JOHN O. ANDERSON, HUANN-SHYANG LIN, DAVID F. TREAGUST,
SHELLEY P. ROSS and LARRY D. YORE

USING LARGE-SCALE ASSESSMENT DATASETS FOR RESEARCH IN SCIENCE AND MATHEMATICS EDUCATION: PROGRAMME FOR INTERNATIONAL STUDENT ASSESSMENT (PISA)

Received: 30 May 2007; Accepted: 30 May 2007

ABSTRACT. Large-scale assessments of student achievement provide a window into the broadly defined concepts of literacy and generate information about levels and types of student achievement in relation to some of the correlates of learning, such as student background, attitudes, and perceptions, and perhaps school and home characteristics. This paper provides an overview and outlines potential research opportunities of one such assessment—the Programme for International Student Assessment (PISA). In order to provide examples of the work that can be accomplished with these data, we describe and discuss the results generated from PISA 2000 and PISA 2003 in terms of international comparisons of achievement and the models of relational patterns of student, home, and school characteristics. We provide insight from the recent pilot testing conducted in Taiwan for PISA 2006, which has a focus on scientific literacy. This is followed by a discussion of the implications and potentials of the 2000 and 2003 datasets to facilitate research on scientific and mathematical literacy. The paper concludes with a look ahead to PISA 2006 and what researchers should be attending to in the research reports generated from the OECD and the research interests that they could follow given access to the datasets generated.

KEY WORDS: future opportunities, large-scale assessments, PISA results, secondary analyses

Literacy in science, mathematics, and reading has become a fundamental focus in public education. What this means in terms of student performance and achievement can vary from one definition or measure to another (Laugksch, 2000; Norris & Phillips, 2003), but the importance of literacy to future success in education and employment is constantly significant (DeBoer, 2000). Large-scale assessments of student achievement provide a window into the broadly defined concepts of literacy and generate information about levels and types of student achievement in relation to some of the correlates of learning, such as student background, attitudes, and perceptions, and perhaps school and home characteristics.

This paper provides an overview of one such assessment—the Programme for International Student Assessment (PISA).
Assessment programs such as PISA certainly attract public attention for the comparative listings of country mean performance. For example,

International Journal of Science and Mathematics Education (2007) 5: 591–614

there was considerable public and political debate generated in Germany in response to the publication of the PISA 2000 results (Fertig, 2003). However, as Fertig pointed out, "The publicly available background information collected in PISA 2000 together with the test results, family and individual characteristics and a rich set of school-related variables allows for a deeper analysis" (p. 3)—meaning deeper than the simple listing of average achievement scores that are reported in the public media. PISA was created by the Organization for Economic Cooperation and Development (OECD) in 1997 to evaluate the achievement of students nearing the end of their public school careers (15-year-old students). The achievement domains targeted are the literacies associated with reading, mathematics, and science primarily with some attention paid to problem solving. PISA uses the term *literacy* to encompass a broad range of competencies relevant to coping with adult life. These competencies were based on the relevance and applicability to adult life with no specific linkage to curricula of the participating countries. The assessment focuses on young people's ability to apply their knowledge and skills to real-life problems and situations. Along with the achievement estimates for students in the OECD countries and other nations participating in these studies, information is collected on student attitudes and perceptions related to schooling, home background variables, and school information. These rich datasets offer researchers the opportunity to investigate relationships amongst the correlates of learning and achievement and do so from an internationally comparative perspective.

PISA operates on a 3-year cycle in that all achievement domains are assessed in each cycle but one domain is the focus of the assessment for that cycle. In PISA 2000, reading performance was the focus of the assessment, whereas in PISA 2003, mathematics was the domain of focus, and PISA 2006 pays particular attention to science achievement. This cyclic approach is structured on the PISA framework of contemporary literacy and problem solving, and facilitates research and secondary data analysis to investigate a range of relationships and structures of student and school traits and student performance in science, mathematics, and reading literacies. The OECD and its 30 member countries administer the program, but a number of non-OECD countries choose to participate. A total of 32 countries participated in 2000, 41 countries in 2003, and 58 countries in 2006.

The basic sample units for PISA are countries, and within each country, students are randomly sampled within schools. There is a target of 5,000–10,000 students from at least 150 schools within each country (OECD 2003, p. 10). However, actual samples range widely from about 3,500 to over 30,000 students in some countries. The sampling procedure constrains analysis to country-level approaches and within-country school comparisons since it is not possible to conduct analysis at the classroom level (OECD, 2005b, p. 39). In order to generate unbiased estimates of country parameters, sampling weights were determined, and replicate weights are provided to allow for the calculation of unbiased standard error estimates of all country parameters.

The concept of literacy used by PISA is "... the capacity of students to apply knowledge and skills in key subject areas and to analyze, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations" (OECD, 2003, p. 13). The domains assessed by PISA are described as follows:

- *Mathematical literacy* - An individual' s capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments, 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.
- *Reading literacy* - An individual's capacity 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.
- *Scientific literacy* - The capacity to use scientific knowledge, to identify questions, and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity.
- *Problem-solving skills* - An individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the literacy domains or curricular areas that might be applicable are not within a single domain of mathematics, science, or reading.

(OECD, 2003, p. 14)

The test items used in PISA consist of an approximately even split of multiple-choice and constructed-response items. Sample items for the various domains and different levels of proficiency are provided in PISA documentation that is available for download at www.pisa.oecd.org. The tests consist of items and tasks that require 2 hours of student time to complete. Test booklets are derived from a larger pool of items (total testing time of 7 hours) that represent the domain of performance in the literacies of mathematics, science, and reading, which means that different students respond to different sets of items. The test each student completes consists of two-thirds of items for the domain of focus (reading for PISA 2000, mathematics for PISA 2003, science for PISA 2006). In order to provide estimates of student performance on a common metric, scoring and analysis is based upon item response theory (Hambleton & Swaminathan, 1985). The performance estimates are scaled to have an OECD mean of 500 with a standard deviation of 100. It should be noted that each student has performance estimates for all domains; for the so-called minor domains (e.g., all non-mathematics domains for PISA 2003), these estimates are based on small numbers of test items.

