Measuring Learner Outcomes

Eugene H. Owen
National Center for Education Statistics
U.S. Department of Education

Jay Moskowitz
American Institutes for Research

Maria Stephens
American Institutes for Research

This article describes the work to develop information on learner outcomes being undertaken through the Organization for Economic Cooperation and Development's (OECD) International Indicators of Education Systems (INES) Project. It first sets the context for this work, describing the increasing interest in many countries in comparative assessments and the use of performance data to inform education policy. The article then describes early development work on learner outcomes, including the establishment of technical quality standards for the presentation of data; development of a data strategy to collect regular and policy-relevant information on learner outcomes; and recent work to explore new areas for measurement and to expand the range of indicators on learner outcomes. In doing so, it gives particular attention to the development of the Programme for International Student Assessment (PISA) and details initial findings on students' overall performance in reading literacy and on the variation between and within schools in reading literacy.

Countries have long collected important information on students' access to education, their progression through school, their completion rates, and the human and financial resources used in education. In past years, however, there has been an increasing trend to supplement basic education data with information on students' *actual learning* in various

subject areas and also with information on those individuals who have left school and comprise the workforce (UNESCO, 1990). The importance of assessment of students and adults has grown at both the national and international levels (Owen et al., 2003, *pending*; Ministry of National Education, 1994). In particular, evaluations of performance that focus on system-level results—on the overall product of a school's, jurisdiction's, region's, or country's education and training system—have become important for policy makers and planners. These include national assessments such as the United States' National Assessment of Educational Progress (NAEP) and Canada's School Achievement Indicators Program (SAIP) as well as international assessments, which allow countries to examine their results in the broadest possible context.

Results of such outcomes assessments in the United States, Canada, and around the world are used to inform policies, benchmark achievement across countries (or regions or jurisdictions, in the case of national assessments), identify relative strengths and weaknesses in education systems, identify factors for improving achievement, assist teachers' efforts to raise student achievement, inform curriculum development, and increase public awareness and stimulate political debate (Owen et al., 2003, *pending*).

This article describes the role of the International Indicators of Education Systems (INES) Project's Network A in supporting international assessment and the development and presentation of information on learner outcomes. It describes how the concept of outcomes has broadened over the life of the project and some of the technical and measurement issues that have arisen. Initially focused on presenting information on students' achievement in school, in wrestling with technical issues such as how to measure students' opportunity to learn the subject matter being assessed and make meaning of results, and in trying to develop indicators that would be policy relevant and meet high technical standards, the network's focus shifted in a way that can be characterized as a broadening from examining *education* outcomes, i.e., what have students learned in school, to examining *learner* outcomes, i.e., what knowledge and skills do students and adults possess.

Twenty-eight countries are voluntary members of Network A, and the United States holds the Chair.¹ Like the other networks in INES, Network A is charged with developing information to support the international system of indicators that has come to be important to

many participating countries. While Networks B and C focus, respectively, on labor market outcomes and on schools and school processes, Network A focuses on developing indicators of learner outcomes.

Past Work on the Development of Outcomes Indicators

A major focus for Network A in the early years of the INES project was on establishing standards for indicator work related to learner outcomes. These quality standards (unpublished) addressed issues such as comparability of instruments, procedures and samples, validity, and reliability, and set minimum acceptable criteria for the inclusion of information on learner outcomes in *Education at a Glance* (EAG). For example, the network recommended that country mean scores of achievement for entire populations or selected groups should have a reliability score of 0.80 or higher and that attitudinal scales should have at least a 0.70 reliability (where the reliability score is the ratio of true score variance to total observed score variance). The matter of response rates and population exclusions received similar attention. The network recommended a minimum requirement of 85 percent for school response rate and 90 percent for student response rate and determined that results should not be reported for countries where more than 10 percent of the population had been excluded from the assessment. It also recommended that indicators provide information on the content validity of assessments from which data were drawn and on rationales and specifications of scaling procedures. Although such standards for reliability, validity, and comparability, and for subsequent reporting seem more ubiquitous now, they were not always so at that point in time, when large-scale assessment was still developing, albeit rapidly in the 1990s. These key technical issues under discussion influenced many later decisions related to a data strategy for outcomes indicators (see next section).

