	✓	✓
Iceland	✓	✓	✓	✓
Indonesia	✓	✓	✓	✓
Ireland	✓	✓	✓	✓
Israel	✓		✓	✓
Italy	✓	✓	✓	✓
Japan	✓	✓	✓	✓
Jordan			✓	✓
Kyrgyzstan			✓	✓
Latvia	✓	✓	✓	✓
Liechtenstein	✓	✓	✓	✓
Lithuania			✓	✓
Luxembourg	✓	✓	✓	✓
Macedonia	✓			
Mexico	✓	✓	✓	✓
Netherlands	✓	✓	✓	✓
New Zealand	✓	✓	✓	✓
Norway	✓	✓	✓	✓
Peru	✓			
Poland	✓	✓	✓	✓
Portugal	✓	✓	✓	✓
Qatar			✓	✓
Republic of Korea	✓	✓	✓	✓
Romania	✓		✓	✓
Russian Federation	✓	✓	✓	✓
Scotland	✓	✓	✓	✓
Montenegro			✓	✓
Serbia		✓	✓	✓
Singapore				✓
Slovak Republic		✓	✓	✓
Slovenia			✓	✓
Spain	✓	✓	✓	✓
Sweden	✓	✓	✓	✓
Switzerland	✓	✓	✓	✓
Thailand	✓	✓	✓	✓
Trinidad and Tobago				✓
Tunisia		✓	✓	✓
Turkey		✓	✓	✓
United Kingdom	✓	✓	✓	✓
United States of America	✓	✓	✓	✓
Uruguay		✓	✓	✓

[1] Includes participation in PISA plus (OECD 2003a), which was a replication of PISA 2000 run in 2001.
[2] The list of 2009 participants is provisional.

analysis of the data and their subsequent reporting form a further major set of issues. PISA scaling employs models based on Rasch methodologies. The use of described proficiency scales as a basic tool in reporting PISA outcomes also are derived using Rasch analysis. Each of these technical areas will be briefly discussed in turn in the following sections. Greater detail is provided on each of these areas in the Technical Reports for PISA 2000 and PISA 2003 (Adams and Wu, 2002, OECD 2005a).

Quality standards within each of these technical areas are defined, monitored and assured through the use of a set of technical standards. These standards have been endorsed by the PGB, and they form the backbone of project implementation in each participating country and of quality assurance across the project.

Test Design and Development

PISA has so far been implemented using pencil-and-paper tests.[2] Students are expected to undertake two hours of testing in the *cognitive* test that covers the domains of reading, science and mathematics. For the PISA 2006 assessment, a number of test items aimed at exploring student attitudes to science were embedded in the cognitive part of the test. In addition, students complete a short questionnaire designed to gather relevant background data about the student's personal characteristics, opinions, preferences and aspirations, some characteristics of his or her home and family environment, and some characteristics of his or her school environment. This is designed

[2] PISA 2006 included an optional computer-based assessment of scientific literacy. This option was taken up by three countries.

for students to complete in 20-30 minutes. School principals also complete a short questionnaire about broader aspects of the school context.

The cognitive part of the test must provide suitable coverage of each test domain and must generate data related to the several constructs laid out in each of the assessment frameworks (OECD 2006). In other words, the development of assessment instruments commences with an explicitly stated set of constructs to be targeted by specific test items. The constructs encompass a range of aspects of subject content within each domain, a range of learning processes relevant to each domain, and a variety of contexts that are used in the presentation of test items to ensure that no particular set of interests and experiences is unfairly over-represented and so that the wide variety of student experiences in different national contexts is adequately represented in the test. The PISA tests are designed to measure the extent to which students can use the range of knowledge and skills they have acquired at school, as they attempt to solve the kinds of problems they will confront in non-school contexts. The emphasis is not on assessing specific curricular outcomes, but on the application of acquired knowledge in a variety of real-life contexts.