In addition to assessing student cognitive proficiency, PISA collects information about student attitudes, perceptions, and background. These student variables and indices include: gender, grades, and level of courses taken in school, home resource indices, home traits (parent's education and employment), self-regulated cognition, student motivation and engagement, emotional factors, and perceptions of school traits.

Information about schools (such as instructional processes and school organization) is provided by a questionnaire that school principals complete. Variables from the school questionnaire include: instructional practices, school organization, school decision-making, school resources, school climate, and teacher and school autonomy.

The design and development of the PISA datasets lead to analytic approaches that are very attentive to accurate estimates of student performance and of the magnitude of error associated with each parameter estimated. The datasets were specifically developed with replicate sampling weights to better estimate country parameters and plausible values for unbiased estimates of student performance estimates at the country level. These features, along with the inevitable missing data associated with multivariate analysis, routinely result in replicate analyses of over 400—and sometimes over 2,000—replications of computation within a given analysis. This cumbersome and at times daunting analysis requires a relatively high overhead to begin the engagement in analysis as well as substantial run-times that many analysts will not have experienced in contemporary computing. However, this analytic approach is well supported by a detailed technical manual and the statistical subroutines (macros) developed by OECD for both SAS and SPSS. These are freely available for download from the OECD/PISA website along with the data and reports (www.pisa.oecd.org).

The design of PISA has not been without criticism. Goldstein (2004), for example, pointed out the limitations of modeling achievement performance as a single dimension as is explicitly done in PISA by using a one-parameter item response model to develop test items and analyze student responses. This design element constrains analysis if the data are better modeled as multidimensional. Further, it is pointed out that in order to better investigate student performance and the correlates of achievement at the student, school, and home levels, it would be better to have a longitudinal dataset. However, the PISA project group chose to design the test items to fit the one-parameter item response model on the basis of theoretical considerations (OECD, 2005b) and to provide cross-sectional data with common anchor items from one cycle to the next. The use of anchored cross-sectional data allows for more feasible sampling of 15-year-old students in a varying number of participating countries and avoids problems associated with attrition over the 3 years between test administrations.

In this paper we describe and discuss the results generated from 2000 and 2003 in terms of international comparisons of achievement and the models of relational patterns of student, home, and school characteristics. This description will pay particular attention to the models developed for Australia and Canada. This is followed by a discussion of the implications and potentials of these datasets to facilitate research on scientific and mathematical literacy. We illustrate these potential analyses with data generated by the pilot testing conducted in Taiwan for PISA 2006. The paper concludes with a look ahead to PISA 2006 in which scientific literacy is the key focus and what researchers should be attending to in the research reports generated from the OECD and the research interests that they could follow given access to the data generated from PISA 2006.

PISA 2000 - Focus on Reading Literacy

Reading literacy was the focus domain for PISA 2000. Science literacy, mathematics literacy, and problem solving were also assessed, although the majority of questions focused on individuals' capacity to understand, use, and reflect on written texts that develop knowledge, personal potential, and informed participation in society. Student achievement was reported as overall reading literacy and in three content dimensions: interpreting texts, reflecting and evaluating, and retrieving information. Five levels of student proficiency are described for each domain, and students are classified within these levels based on how well they perform on each task (OECD, 2002, chapter 16). The *PISA 2000 Technical Manual* provided classification procedures with a summary of the distinguishing features of tasks at each level. For example, the content dimension retrieving information at proficiency level 3 (score range: 481 to 552; 60% of the OECD sample performed at this level or higher) requires that students “locate and in some cases recognize the links between pieces of information, each of which may be required to meet multiple criteria. Typically there is prominent competing information.” (OECD, 2002, p. 204). Released items, sample responses for each level of proficiency, and item difficulty are provided in the *Knowledge and Skills for Life* PISA report (OECD, 2001).

The publications produced by the OECD (2001, 2004) focus on the comparison of countries across overall reading literacy and for each of the three content dimensions. No comparison to earlier achievement was possible since this was the first cycle of the PISA program. Country rankings are given for overall reading literacy, mathematical literacy, and scientific literacy.

At the country level, there was a positive linear relationship (Table I) between average performance across the combined literacy score (reading, mathematics, science) and gross domestic product (GDP). As a country’s GDP increased, so did the combined literacy scores for that country (OECD, 2001, p. 91). Per capita GDP accounted for 27% of the variance in student performance on PISA 2000.

A similar trend was found for the amount of spending on education and average combined performance on the literacy scales that accounted for 17% of the variance in student performance (OECD, 2001, p. 91). Furthermore, at the school level, students attending high socioeconomic status schools had significantly higher composite reading literacy scores than did students attending low socioeconomic status schools (OECD, 2001, 2004). On average, the difference between the 75th and 25th quartiles on the school mean socioeconomic index was 0.72 of a standard deviation (OECD, 2001). At the individual level, students from more advantaged socioeconomic homes tended to perform better on all PISA 2000 measures (OECD, 2004). This trend was most notable in the German data, where a one standard deviation increase in socioeconomic

TABLE I

Correlations of OECD country mean performance reading, mathematics, and science (PISA 2000) to per capita GDP and education spending

Performance Domain	Correlation to GDP 2000	Correlation to Education Spending
Reading literacy	0.59	0.44
Mathematical literacy	0.55	0.47
Scientific literacy	0.39	0.29

status was associated with a 67-point increase in student performance compared to an average increase of 28 points in other OECD countries. However, the socioeconomic effects of home for Brazil, Finland, Greece, Iceland, Ireland, Korea, Norway, and Sweden were not statistically significant. The differential trends illustrate the nation-specific character of the correlates of learning and serve as a caution that a single model is unlikely to fit all countries.