The establishment of standards then led to the development of templates for specific indicators related to student achievement (e.g., comparative distribution of scores). The indicators in the earliest editions of EAG presented mainly mean scores, multiple comparisons of mean scores, and student differences, both between sexes and among students at various percentiles of performance within countries for reading, mathematics, and science. Although Network A was interested in more in-depth indicators, such as on the sources of variance in achievement, it was not until the 1998 edition that such an indicator appeared. Up until that point, the network had determined the

indicator to be too crude – finding the sources of data for constructing the indicator incomplete or not comparable given the methods of calculation that had been used, and believing it to be too complex for readers and requiring a level of quality control that could not be provided (Grisay, 1995).

In addition to work to develop standards and improve the quality of presentations of existing information, the network began to think about expanding the information available on outcomes and the context for outcomes. For instance, in response to growing activities in many countries to have broad and specific educational goals related to educational reform and outcomes, the network's Goals Orientation and Attainment in Learning Systems (GOALS) project conducted a feasibility study and consequently administered a survey and analyzed and reported results on this topic (Granheim & Pettersson, 1995; Knutsen, 1995). Undertaken from 1994 to the completion of the final report (OECD, 1997a), this work was important in that it helped to further an understanding of education systems in their entirety and to provide a context in which to assess and understand outcomes (i.e., in relation to stated goals). A key finding of this work was the importance many countries placed on non-academic outcomes.

In keeping with this finding, the network expressed an interest in developing instruments to measure cross-curricular competencies (CCCs), which refer to those competencies that are learned and applied across curricular contexts and that are thought to be important for individuals' successful functioning in society. A feasibility study was conducted in 1994-95, described in the resulting publication *Prepared for Life* (OECD, 1997c) and other sources (Trier & Peschar, 1995). The study showed that it would be possible to develop indicators on CCCs and several of the network's developmental activities since then have focused on areas identified for short-, medium-, and long-term development. It also sparked a parallel activity, outside the network, that focused on exploring the conceptual underpinnings for competencies.

Parallel to these activities and in accordance with the priority to establish international benchmarks for outcomes expressed by countries at the third INES General Assembly in Lahti, the network began to think more broadly about how the collection of information on learner outcomes could become part of OECD's mainstream data collection efforts. Throughout the mid-1990s, Network A countries developed a long-term data strategy for the regular collection of policy-relevant

information on students' performance. This was the beginning of PISA, the Programme for International Student Assessment.²

PISA – Programme for International Student Assessment

The data strategy for student achievement outcomes (OECD, 1997b), which is now being implemented as PISA, was a major accomplishment for Network A and the culmination of many years of discussions and consensus-building about what it was that OECD countries wanted to measure, over and above, or distinct from, information already available through other sources. Most of the data available at the time focused on assessing student achievement as it related to school curricula and were developed for educational research purposes.³ While these data were (and still are) of importance to countries, countries also were becoming keenly interested in assessments that would go beyond schools and curricula and that would be suited to an international indicator program related to education policy. Hence, the core of the data strategy is that the OECD assessment program should support the development of policy-relevant indicators on the yield of learning of students at a given age that would be available with regularity. The key design features of PISA relate to, and elaborate on, this core concept.

Policy-relevant indicators...

PISA collects information on students' reading, mathematical, and scientific literacy (i.e., subject-bound domains), as well as on cross-curricular competencies, such as their abilities to regulate their own learning, to solve problems outside of single curricular contexts, and to use information and communication technologies (ICT), which also are thought to be important to students' future success in school and life.

PISA's focus on policy relevance has implications for reporting. For example, in addition to presenting overall mean scores and distributions across percentiles of performance, PISA identifies graduated levels of proficiency and describes the skills associated with those levels in order to provide policy makers with a more descriptive picture of what students can do. PISA is the first international assessment of a student population to identify proficiency levels. (The proficiency levels are described in more detail in the following section.)