A large volume of test material is therefore developed in each assessment cycle to ensure suitable coverage and balance across the various constructs and aspects of each domain framework, and the material is distributed across a number of test booklets in a rotated test design (a balanced incomplete block design). Each sampled student is randomly assigned one of the test booklets.

Rasch analysis is used during the test development process to check the characteristics of the items developed prior to their finalization and selection for inclusion in the main survey instruments. In particular, the extensive field trial that takes place in the year preceding the main survey in each assessment cycle generates student response data on all items that are being considered for inclusion in the main survey item pool. ConQuest (Wu, Adams and Wilson, 1998) is used as the main analysis tool. The standard item statistics generated by ConQuest as well as various Rasch fit statistics and diagnostic indicators are used as primary tools in reviewing item performance. These include indices of discrimination and fit to the model, point-biserial correlations, the mean ability of students by response category, a check of category ordering for partial credit items and the consistency of these indices across countries, the expected and observed score curves by gender and by country, and the expected and observed item characteristic curves by response category.

The information from these analyses is used as the basis of item selection for the main survey. It is also typically used as the basis for identifying items that have characteristics that render them unsuitable for use in the PISA assessment instruments, or items that need to be revised (for example evidence from an unexpected item performance in an individual country that leads to uncovering of a translation error). Information about item difficulty is particularly important in the construction of survey instruments, since the instruments should contain test items with an acceptable mix of difficulties within each of the relevant framework categories.

PISA test items are presented in several formats: multiple-choice, short-answer, and extended response. Multiple-choice items are either standard multiple-choice with a limited number of responses (usually four) from which students are required to select the best answer or complex multiple-choice presenting several statements for each of which students are required to choose one of two or more possible responses (true/false, correct/incorrect, etc.). Short-answer items include both closed constructed-response items that generally require students to construct a response within very strict constraints (such as mathematics items requiring a numeric answer), and items requiring only a word or short phrase. Short response items are similar to closed constructed-response items, but for these a wider range of responses is possible. Extended constructed-response items have a much wider range of acceptable responses. They typically require more extensive writing, or showing a calculation, or demonstrating a chain of reasoning, and frequently demand some explanation or justification.

In previous PISA cycles, it has been successfully argued by the test developers and by domain experts, and reinforced by national feedback on items as they were in development, that the range of cognitive processes that can be exposed and tapped is much greater when open formats are used than would be the case if only closed form items such as multiple-choice and short response items were used. The PGB has taken the view that the additional costs involved with coding and processing responses from these more open items are justified by the increased power and richness of data derived from using a wider range of test item formats. This view has also been reinforced by research on item format using PISA data (Routitsky and Turner, 2003) that has shown the importance of using a variety of test item formats to cater for the full range of student abilities typically sampled in PISA.

Student responses to more than half of the cognitive test items in the PISA 2006 main survey were able to be processed by computer. The remainder, a total of 80 of the 185 items (that is, 43%) required intervention by a trained coder in order to process student responses.

A common battery of questionnaire items was chosen for the background questionnaires. The purpose of the background questionnaires was to identify social, cultural, economic and educational factors that are associated with student performance. This would make it possible to explore the relationships between student performance outcomes on the cognitive tests and various student-level and school-level factors, and to see how these factors might vary across systems, across countries, and across time.

Sampling in PISA

PISA sampling is carried out in two stages according to a procedure that is designed to assign all eligible students in each participating country a known probability of being chosen to participate. The international population definition enables construction of a sampling frame that comprises all 15-year-old students in school that are in grade seven or higher. First, schools that contain eligible students are randomly sampled with probability proportional to size. Then 35 eligible students are randomly sampled from within each sampled school. In other words, the students who are sampled for the PISA tests are randomly selected and truly represent the population of 15-year-old students in school in each participating country.