The difference between reading literacy for males and females was statistically significant for all countries favoring females, although the spread in scores between the two genders varied with the largest difference reported in Finland (a 51-point difference) and the smallest in Korea (a 14-point difference) (OECD, 2001). On average, the difference in scores between males and females was one-half of a proficiency level (OECD, 2001). The magnitude of these differences varied across the content dimensions. The largest gender differences were on the reflection and evaluation dimension, where the average difference between male and female students was 45 points. This result is compared to an average difference of 29 points between genders on the interpreting texts dimension and 24 points on the retrieving information dimension (OECD, 2001). Gender differences were not as pronounced for mathematics literacy and science literacy. In mathematics literacy, about half of the countries tested showed a statistically significant difference, with male students outperforming female students. There was no consistent pattern of gender differences in science literacy performance (OECD, 2001).

The results from PISA 2000 for all participating countries revealed a positive relationship between mothers' education and student performance on reading literacy. Students whose mothers had completed upper secondary education or higher demonstrated reading literacy scores higher than those students whose mothers had not completed secondary education (OECD, 2001). The relationship between mother's education and student performance was most easily described by the *disadvantage* to students whose mothers had not completed secondary education. Mean scores were on average one-half of a standard deviation lower in reading, mathematics, and science literacies. The largest differences were seen in the German data, where students whose mothers had not completed secondary education scored on average 99 points lower.

Parental occupation also showed a significant relationship with student scores. Parental occupation was ranked by socioeconomic index of occupational status on a 90-point scale (OECD, 2001, p. 139) where higher values signify higher status. Students whose parents were in the top quarter of a country's average socioeconomic index averaged 45 points above the OECD average on reading literacy. The average gap between students whose parents ranked in the top quarter for their country and students whose parents ranked in the bottom quarter for their country was one full proficiency level on the composite reading literacy measure or just over one standard deviation in student score (OECD, 2001). However, not all countries showed identical patterns. In Korea, Finland, and Iceland, the differences between the two groups of students from families in the top and bottom quarters for their country were much smaller than average. In Belgium, Germany, and Switzerland, the gap was more than twice as large as the average for all participating countries. The socioeconomic index of occupational status accounted for 11% of the total variance in student performance on the composite reading literacy score, and these results were consistent for mathematical literacy and scientific literacy (OECD, 2001).

Other factors that influenced student performance in reading literacy were whether or not a student was native-born in the country in which he or she wrote the test. Ten of the countries assessed in PISA 2000 showed statistically significant differences in student performance on reading literacy, with native-born students outperforming first generation and foreign-born students. There was variation between countries; one notable exception was Ireland, where foreign-born students performed better than native-born students. This again illustrates the need for careful, jurisdiction-specific development and interpretation of student performance models.

PISA 2003 - Focus on Mathematics Literacy

Mathematical literacy was the domain of focus for PISA 2003. In addition to overall mathematical literacy, student mathematics performance was reported in four content dimensions: quantity, space and shape, change and relationships, and uncertainty. For each content dimension and the omnibus mathematics literacy, six levels of student proficiency are defined in relation to the kind of tasks students are able to successfully perform and the associated score range (OECD, 2005a, chapter 2). These content dimensions and their associated proficiency levels are described with sample items that clearly articulate the concept and definitions of mathematics literacy as operationalized by PISA. Sample items and their scale values (level of difficulty) are presented to illustrate the tasks related to the assessment of each domain. For

TABLE II

Correlations of OECD country mean performance in reading, mathematics, and science (PISA 2003)

Performance Domain	Science	Mathematics	Reading
Mathematics	0.97		
Reading	0.93	0.94	
Problem-solving	0.97	0.99	0.94

example, for the content domain of space and shape at proficiency level 4 (score range: 544 to 607 at which 30% of the OECD sample performed), students are described as being able to "solve problems that involve visual and spatial reasoning and argumentation in unfamiliar context; link and integrate different representations; carry out sequential processes; apply well-developed skills in spatial visualization and interpretation" (OECD, 2005a, p. 55).

A main thrust of the general report on PISA 2003 (OECD 2005a) is comparison of countries across the overall mathematics means and for each of the four content domains. The mathematic performance is largely unchanged between 2000 and 2003 although a small number of countries showed some change. At the country level, there is also stability in terms of performance across the literacy domains. When the country means from PISA 2003 are correlated (Table II), we see that there is a close association ($r > 0.9$) between all domains—science, mathematics, reading, and problem solving. This means that, if a country scores high in mathematics, it will also have high scores in science, reading, and problem solving; and, if a country is at OECD average in mathematics, it will likely be average in the other domains as well. These strong positive relationships hold at the student level albeit with a somewhat reduced strength of relationship (Table III).

TABLE III

Correlations of OECD student means in reading, mathematics, and science[a] (PISA 2003)

Performance Domain	Science	Mathematics	Reading
Mathematics	0.82		
Reading	0.82	0.78	
Problem-solving	0.82	0.87	0.82

[a] These correlations were calculated using all participating students' plausible values.

TABLE IV

Correlations of OECD country mean performance in reading, mathematics, and science (PISA 2003) and per capita GDP

Performance Domain	Correlation to GDP 2003
Science	0.66
Mathematics	0.72
Reading	0.76
Problem-solving	0.70

At the country level, there was also a moderately strong positive relationship between economic wealth as indexed by per capita GDP for 2003[1] (Central Intelligence Agency, 2003) and mean performance in all domains (Table IV). The wealth of a nation as indexed by GDP accounts for approximately half of the variance in country-level performance in all domains of performance with the exception of science, which accounts for about 44% of variance. At the individual student level, socioeconomic variables were important correlates of the mathematics performance of the 15-year-old students who completed the PISA tests. The strength of the relationships varied across countries and across the four content domains of mathematics.