... on the yield of learning

PISA aims to advance understanding of how well equipped young people are for their future lives. Thus, the content of the assessment is drawn, not from common elements of school curricula, but from a framework agreed to internationally on what it means for a student to be *literate* in (i.e., to be able to use and apply) PISA's three key subject areas. Figure 1 (p. 40-41) illustrates the definitions and components of the assessment frameworks for reading, mathematical, and scientific literacy.

Countries recognize, however, that further work is needed to better understand this concept and to validate that the literacies measured in PISA do, in fact, equip young people for success in their future lives. The PISA longitudinal study and linkages at the item level to the International Adult Literacy Study (IALS) are two opportunities to explore this issue further.

One of the implications of the yield notion is that the PISA items, while based in their respective subject areas, especially focus on presenting "real-life" problems and contexts. As discussed above, each of the PISA frameworks has as one of three major dimensions the contexts in which knowledge and skills are drawn upon, with personal, social, and occupational contexts having as much importance as educational ones. Also, PISA uses a combination of multiple-choice and constructed-response items, with a greater emphasis on the latter than in some past international studies of student achievement. An expert panel review in the United States of the mathematics and science items from PISA 2000, TIMSS 1999, and NAEP, for example, found that PISA items require multi-step reasoning more often than do the items from the other programs and that PISA 2000's mathematics and science literacy items more often involve the interpretation of charts and graphs or other real-world stimulus material (Nohara, 2001). The reading literacy assessment's use of both continuous (e.g., literature, textbooks, newspapers) and non-continuous (e.g., charts, tables, instructions) texts also presents a range of reading material to students. To the extent that open-ended items and authentic stimulus material from wide-ranging sources are more representative of real-life situations, PISA makes advances in this area.

... at a given age

Because PISA focuses on the overall outcomes of learning (rather than, for example, the outcomes of schooling or the curriculum per se), it uses an age-based sample, which reinforces the idea that what is being measured goes beyond school boundaries and also allows easier comparisons across countries. Although in the United States, Canada, and some other countries, most students continue beyond compulsory education an additional year or two to complete secondary school and perhaps go on to higher education, PISA shows how 15-year-olds in each country perform compared to their counterparts, all of whom will soon face the transition from school to work or to further education. PISA first samples schools (a minimum of 150 for most countries), then samples some 35 students within each school. For those countries that are interested, however, there is an option to also include a grade-based supplement to the PISA design.

... available with regularity

Finally, PISA employs a cyclical structure that allows for a rotating emphasis among the three key assessment domains (i.e., reading, mathematical, and scientific literacy) every three years. Each domain is assessed in each data collection, but the design distinguishes between “major” and “minor” domains. The major domain assessment includes more items and requires more testing time and provides a full measure of the domain. A minor domain assessment includes fewer items (though covering the ranges of the framework) and provides a data point for the examination of trends. Reading literacy was the major domain in 2000; mathematical literacy will be the major domain in 2003 and scientific literacy will be in 2006. In 2009, the cycle will begin again with reading literacy. This allows trend lines to be established in every assessment domain without overburdening students with the testing time or countries with the sample sizes that would be required for equal emphasis in every assessment year.

The use of proficiency levels, collection of information on cross-curricular competencies, use of the literacy concept (at least in mathematics and science) and age-based sample, and delineation of major and minor domains on a three-year cycle are the features that distinguish PISA from other well-respected international student assessment programs.