A minimum of 150 schools are sampled in each country (or all schools if there are fewer than 150 containing eligible students). The target student sample size of 35 per school means that a minimum of 5250 students from each country would be sampled, with the expectation that a minimum of 4500 students would be assessed. If fewer than 35 students are available in a large enough number of schools, then additional schools are sampled to ensure an adequate minimum total student sample size.

Standards are applied to ensure i) adequate coverage of the eligible population (involving strict rules about which schools and students could be legitimately excluded), ii) adequate accuracy and precision in the estimates derived from the sample (involving strict rules about the required sample size), and iii) adequate response rates for both schools and students (involving strict rules about response rates, including procedures for using replacement schools where needed to reach acceptable school response rates and decision rules about inclusion or exclusion of student data depending on student response rates).

Translation and Cultural Appropriateness of PISA Material

PISA is one of the largest ever international studies of its kind. In PISA 2006, for example, it involved test administration in at least 150 different schools in each of the 57 countries participating, with 81 national versions of assessment instruments in 42 different languages. In such an enterprise, the need to ensure comparability of the test material across all test administrations is no small matter. The first part of achieving this lies in ensuring that the test materials themselves are appropriate for use across such culturally diverse settings and that the different versions used in those different settings are equivalent.

The approach to ensuring cultural appropriateness is first to use a wide variety of materials, representing different cultural experiences and contexts, then to process and refine those materials to ensure that different interests are well balanced, and to empirically test that all selected materials work well in all countries. The mechanisms used to ensure that materials developed for use in PISA are culturally appropriate include the following:

- Test materials (questionnaire and test items) are sought from the widest possible range of sources, including seeking national submissions from all participating countries. Test development procedures are used that are overseen by international experts in the relevant field and conducted by test development experts from a variety of countries and cultural contexts.
- Several opportunities are provided for all participating countries to review and comment on the material under development; and material is also reviewed by panels of international experts in each development area.
- Cognitive laboratory and other pilot activities are conducted with material under development using real students in several different countries.
- A large-scale formal field trial is conducted with students in all participating countries to test the functioning of material under development.
- The results of statistical analysis of field trial data are used to empirically evaluate the material as implemented in all countries and to detect instances of test items behaving differently in different countries.
- All material is revised on the basis of information received from each of these different mechanisms and only material that is demonstrated to work is selected for use in the main PISA assessment.
- The decision about which material is finally selected is reviewed by NPMs, and endorsed by the PGB.

In parallel with the development of culturally appropriate material for use in the PISA assessment instruments, source versions of all material are prepared in both English and French as a precursor to the development of equivalent national versions. From these two source versions, national centres in each participating country then prepare their own national versions of the test material, using a tightly controlled process of double independent translation, expert reconciliation of the two versions so produced, and independent international expert verification of the final translated versions.

An extensive field trial is conducted to test the translated materials of each country and analyses of field trial data are used to empirically evaluate the equivalence of the different language and national versions. Information about the characteristics of items as they occur in the different language instruments is obtained from the Rasch analyses mentioned previously. One particularly important output is the item-by-country interaction data (a form of DIF analysis) that are used to expose any items that behave differently when presented in a particular language or culture. This is essential in ensuring that the test instrument used for the main survey comprises items that together are capable of generating a single international set of item difficulty parameters that can be used for measurement of students.

Materials are revised on the basis of the field trial data analyses and the final selection of material for the main PISA study is chosen to ensure that only fully functioning test items across all national and language versions are selected. At the conclusion of this process, each national version that is produced is as linguistically and psychometrically equivalent as possible to the source versions, and therefore capable of contributing to the estimation of a single international set of item parameters.

Field Operations

The second major way in which comparability of test results across such a diverse range of countries and settings is ensured lies in the standardization of test administration procedures. An extensive array of procedures has been developed and documented to assist all participating countries to administer PISA test sessions in a way that facilitates the generation of internationally comparable test data.