In mathematics, males generally outperformed females in most countries, but it would be incorrect to claim this is a universal relationship. A clear exception to this is in Iceland where females scored on average 18 points higher than males. In a number of countries, the difference between males and females, although favoring males, is not statistically significant. In addition, the magnitude of difference varies across the four content dimensions. Change and relationship, for example, had lower levels of male–female difference than did space and shape; and quantity had the lowest level of difference between males and females generally across countries. In a recent paper, the interactions between student gender, content dimension, item formats, and task demands were explored (LaFontaine & Monseur, 2006); the female advantage in open-ended constructed responses and the male advantage in terms of selection item formats were supported. Further, in their analysis of reading results from PISA 2000, LaFontaine and Monseur found that the female advantage is greatest for reflective tasks and least for tasks involving noncontinuous texts.

PISA 2003 results clearly show the positive relationship of parental education and vocational status to student performance in mathematics

(OECD, 2005a). On average, there is a 50-point gap between students whose mothers have completed upper secondary school and those whose mothers have not, and a 93-point gap between those students whose parents are in the highest vocational level and those in the lowest (OECD, 2005a, p. 165). But there is a clear variation in these relationships across countries. Some countries, such as Finland and Canada, have both high levels of performance and equity (in the sense that the performance gaps on these variables were relatively low). Performance gaps were also found between students who were native born and those who were not (students were asked if they were born in the country in which they were taking the test), which favored the native born. Again, there were substantial differences between countries. For example, there were very modest gaps in some countries (such as Australia, Canada, and the United Kingdom) and large gaps in other countries (such as Belgium, Germany, the Netherlands, Switzerland, and Sweden) (OECD, 2005a, p. 168).

The data also allow for the investigation of relationships between school characteristics and performance. After modeling the data for all OECD combined, it was found that school disciplinary climate, student–teacher relations, and academic record utilization were relatively strongly related to student performance even when conditioned on student gender, origin, and language and both student and school level socioeconomic status. There was also substantial variation in these relationships across countries. These relationships have significance for educators and policy makers, and the PISA datasets provide opportunities to explore them within each participating country.

A generally overlooked yet useful finding from this international survey is not only the extent to which countries vary in terms of average student performance but also the nature of the variance in performance measures that is specifically associated with schools as opposed to individual students. On average, schools accounted for 34% of the variance in mathematics performance (the intra-class correlation), but this varied widely across countries. Schools in Iceland, for example, accounted for about 4% of variance whereas schools in Belgium and the Netherlands accounted for about 60% of the variance in mathematics performance. This means, of course, that when developing models of the correlates of learning, each country would have to be modeled separately—a 'grand PISA model' would not reflect the nature of the relationships. Also, for those countries with low intra-class correlations, models that used schools as the level-2 units did not have much variance to model; those models are constrained and should be interpreted differently than models from countries in which a high proportion of mathematics performance variation is accounted for by schools. Further, the underlying structural and policy differences across countries that may have led to this discrepancy in school-level variance offer another potentially significant line of inquiry in addition to consideration of mean differences in mathematics performance between countries.

Work continues on the analyses within the CRYSTAL Pacific Project at the University of Victoria with a focus on hierarchical linear modeling (HLM) of mathematics performance with student, home, and school correlates. A recent study (Gu, 2006) showed that, although student self-concept is positively related to mathematics performance for two high-performing countries (Canada and Hong Kong), the mean student self-concept is much lower for Hong Kong students than for Canadian students. This occurs even though Hong Kong has a higher overall mean performance in mathematics. Another study (Goh, 2006) that modeled the relationships of student intrinsic motivation, teacher support, student-teacher relations, and mathematics performance for Canadian students demonstrated that motivation is positively related to achievement. If student perceptions of teacher support are added to the model, the level of teacher support is negatively related to mathematics achievement. Although this finding may appear counterintuitive, it is consistent with some other research that indicates parental support, such as assistance with homework, is negatively related to student achievement (Anderson, Rogers, klinger Ungerleider, Glickman & Anderson, 2006; Ho & Willms, 1996). We have speculated that in reporting levels of perceived support those students achieving at lower levels may require more interaction with supportive adults and so report more frequent experiences of support. Current investigation is centered on the modeling of student motivational correlates in association with other student and parental traits.

PISA 2006 - Focus on Science Literacy

PISA 2006 emphasized science literacy while continuing to collect data on mathematical literacy, reading literacy, and problem solving. The same general test design, sampling, and statistical analysis were used for PISA 2006 as were used for 2000 and 2003. However, at the time of writing this article, the actual 2006 data and results were not available so the authors relied on the field trial test and results from Taiwan (described in the next section of this article) for much of their insights and suggestions. The field test items of science were designed to assess student abilities of applying scientific knowledge in problem-solving settings rather than the memorization of factual knowledge. These items have a greater emphasis on the application of scientific knowledge in the context of life situations than traditional school science tests and most standardized examinations. In other words, the inquiry abilities of making hypotheses, defining researchable questions, and drawing evidence-based conclusions were emphasized in the development of test items. Significantly, this emphasis is consistent with the major trend of national curricular reform in many countries, e.g., the National Statement on Science for Australian Schools 1994, the National Curriculum Orders in the United Kingdom 1994, the National Science Education Standards in the United States 1996, and The Pan-Canadian Framework for Science 1997. For example, the U.S. National Research Council (NRC, 1996) indicated that:

Students should develop sophistication in their abilities and understanding of scientific inquiry. (p. 173)

Students should formulate a testable hypothesis and the design of an experiment. (p. 174)

These inquiry standards and related habits of mind imply that science learning should be based in a constructivist context where students integrate prior knowledge, concurrent experiences, and canonical knowledge to construct understandings of the nature of science, scientific inquiry, and the big ideas of science. Precisely what this classroom climate and context might be is not completely described in the reform documents.