Figure 1: Summary of PISA Assessment Frameworks (2000)

Domain	Reading Literacy	Mathematical Literacy	Scientific Literacy
Definition	Understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.	Identifying, understanding, and engaging in mathematics and making well-founded judgments about the role that mathematics plays, as needed for an individual's current and future life as a constructive, concerned and reflective citizen.	Combining scientific knowledge with the drawing of evidence-based conclusions and developing hypotheses to help make decisions about the natural world and the changes made to it through human activity.
Components	Reading different kinds of text: continuous prose sub-classified by type (e.g., description, narration) and documents, sub-classified by structure.	Mathematical content – primarily mathematical “big ideas”. In the first cycle, these are change and growth, and space and shape. In future cycles, chance, quantitative reasoning, uncertainty and dependency relationships also will be used.	Scientific concepts – e.g., energy conservation, adaptation, decomposition – chosen from the major fields of physics, biology, chemistry, etc. where they are applied in matters to do with the use of energy, the maintenance of species or the use of materials.

Domain	Reading Literacy	Mathematical Literacy	Scientific Literacy
	Performing different kinds of reading tasks, such as retrieving specific information, developing an interpretation or reflecting on the content or form of the text.	Mathematical competencies, divided into three classes: (i) carrying out procedures; (ii) making connections and (iii) mathematical thinking and generalization.	Process skills – e.g., identifying evidence, drawing, evaluating, and communicating conclusions. These do not depend on a pre-set body of scientific knowledge but cannot be applied in the absence of scientific content.
	Reading texts written for different situations, e.g., for personal interest, or to meet work requirements.	Using mathematics in different situations, e.g., problems that affect individuals, communities, or the whole world.	Using science in different situations, e.g., problems that affect individuals, communities, or the whole world.

NOTE: The frameworks for mathematical and scientific literacy are being further developed for the cycle in which they will be the major domain.

SOURCE: From Table 1 (OECD, 1999), with information on mathematical and scientific literacy updated from various OECD working papers.

Technical Notes and Initial Results

OECD released the first results from PISA 2000 in December 2001, and the complete database is also now available.⁴ This section provides a sample of indicators from the PISA international report, which also were in EAG (OECD, 2002a), including examinations of overall student performance in reading literacy and of the sources of variation in student performance in reading literacy.

First, a few technical notes are in order. As noted before, PISA is the first international assessment of a school-age population to identify levels of proficiency for interpretation and reporting purposes. The thresholds, or cut points, for the levels were set such that all students within a level would be expected to answer at least half of the items from that level correctly, with those at the top of the level having a 62 percent chance of success on the hardest items in the level and those at the bottom having a 62 percent chance of success on the easiest items in the level. This is possible because of the use of Item Response Theory (IRT), which ranks items on a scale of difficulty according to how well students actually performed on them and then assigns scores to individuals according to how well they perform on a number of tasks of varying difficulty. The scale point assigned to each task is the point at which individuals with that proficiency score have a given probability of responding correctly. In PISA, the response probability was set at 62 percent (as above).

IRT scaling is used in most international assessments, as it allows performance to be summarized on a single scale even when different students are administered different items as happens in order to maximize assessment items and minimize testing time (see Appendix). However, the specific procedures may differ in detail from assessment to assessment. For instance, PISA 2000 used a mixed coefficients multinomial model, whereas IALS and TIMSS used somewhat more familiar two- and three-parameter models. This was a matter both of practicality, as the results ultimately did not differ between the models and the former was thought to adequately represent the data, and of preference for what is, in Europe, the more traditional model.

Figure 2 (p. 44) shows the percentage of students that reach each of the five proficiency levels and those that fall below the lowest level on the combined reading literacy scale overall and in selected countries.^{5,6} Looking at the final column, which shows the average for OECD countries, we see that 9 percent of all students in PISA reach

proficiency level 5 on the combined reading literacy scale. This indicates that these students have high-level skills such as evaluating information and building hypotheses, drawing on specialized knowledge, and accommodating concepts contrary to expectations. At the other end of the spectrum, 12 percent of students reach only level 1, which indicates more limited, basic skills such as the ability to recognize the main theme in a familiar topic or to make simple connections. Six percent of students did not reach proficiency level 1 and are at high risk—despite perhaps being able to read in the literal sense—that their reading skills may be insufficient for enabling further learning and thus limiting life and career opportunities.