A NPM's manual describes all procedures to be developed and implemented by each national centre, including involvement in various consultation and review procedures, the implementation of sampling procedures, implementation of all procedures related to the preparation, production and dispatch of test materials, recruitment and training of test administration personnel and oversight of test administration, assistance with implementing quality monitoring procedures, recruitment and training of personnel to code student responses, management of the coding of student responses and the entry of student response data, processes related to the capture and preparation of all PISA data for submission to the international study centre and subsequent processes related to assisting with the analysis of data and the reporting of results.

Separate manuals cover specific operational procedures related to sampling, translation, test administration, test centre coordination, coding of student responses, data management, and related to the specialised data capture software used in the project. In addition to this extensive documentation, the international study centre conducts several meetings of NPMs and other key national centre staff with the main purpose to provide information and training related to field operations and all other aspects of project implementation in each country.

Through the extensive field operations documentation, the scheduled training and briefing meetings, and by using regular communication via e-mail and telephone, the international study centre ensures that PISA testing procedures could be consistently applied in all participating countries. To check the extent to which those procedures were in fact applied consistently, a variety of quality-monitoring procedures are implemented and are described in the following section.

Monitoring Quality

Implementation of the PISA project is built around a set of quality standards that relate to the various aspects of the project. Standards exist in relation to the definition of the target population, sampling, language of testing, preparation of tests and manuals, test administration, print quality of test materials, security of materials, and so on.

For each such area, a plan is established that describes the manner in which the standard is to be met by each participating country. Evidence is generated as the project proceeds that show whether each standard has been met. Control processes exist through which national PISA centres continuously ensure the quality of project implementation. External monitoring procedures are applied that either enable early intervention by the international study centre in cases for which it appears a standard may not be met in a particular participating country or intervention after the fact to determine if the data collected can safely be used for the intended purpose.

Ultimately, an adjudication process is in place through which a final determination is made as to whether or not the data from a participating country are fit for use and can therefore be included in the international data set. If they are not deemed to be fit for use, then a country's data can be excluded from some or all of the international reporting of PISA outcomes. The adjudication process involves senior consortium staff, members of the technical advisory group including the sampling referee, and the OECD Secretariat. Recommendations from that group are made to the PGB for final decision.

Scaling and Analysis of Data

The scaling of PISA data rests on a simple assumption that there is some underlying trait or set of underlying traits of interest (these traits are defined in the assessment frameworks across the

several test domains) that each form a continuum, or scale; that test items can be developed that embody each underlying trait (by demanding different amounts of that trait); and that the amount of the trait possessed by students can be estimated by observing their responses to the test items. Typically, we think of the continuum as a line, with *more* of the trait in one direction, and *less* of the trait in the other direction; that test items can be placed along the line according to the amount of the underlying trait that they demand; and that students can also be located along the line according to the amount of the underlying trait that they possess. The Rasch analysis provides a means for constructing an interval scale for these measures. Figure 2 provides a graphic characterisation of a PISA literacy scale, labelled as the Mathematics scale, but the same idea applies to scales developed in any of the PISA test domains.

PISA uses a Rasch-based form of item response modelling in order to scale the student data to derive the various comparative measures that are produced and reported by the OECD. The model is a mixed coefficients multinomial logit model, which is a generalized form of the Rasch model. Essentially it uses student responses to a set of test items to simultaneously derive estimates of the 'difficulty' of the test items, and of the 'ability' of the sampled students, enabling both test items and students to be located along a continuum that is defined by the underlying trait being measured. Details of the model used and the scaling methods applied can be found in the PISA 2000 Technical Report (Adams and Wu 2002), and Adams and Wu (2006) and Adams, Wu and Carstensen (2006).