In the study conducted in Taiwan, the focus was on field testing instrumentation. The participants in this study attended 40 schools randomly selected from the 477 schools in Taiwan identified by the PISA 2006 consortium. Once the schools were identified, 40 students were randomly selected from each school resulting in a sample of 1,434 15-year-old students. A feature of this pilot study was the enhanced attention paid to students' perceptions of their learning environment in addition to their science literacy.

The importance and influence of the classroom learning environment on the process of education has been recognized by researchers (Aldridge, Fraser & Huang, 1999; Alridge, Laugksch, Seopa & Fraser, 2006; Fraser, 1998; 2002; Goh & Khine, 2002). A number of learning environment questionnaires have been carefully validated and widely used including: the Learning Environment Inventory (LEI) to assess the environment of high school settings (Walberg, 1979), Questionnaire on Teacher Interaction (QTI) for assessing students' perceptions of their teachers' interpersonal behavior (Wubbels & Levy, 1993), Constructivist Learning Environment Survey (CLES) to measure the extent to which constructivist approaches are being adopted in classrooms (Taylor, Fraser & Fisher, 1997), Science

Laboratory Environment Inventory (SLEI) for the assessment of science laboratory learning environment (Fraser, Giddings & McRobbie, 1995), and What Is Happening In this Class (WIHIC) (Fraser & McRobbie, 1995).

WIHIC was used for the Taiwan PISA study because it had been used to assess student perceptions of learning environment in a number of different subject areas, at a range of grade levels, and in more than nine countries (Aldridge et al., 2006). In addition, it had been translated into Chinese, back-translated into English, and checked for accuracy (Aldridge et al., 1999). More importantly, WIHIC has an inquiry subscale, which is one of the emphases to be explored by the PISA 2006 field trial test and is the research focus of the field trial study of the relationship between Taiwanese students' perceptions of learning environment and their science literacy performance. The 56-item Chinese version of the WIHIC questionnaire (Aldridge et al., 1999) designed to measure students' perceptions of their classroom environment has seven 8-item subscales: cohesiveness, teacher support, involvement, inquiry, task orientation, cooperation, and equity. Each item contains a statement (e.g., I carry out investigations to test my idea) and is followed by a five-point Likert-type scale of (5) *always*, (4) *often*, (3) *sometimes*, (2) *seldom*, and (1) *never*.

For measuring performance in science literacy, all participants were randomly assigned 1 of the 12 trial test booklets designed to assess science literacy and containing 51–57 test items. The test items can be classified into three types: (1) science items, which ask students to identify questions, explain scientific phenomena, and draw evidence-based conclusions about contextual scientific problems; (2) level of interest items, which ask students to express their interest level toward scientific topics; and (3) degree of support items, which solicit students' opinions of science-related issues. The science items included multiple-choice and open-ended questions. The level of interest items and the degree of support items are all in Likert-scale form. Participants' responses were coded based on the coding guide and scoring rubrics developed by the PISA 2006 consortium.

The data analyses revealed that at the school level student performance in science, level of interest, degree of support, and general learning environment (total WIHIC), and learning environment of inquiry (inquiry subscale of WIHIC) were significantly correlated ($p < 0.001$). The correlations are shown in Table V.

Further analyses revealed that male students significantly outperformed female counterparts in scientific literacy, expressed higher level of interest, and perceived higher degree of support. However, there were no significant ($p > 0.05$) differences between male and female students on

TABLE V

Correlations of students' scientific literacy and learning environment

Learning Environment	Science	Interest	Support	WIHIC
Interest	0.75			
Support	0.69	0.76		
WIHIC	0.65	0.72	0.50	
Inquiry	0.66	0.72	0.56	0.87

their performances of reading and mathematics. The independent *t*-test results are shown in Table VI.

These initial findings from the PISA 2006 trial results in Taiwan have important implications for science teaching in that a learning environment of inquiry is significantly correlated with students' scientific literacy, interest, and perceived support. These results support the curricular reforms that emphasize the importance of inquiry since increased levels of student-perceived use of inquiry approaches are associated with improved PISA performance. Despite the emphasis on inquiry for students and the encouragement of inquiry-based teaching in the science education reform documents, authors and publishers of textbooks tend to promote the traditional cookbook approaches to experimentation in which limited opportunities for open investigation are provided. Friedler & Tamir (1986) indicated that students are rarely required to identify problems, formulate hypotheses, design experimental procedures, and work according to their own designs. Science teachers and science education researchers may not

TABLE VI

Independent *t*-test results of student performance between male and female

Variable	Gender	Mean	*SD*	*t*	*P*
Science	Female	58.64	15.89	3.12	0.002
	Male	61.25	15.62		
Reading	Female	62.54	18.03	1.10	0.274
	Male	58.64	21.10		
Mathematics	Female	65.99	20.95	1.54	0.127
	Male	71.30	16.46		
Interest	Female	72.26	13.27	2.18	0.030
	Male	73.90	14.99		
Support	Female	81.89	9.22	3.07	0.002
	Male	83.52	9.97		

be sufficiently aware of the importance of assessing students' scientific process skills in the area of 'designing experimental procedures' and 'making hypotheses' as part of both the fundamental and derived senses of science literacy (Norris & Phillips, 2003). If the National Science Education Standards (NRC, 1996) represent consensus on the international reform goals for science teaching, science teachers should pay more attention to the inquiry ability of students that have long been ignored. The PISA results from the Taiwanese study support the view that lock-step, directive instructional approaches do not yield improved achievement in science literacy.