Looking at the columns together, we also see variation in the patterns of proficiency between countries. For instance, Canada (like Finland and, to some extent, the United Kingdom) is a country in which a relatively large proportion of students function at proficiency level 5, a relatively large proportion function at least at level 3, and a relatively small proportion are at high risk, below level 1. Contrast this with Korea, another high-performing country on average, where there also are relatively few students at the lowest levels of reading literacy, but also relatively few at the highest levels. Korea has the largest proportion of students at level 3 of any participating country. The United States, an average overall performer in reading literacy, presents the opposite case. It has a relatively large percentage of students at level 5, but a large percentage at the lower extreme as well.

PISA 2000 also shows that there are considerable differences in performance in reading literacy within many countries (see Figure 2.5 in OECD, 2001). A variety of factors may contribute to this variation, such as differences in the socio-economic status of students or intake of schools, human and financial resources available, organizational structures, selection policies and practices, curricula, and instructional practices. Figure 3 (p. 46-47) shows the extent of variation associated with differences between schools and with differences within schools. For each country, variation either within or between schools is expressed as a percentage of the total OECD average variation in student performance (SP), or the square of the standard deviation, a technique which allows for the decomposition of the components of SP. Information from PISA is also available on the extent to which differences between schools are associated with geographical, systemic, and institutional factors. This can be enlightening for policy makers aiming to identify factors associated with variation in student performance.

Across OECD countries, differences within schools are associated with 65 percent of the average variation in student performance across countries in reading literacy, and differences between schools account for 36 percent of the variation. In other words, overall, there is more

Figure 2: Percentage of Students Reaching Each Proficiency Level on the Combined Reading Literacy Scale (2000)

NOTES: Countries are presented in descending order of the percentage of students at Level 5, the highest level of proficiency. Data are presented for select countries: Canada, Finland, France, Germany, Italy, Japan, Korea, Mexico, United Kingdom, and United States. This includes the three North American countries, the G-8 countries (except Russian Federation), and the highest performing Asian and European countries. "Ave" refers to the average of all OECD countries.

SOURCE: From Table 2.3 (OECD, 2001).

variation from student to student within schools than there is from school to school within countries.

This pattern, however, varies across countries. For example, in Finland, differences between students within schools are 77 percent of the average variation, and differences between schools are less than 11 percent of the average variation. This suggests that schools in Finland have similarly heterogeneously skilled populations – that most differences in performance happen at the individual student level and that schools do not systematically differ in comparison to other countries. Germany, on the other hand, has a between-school variation that is equal to 75 percent of the average variation in student performance across OECD countries, and within-school variation that is 50 percent of the average variation. Where there is substantial variation between schools and less variation within schools, as is the case in Germany, it suggests that students tend to be in schools in which their skill levels are similar to those of the other students. In Germany, this may relate to the division between vocational and academic schools at the secondary level.

Korea also provides an interesting example because of the relatively small percentage of variation accounted for by either between-school differences or within-school differences. Perhaps this is not surprising; however, as the previous paragraphs showed, Korea is a country with few students at the extremes and thus has low overall variation in student performance.

Recent and Future Development Work on Learner Outcomes

Since the implementation of PISA, the network has served as a “think tank” to inform PISA and decisions taken by the Board of Participating Countries regarding PISA. When there is interest in expanding PISA or considering how information from it can be used, Network A, with responsibility for the feasibility and relevance of the data strategy over time, provides guidance and assistance. Development work and technical assistance for PISA has focused on three areas: planning and analytical work, exploring new domains, and supporting the evaluation and continuation of the data strategy that underpins PISA. Additionally, the network undertakes activities not specifically related to PISA but rather related to the support and expansion of INES more generally. In particular, the work to explore new domains for PISA and to support INES in general is described more fully below.

Figure 3: Percentage of the Average Variation in Student Performance on the Combined Reading Literacy Scale that is Between Schools and Within Schools, Across All Participating OECD Countries, Expressed Relative to the Overall Variation in OECD Countries (2000)

NOTES: Countries are presented in ascending order of the percentage of variation that is between schools. Data are presented for select countries: Canada, Finland, France, Germany, Italy, Japan, Korea, Mexico, United Kingdom, and United States. This includes the three North American countries, the G-8 countries (except Russian Federation), and the highest performing Asian and European countries. Data for France are missing. "Ave" refers to the average of all OECD countries.