In PISA, the application of these models also permits observed student responses to test items, or more specifically the ability measures obtained from these, to be linked through regression equations to various background variables, such as gender, socioeconomic background, and so on. One outcome of these analyses is the generation of estimates of population means and

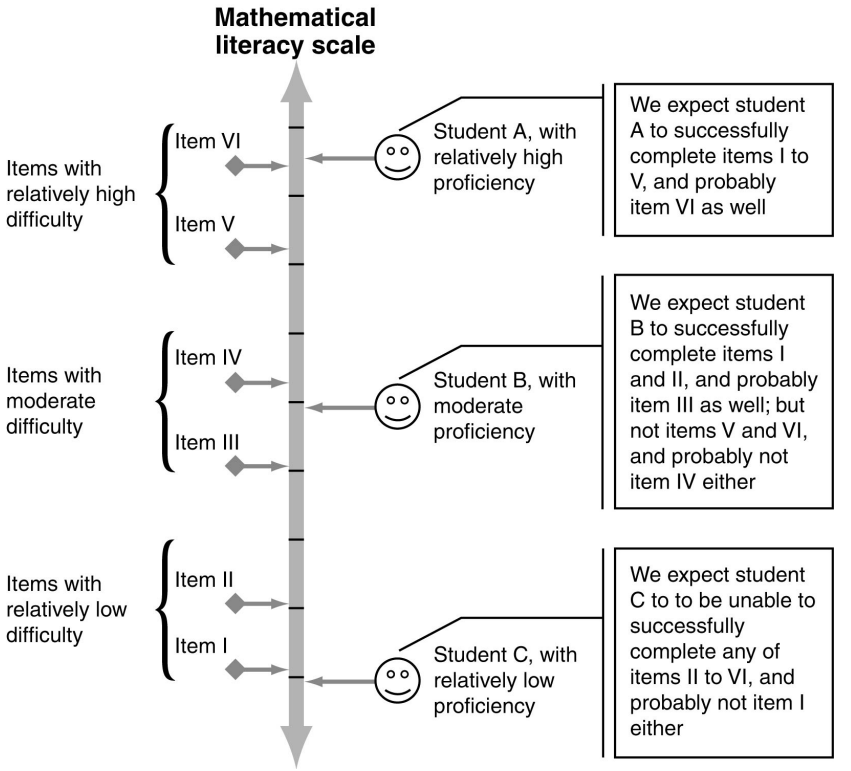

Figure 2. A characterisation of a PISA literacy scale

other statistics that enable comparisons among groups of students between and within PISA sampled populations. It is those comparisons among groups of students, linking performance in one or more of the cognitive test domains with various background variables, that enable policy-oriented researchers to investigate factors that may influence student performance and to consider any implications for school management or other school organizational features, teaching and learning practices, and so on.

The design of PISA test instruments, along with the scaling and analysis of PISA data, also permits monitoring of trends in test outcomes across PISA assessment cycles. Some of the test material from PISA 2000 was also used in PISA 2003, and similarly there is test material common to the PISA 2003 and PISA 2006 assessments. This enables measures of trends to be derived, through which the monitoring of changes over time can occur. It might be expected that such changes would be negligible from one cycle to the next. However, as PISA continues across further cycles, it can be expected that adjustments to policy settings and to teaching practices in different countries might lead to more noticeable changes in PISA test outcomes over longer time periods. In the first PISA test, occurring in 2000, reading was the major test domain. Any changes in reading outcomes related to changes in policy and practice that were effected in response to PISA 2000 results (or indeed other unrelated changes) might be expected to have an impact on PISA 2009 results, when reading is again the major test domain.

Reporting of PISA Outcomes

Following data collection in each PISA assessment cycle so far conducted, the OECD has produced a comprehensive report that captures the major outcomes from an international perspective. Two such reports have been produced, on the PISA 2000 outcomes (OECD 2001) and the PISA 2003 outcomes (OECD 2004b) respectively.

Those main international reports have provided an overview of the PISA project, details that help understand key features of the major assessment domain (reading in the 2000 report, mathematics in 2003) and how results should be interpreted in relation to the framework, cross-country comparisons of results in the major assessment domain and their relationship with some of the key student background variables, an overview of outcomes in the minor assessment domains, an overview of the information gained from the various student-level and school-level background variables captured and differences in these among countries, international comparisons of aspects of the learning environment and the organization of schooling, and some discussion of the policy implications of these various aspects of the study.