IMPLICATIONS FOR RESEARCH ON SCIENCE LITERACY AND MATHEMATICS LITERACY

Many of the results from the PISA project have been widely published in the national and international press, and the findings have received varied responses. For example, Donelly (2005) claimed that: "Based on PISA, it appears that education in Australia is on the right track and that standards are high. A closer view of the PISA test shows that the opposite is the case as the test, when compared to TIMSS, is substandard and flawed."

This view tends to be supported by the Australian federal government, which is concerned that current school curricula, not only in science and mathematics but other subjects as well, are not fulfilling the nation's needs and subsequently has called for a back-to-basics in classrooms (Ferrari, 2006). From the 2003 PISA data with an emphasis on mathematical literacy and the 2000 data with an emphasis on reading literacy, a large number of findings have been reported, several of which have been brought together as separate reports. Of particular interest have been results on immigrant students (OECD, 2006a), Australian indigenous students (De Bortoli & Cresswell, 2004), and socioeconomic issues (OECD, 2006b).

Two groups of immigrant students were identified in the analyses: first generation students who were born outside the country of assessment and whose parents were also born in a different country, and second generation students who themselves were born in the country of assessment but whose parents were born in a different country. The latter students completed all their schooling in the country of assessment. The PISA report also compared immigrant students to native students born in the country of assessment and who had at least one parent born in that country. The analyses involved 17 OECD countries that had significant immigrant populations: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Luxembourg, The Netherlands, New Zealand, Norway, Sweden, Switzerland, and the United States as well as partner countries Hong-Kong-China, Macao-China, and the Russian Federation (OECD, 2006a, p. 7).

The data from PISA 2003 showed that immigrant students are motivated learners and have positive attitudes towards school. But as stated in the foreword, "Despite these strong learning dispositions, immigrant students often perform at significantly lower levels than their native peers in key school subjects, such as mathematics, reading and science as well as in general problem solving skills" (OECD, 2006a, p. 3). However, there were considerable differences between participating countries. There was little difference between performance of immigrant and native students in Australia, Canada, and New Zealand—traditional settlement countries where immigrant and native students attend a school with similar resources and climates—as well as Macao-China. However, the performance of immigrant and native students was most pronounced in Austria, Belgium, Denmark, France, Germany, The Netherlands, and Switzerland where immigrant students often attend schools with relatively disadvantaged native student populations in terms of economic, social, and cultural backgrounds. A major issue arising from this report is that in the majority of countries at least one in four immigrant students did not demonstrate basic mathematics skills as defined by the PISA assessment, giving rise to future challenges of employment.

Other data from PISA 2003 revealed that high levels of immigration do not necessarily impair integration and that schools can build on the strong learning dispositions of immigrants. The issue of language support in schools is of particular importance where immigrant students do not speak the language of instruction at home. Although the language spoken at home does not fully account for variations in immigrant students' relative performance levels, immigrant students who do not speak the language of instruction at home tend to be lower performing in mathematics in several countries, even accounting for parents' educational and occupational status.

On the available information from PISA 2003, it is not possible to determine the extent to which the different language support programs contribute to relative achievement levels of immigrant students. However, in countries with relatively small achievement gaps between immigrant and native students or smaller gaps for second-generation students compared to first generation students, long-standing language support programs exist with relatively clearly defined goals and standards. These countries include Australia, Canada, and Sweden (OECD, 2005a, p. 167). In a few countries where immigrant students perform at significantly lower levels, language support tends to be less systematic.

The Australian PISA 2000 data sampled approximately 500 Australian indigenous students, and comparisons were made with other Australian students and other countries' students (De Bortoli & Cresswell 2004). In reading literacy, mathematical literacy, and scientific literacy, Australian indigenous students performed at a lower level than nonindigenous students and were below the OECD mean. Female Australian indigenous students outperformed males in reading literacy, but there were no significant differences between Australian indigenous male and female students in mathematical literacy and scientific literacy (Lokan, Greenwood & Cresswell, 2001). The majority of indigenous students were in either year 9 or 10, in a general academic program (62%), and in a program with general and academic subjects (19%). An equal but smaller percentage of year 11 indigenous students were in programs aiming to attend university (7%), leading to apprenticeships (7%), and allowing direct job entry from school (5%).

Taking into account the influence of home background factors on performance, resources (such as books) in the home were fewer in homes of indigenous students compared to nonindigenous students. Indigenous students reported spending less time on homework, reported fewer disciplinary problems in their English lessons, and had a more positive sense of belonging at school compared with nonindigenous students (Lokan et al., 2001; Zubrick , Silburn, Gurrin, Teoh, Shepherd, Arlton & Lawrence, 1997).

Indigenous students had less preference for a competitive learning environment and a greater preference, especially females, for a cooperative style of learning. Indigenous students were less likely to use elaboration strategies that relate to the degree to which students apply knowledge in new situations and executive control strategies that relate to monitoring their own learning. However, both indigenous and nonindigenous students reported using elaboration strategies more often than the OECD average, with indigenous females using these strategies less than male indigenous students.

One of the issues to arise from the PISA data requiring further exploration is how the structure of schooling influences the quality and equity of educational outcomes (OECD, 2006b). Aspects of interest include the grouping of students, segregation of schools, management and financing, school resources, and the instructional climates. The results of the PISA data show that the school that students attend is strongly predictive of their performance and that this is most evident in the school's socioeconomic composition. As might be anticipated, there is evidence that schools with a more advantageous intake often have better educational resources. A positive school climate, in particular a strong disciplinary climate, is associated with better student performance (Doig, 2001).

Comparing the private school and public/state school education systems revealed that the mean performance in the private education systems was on average higher than in public/state systems although this appears to be largely due to an advantage student intake (OECD, 2006b, p. 91). However, the PISA data do not provide any evidence that public/state systems are more equitable in terms of the total variation in student performance. In many of the participating countries, there is some degree of school autonomy with regards to school policies, financial resources, curriculum, and instruction but not personnel management, although there is more responsibility for this in private schools, which in turn is associated with better school performance.