The total variation in student performance (SP) shown is the statistical variance and is obtained as the square of the standard deviation as shown in Table 2.3 (OECD, 2001). The statistical variance in SP and not the standard deviation is used for this comparison to allow for the decomposition of the components of variation in SP. The sum of the between- and within-school variance components may for some countries differ slightly from the square of the standard deviation for reasons explained in the PISA 2000 Technical Report (OECD, 2002b).

For each country, variation either within or between schools is expressed as a percentage of the total OECD average variation in SP, which is why between- and within-school variation can sum to over 100 percent in a given country. In countries where the sum is greater than 100, this indicates that variation is greater in that country than the average variation across countries.

SOURCE: From Table 2.4 (OECD, 2001).

Exploring New Domains

Network A supports the development of new domains for PISA, which has focused mainly on cross-curricular competencies (CCCs), in both the practical and theoretical realms. For example, PISA 2000 included a self-report instrument on self-regulated learning, which was developed under the sponsorship of several Network A countries (Belgium [Flemish community], Germany, Netherlands, and Norway) and which built on the earlier CCC feasibility study. The instrument collects self-reported information on students' learning strategies, motivation, and self-related cognitions, which together can provide an indication of students' preparedness for continual and self-directed learning, as well as a set of variables to relate to students' performance in the major and minor domains (Baumert et al., 1998).

More recently, for PISA 2003, Network A guided the development of a framework to assess students' problem solving ability, which is widely recognized by countries as central for citizens' academic, professional/vocational, and personal success. The framework focuses on students' ability to recognize, understand, formulate, strategize, solve, and reflect upon and communicate solutions to complex, real-life problems that require transfer of knowledge across curricular boundaries (Dossey et al., 2000).

Currently, the network is overseeing a group of experts charged with developing a framework for assessing students' information and communications technology (ICT) literacy in PISA 2006. Like the other PISA frameworks, the ICT framework will focus on literacy, or the ability to apply ICT skills for broader purposes. Though still under development, the framework will likely be organized similarly to the other domains with processes (e.g., access, manage), technologies (e.g., digital technologies, communication tools), and contexts (e.g., personal, educational) as organizing components.

Finally, simultaneous to these design endeavors, the Definition and Selection of Competencies (DeSeCo) Project worked on a separate track under the sponsorship and leadership of Switzerland, with the assistance of the United States and Canada. Growing from discussions in Network A, the goal of the DeSeCo Project was to provide a conceptual framework for defining and selecting competencies (and in particular a set of key competencies), which would be rooted in theory, interpretable in policy terms, and useful for informing the possible future assessment of such key competencies at the international level

(Rychen, 2001). DeSeCo used a multi-disciplinary approach, soliciting input from different fields (e.g., psychology, philosophy, education) and validating findings through two international symposia and against the backdrop of a review of competencies-related activities at the country level (Rychen, 2001; Salganik & Stephens, 2003, pending).

Supporting the Continual Development of INES

While activities to support and inform PISA require much of the network's attention, it is not the sole focus; the network also works to support and expand the INES Project more broadly. Countries have agreed that not only could PISA not bear the burden of all areas in which members are interested, neither should it, as some topics may be better explored outside the confines of PISA. This is the case with information on teaching and learning, which countries identified as a major gap in the INES project. Adult skills is another broad area in which the network may be interested for future development work.