In reporting the literacy outcomes within the major assessment domain, emphasis has been placed on the profile of student results in each country in relation to the scales and subscales that come out of the relevant framework. For example, in the case of mathematical literacy, results are reported for an overall mathematical literacy scale and the results are also 'pulled apart' and reported for sub-scales that are based on the four 'content' areas of the mathematics framework. Central to the profile of student results is a set of descriptions of what students located at various points along the literacy scale would typically be able to do. The mathematics proficiency descriptions provide a clear picture of the way students are able to draw on the various mathematical competencies that are described in the framework. They describe growth in mathematical literacy in relation to an increasing student capacity to demonstrate and draw on those competencies. The main international OECD report places significant emphasis on the relative proportions of students in each country performing at various levels along the literacy spectrum.

In reporting on the background variables (student, family, school and system factors), differences among countries with respect to those variables are described, and the relation between some of those variables and student cognitive outcomes are analyzed in an attempt to describe how these factors play out in different countries.

In addition to the main international reports, the OECD produces or promotes a number of additional reports. Following the first assessment cycle, detailed reports on outcomes of the assessment of reading were published (OECD 2002c, 2003b). A number of other reports on particular aspects of the PISA 2000 assessment outcomes have also been published by the OECD, covering such matters as student engagement, student approaches to learning, school factors related to quality and equity (OECD, 2003b, 2004a, 2005b), and others. In addition, the OECD publishes a number of more technical documents including the technical report, a database manual, sample items, and the framework documents (OECD 2002a; OECD 2002b). A similar range of reports relating to the outcomes of the PISA 2003 survey cycle have also been published by the OECD, or are in preparation. These and other PISA publications can easily be accessed through the OECD's web site at http://www.pisa.oecd.org/.

Many PISA countries also produce their own national reports, giving greater detail of various outcomes within the country, or at least providing a clearer national perspective on the results, and providing in many cases more detailed analyses and interpretations that take into account various factors operating in the particular country.

References

Adams, R. J., and Wu, M. L. (2002). *PISA 2000 technical report*. Paris: OECD Publications.

Adams, R. J., and Wu, M. L. (2006). The mixed-coefficient multinomial logit model: A generalized form of the Rasch model. In M. v. Davier and C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 57-76). New York: Springer Verlag.

Adams, R. J., Wu, M. L., and Carstensen, C. H. (2006). Application of multivariate Rasch models in international large scale educational assessment. In M. v. Davier and C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 271-280). New York: Springer Verlag.

OECD (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publications.

OECD (2002a). *Manual for the PISA 2000 database*. Paris: OECD Publications.

OECD (2002b). *Sample tasks from the PISA 2000 assessment reading mathematical and scientific literacy*. Paris: OECD Publications.

OECD (2002c). *Reading for change: Performance and engagement across countries*. Paris: OECD Publications.

OECD (2003a). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris: OECD Publications.

OECD (2003b). *Student engagement at school, a sense of belonging and participation: Results from PISA 2000*. Paris: OECD Publications.

OECD (2004a). *What makes school systems perform: Seeing school systems through the prism of PISA*. Paris: OECD Publications.

OECD (2004b). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publications.

OECD (2005a). *PISA 2003 technical report*. Paris: OECD Publications.

OECD (2005b). *School factors related to quality and equity: Results from PISA 2000*. Paris: OECD Publications.

OECD (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Paris: OECD Publications.

Routitsky, A., and Turner, R. (April, 2003) *Item format types and their influence on cross-national comparisons of student performance*. Presentation given to the annual meeting of the American Educational Research Association. Chicago.

Wu, M. L., Adams, R. J., and Wilson, M. R. (1998). ACER ConQuest: Generalized item response modeling software [Computer program]. Melbourne: ACER Press.