FUTURE RESEARCH

At the University of Victoria, we are continuing with our investigations of the PISA datasets with the development of multilevel models to further our understanding of systematic relationships between student traits and literacy and the extent to which school characteristics are related to these student level gradients. The work is planned to extend into the data generated by The International Mathematics and Science Study (TIMSS). This work is being conducted under the auspices of the CRYSTAL Pacific Project that is funded by the Natural Sciences and Engineering Research Council of Canada. However, there are a number of significant areas for future research using PISA data.

Socioeconomic status (SES) of the student plays an important role in level of achievement attained. However, a number of factors are used to construct the SES indices used in PISA and in many other studies of educational performance. The effects of schools on student performance vary substantially across countries, and it is likely that the effects of SES variables also influence student performance in different ways and to different extents from one country to another. It appears as if in some countries these SES factors are confounded with ethnicity, nationality, or immigrant status. Unpacking the effects of the individual variables that are used to compose the SES index (e.g., parental educational levels, parental occupational levels, household income, household possessions of varying sorts, etc.)—and doing so comparatively across countries—could lead to better understanding of how these variables correlate to learning and literacy.

The strong correlations between student performances in mathematics, science, and reading literacies (Table III) support the concepts of interacting fundamental and derived senses of science literacy (Norris & Phillips, 2003) and mathematics literacy (Pimm, Tuan & Yore, 2007). Yore & Treagust (2006) speculated that there is a potential cognitive symbiosis between the fundamental and derived senses of scientific literacy that may well exist for mathematics literacy. The measures of science, mathematics, and reading literacies that have been developed for PISA may be more closely aligned to these formulations of literacy than measures associated with other large-scale assessment tests (e.g., Iowa Test of Basic Skills) and programs (e.g., TIMSS). This linkage of PISA data to the aforementioned three literacies offers research opportunities for the investigation of how these vary across countries and how a wide spectrum of learning correlates that are available in the PISA datasets are related to overall student literacy performance.

Another avenue for significant research activity with the PISA data is centered on the differential performances associated with students who have different language and perhaps cultural characteristics than majority students in the participating countries. This would not only be related to the language used in everyday discourse but also to the language associated with schooling—the discourse of instruction and the discourses of science and mathematics. Nonnative students then are confronted with multiple language barriers: the transition from home language to the language of instruction in their new country and then the transition from the instructional discourse to the target disciplinary discourse (Chinn, Hand & Yore, 2007; Gee, 2004; Yore & Treagust, 2006; Yore, 2007). Difficulties of this nature could be further exacerbated for indigenous students who have grown up in homes in which both the language and the culture are substantially different from the mainstream population of the country of testing. It may be that not only the language of discourse creates some impediments to achievement but deeper issues associated with how world views, epistemology, and ontology influence the understanding of the topics of science and mathematics. The PISA data do allow for the segregation of students into language groups which opens some opportunity to investigate this area both within and across countries.

These and many other issues of instructional significance can be investigated through the data provided by PISA. The data, descriptions of the instrumentation, manuals for analysis, and reports of the initial round of findings are freely available for download on the OECD website. We think this offers a solid opportunity for meaningful national and international research and exploration for science and mathematics educators and researchers.

NOTE

[1] Luxembourg was not included in these data since its GDP is deemed to be an outlier at over 3SD above OECD mean

REFERENCES

Aldridge, J.M., Fraser, B.J. & Huang, T.I. (1999). Investigating classroom environments in Taiwan and Australia with multiple research methods. *Journal of Educational Research, 93*, 48–61.

Aldridge, J.M., Laugksch, R.C., Seopa, M.A. & Fraser, B.J. (2006). Development and validation of an instrument to monitor the implementation of outcomes-based learning environments in science classrooms in South Africa. *International Journal Science Education, 28*(1), 45–70.

Anderson, J.O., Rogers, W.T., Klinger, D.A., Ungerleider, C., Glickman, V. & Anderson, B. (2006). Student and school correlates of mathematics achievement: Models of school performance based on Pan-Canadian student assessment. *Canadian Journal of Education, 29*(3), 706–730.

Central Intelligence Agency. (2003). *The world factbook*. Retrieved April 2006 from http://www.cia.gov/cia/publications/factbook/index.html.

Chinn, P.W., Hand, B.M., & Yore, L.D. (2007). Culture, language, knowledge about nature and naturally occurring events, and science literacy for all: She says, he says, they say. *L1-Educational Studies in Language and Literature* (special issue, submitted).

De Bortoli, L. & Cresswell, J. (2004). *Australia's indigenous students in PISA 2000: Results from an international study*. ACER Research Monograph No 59. Camberwell, Victoria: Australian Council for Educational Research.

DeBoer, G.E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*(6), 582–601.

Doig, B. (2001). *Summing up: Australian numeracy performances, practices, programs and possibilities.* Camberwell, Victoria: Australian Council on Educational Research.

Donelly, K. (2005). PISA and TIMSS are like apples and oranges. *ON-LINE opinion* - Australia's e-journal of social and political debate (posted 1 February 2005) http://www.onlineopinion.com.au/view.asp?article=2985.

Ferrari, J. (2006). National syllabus ' a power grab', *The Weekend Australian*, October 7–8, p. 4.

Fertig, M. (2003). Who's to blame? The determinants of German students' achievement in the PISA 2000 study. Discussion paper 739, The Institute for the Study of Labor (IZA), Bonn, Germany. http://ftp.iza.org/dp739.pdf. Retrieved April 2007.

Fraser, B.J. (1998). Science learning environments: Assessment, efforts and determinants. In B.J. Fraser & K.G. Tobin (Eds.), *The international handbook of science education* (pp. 527–564). Dordrecht, The Netherlands: Kluwer.

Fraser, B.J. (2002). Learning environment research: Yesterday, today and tomorrow. In S.C. Goh & M.S. Khine (Eds.), *Studies in educational learning environments: An international perspective* (pp. 1–27). Singapore: World Scientific.

Fraser, B.J. & McRobbie, C.J. (1995). Science laboratory classroom environments at schools and universities: A cross-national study. *Educational Research and Evaluation, 1*(4), 289–317.

Fraser, B.J., Giddings, G.J. & McRobbie, C.J. (1995). Evolution and validation of a personal form of an instrument for assessing science laboratory classroom environments. *Journal of Research in Science Teaching, 32*, 399–422.

Friedler, Y. & Tamir, P.J. (1986). Teaching basic concepts of scientific research to high school students. *Journal of Biological Education, 20*(4), 263–269.

Gee, J.P. (2004). Language in the science classroom: Academic social languages as the heart of school-based literacy. In E.W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives on theory and practices* (pp. 13–32). Newark, DE: International Reading Association.

Goh, M. (2006). *A multilevel analysis of mathematics literacy: The effects of intrinsic motivation, teacher support and student-teacher relations*. Master's Thesis, University of Victoria, Victoria, British Columbia, Canada.

Goh, S.C., & Khine, S.M. (Eds.). (2002). *Studies in educational learning environments: An international perspective.* Singapore: World Scientific.

Goldstein, H. (2004). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education, 11*(3), 319–330.

Gu, Z. (2006). *A comparison of Canada and Hong Kong-China through hierarchical linear models: The relations among students' self-beliefs in math, the learning environment at school, and math performance.* Master's Thesis, University of Victoria, Victoria, British Columbia, Canada.

Hambleton, R.H. & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Norwell, MA: Kluwer.

Ho, E.S. & Willms, J.D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education, 69*, 126–141.

LaFontaine, D. & Monseur, C. (2006). *Impact of test characteristics on gender equity indicators in the assessment of reading comprehension.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, USA.

Laugksch, R.C. (2000). Scientific literacy: A conceptual overview. *Science Education, 84*(1), 71–94.

Lokan, J., Greenwood, L. & Cresswell, J. (2001). *15-Up and counting, reading, writing, reasoning: How literate are Australian students? The PISA 2000 survey of students' reading, mathematics and scientific literacy skills. Report for the OECD.* Camberwell, Victoria: Australian Council for Educational Research.

National Research Council. (1996). *National science education standards.* Washington, DC: Author.

Norris, S.P. & Phillips, L.M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224–240.

Organisation for Economic Co-operation and Development. (2001). *Knowledge and skills for life: First report from the OECD programme for international student assessment.* Paris: Author.

Organisation for Economic Co-operation and Development. (2002). *PISA 2000 technical report.* Paris: Author.

Organisation for Economic Co-operation and Development. (2003). *The PISA 2003 assessment framework - mathematics, reading, science and problem solving: Knowledge and skills*. Paris: Author.

Organisation for Economic Co-operation and Development. (2004). *Messages from PISA 2000.* Paris: Author.

Organisation for Economic Co-operation and Development. (2005a). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

Organisation for Economic Co-operation and Development. (2005b). *PISA 2003 data analysis manual - SAS users*. Paris: Author.

Organisation for Economic Co-operation and Development. (2006a). *Where immigrant students succeed - A comparative review of performance and engagement in PISA 2003*. Paris: Author.

Organisation for Economic Co-operation and Development. (2006b). *School factors related to quality and equity: Results from PISA 2003*. Paris: Author.

Pimm, D., Tuan, H-L. & Yore, L.D. (2007). Language, mathematics literacy, and science literacy: Seeking convergence and cognitive symbiosis. *International Journal of Science and Mathematics Education* (special issue, in press).

Taylor, P.C., Fraser, B.J. & Fisher, D.L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research, 27*, 293–302.

Walberg, H.J. (Ed.), (1979). *Educational environments and effects: Evaluation, policy, and productivity.* Berkeley, CA: McCutchan.

Wubbels, T. & Levy, J. (Eds.) (1993). *Do you know what you look like?: Interpersonal relationships in education.* London: Falmer Press.

Yore, L.D. (2007). Science literacy for all students: Language, culture, and knowledge about nature and naturally occurring events. *L1-Educational Studies in Language and Literature* (special issue, submitted).

Yore, L.D. & Treagust, D. (2006). Current realities and future possibilities: Language and science literacy-empowering research and informing instruction. *International Journal of Science Education, 28*(2–3), 291–314.

Zubrick, S.R., Silburn, S.R., Gurrin, L., Teoh, H., Shepherd, C., Carlton, J. & Lawrence, D. (1997). *Western Australian child health survey: Education, health and competence.* Perth, Western Australia: Australian Bureau of Statistics and the TVW Telethon Institute of Child Health Research.

JOHN O. ANDERSON is a Professor in the Department of Educational Psychology and Leadership Studies at the University of Victoria who works in the area of educational measurement.

HUANN-SHYANG LIN is President of the National Hualien University of Education who works in science education and currently serves as the National Program Manager of PISA 2006.

DAVID F. TREAGUST is a Professor of Science Education in the Science and Mathematics Education Centre at Curtin University of Technology in Perth whose research interests are in science learning and conceptual change.

SHELLEY P. ROSS is a doctoral candidate in the Department of Educational Psychology and Leadership Studies at the University of Victoria who works in measurement and statistical modeling.

LARRY D. YORE is a Distinguished Professor and Chair of the Department of Curriculum and Instruction at the University of Victoria who works in science education with special emphasis on science literacy.

Department of Educational Psychology & Leadership Studies, Faculty of Education
University of Victoria
Victoria, BC V8W 3N4, Canada
E-mail: anderson@uvic.ca