Given PISA's nature as a yield study that examines the cumulative effect of a student's entire educational experience, it would be difficult to incorporate a survey on teachers and teaching, which would capture only part of that experience, to any meaningful result. Therefore, when countries prioritized teaching as one of the key areas in need of development at the most recent INES General Assembly, the networks with related scopes of work (Networks A and C, on schools and school processes) readily decided to undertake this work as a separate track. In January 2002, the two networks jointly established a Task Force on Teaching and Learning, which is charged with developing a long-term strategy to increase the information available to the INES project on teachers, teaching, and the impact teaching can have on learning. The first stage of this activity is to develop a preliminary conceptual framework that identifies the different factors that are believed to be related to teaching and learning (and how they relate to each other) and that draws on the INES framework of multiple levels (e.g., system, school, etc.). Ultimately, the framework should help illuminate the potential for effective student engagement, which is at the heart of the educative process.

Another area that the network may consider for future development is the area of adult skills. While there does exist already a major assessment program, the Adult Literacy and Lifeskills (ALL) Study, developed and organized by Statistics Canada and focused on literacy, numeracy, and analytical reasoning, many OECD countries have

expressed an interest in looking ahead to broadening the scope of assessment of adult skills for future assessments. Network A may be interested in looking at the findings from the DeSeCo Project and building from existing activities to identify a data strategy for a broad range of adult skills. For instance, the DeSeCo Project identified three constructs—acting autonomously, using tools interactively, and joining and functioning in socially heterogeneous groups—that are linked to their theoretical framework and offer a conceptual basis for clustering and describing key competencies in possible future assessments (Rychen & Salganik, 2002; Rychen, 2003).

Summary

In sum, Network A aims to advance the assessment and dissemination of information about learner outcomes. Its early work focused on developing technical standards for indicator work, on establishing a data strategy to ensure the availability of timely and policy-relevant information on student performance in key subject areas at a key stage of life, and on beginning to think about how to measure new domains of interest outside of key subject areas. A major focus of more recent work is overseeing the data strategy that underlies PISA and developing new domains for PISA. However, the network also continues to work in partnership with other groups in areas outside of PISA, such as in developing new ways to think about measuring teaching and learning and, potentially, a broad range of adult skills. Finally, drawing upon these sources and more, the network prepares indicators for dissemination in EAG. With the other INES working groups, Network A aims to develop, produce, and disseminate information that will be useful for informing policy and ultimately improving education systems.

NOTES

This paper is intended to promote the exchange of ideas among researchers and policy makers. The views expressed in it are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education.

1. Members include: Australia, Austria, Belgium (Flemish and French communities), Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, Korea,

Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States.

2. For additional information on the early history of Network A and the genesis of PISA, see Owen, Hodgkinson, & Tuijnman, 1995.
3. The main assessment program at the time with a school-age target population was TIMSS (Trends in International Mathematics and Science Study, then still the Third International Mathematics and Science Study). TIMSS provided, and continues to provide, a wealth of information about the mathematics and science achievement of 4th and 8th grade students in over 40 countries.
4. Additional information on PISA results can be found in the international report, *Knowledge and Skills for Life: First Results from PISA* (OECD, 2001); numerous national reports including *Outcomes of Learning: Results from the 2000 Program for International Student Assessment of 15 Year-Olds in Reading, Mathematics, and Science Literacy* (Lemke et al., 2001) and *Measuring Up: The Performance of Canada's Youth in Reading, Mathematics, and Science* (Bussière et al., 2001); and on the PISA website <http://www.pisa.oecd.org>. Forthcoming reports from OECD will provide more in-depth information, focusing on specific thematic topics, such as students' social background, reading literacy, or high- and low-achievers.
5. PISA also provides information on three subscales (retrieving information, interpreting, and reflecting and evaluating), but these data are not explored here. Combined reading literacy refers to overall performance in reading literacy.
6. The cut points for the proficiency levels were as follows: below level 1 (334 or below); level 1 (335-407); level 2 (408-480); level 3 (481-552); level 4 (553-625); and level 5 (626 and above).

REFERENCES

- Baumert, J., Fend, H., O'Neil, H.F., & Peschar, J.L. (1998). *Prepared for life-long learning: Frame of reference for the measurement of self-regulated learning as a cross curricular competence (CCC) in*

- the PISA project*. Unpublished paper prepared for the Organization for Economic Cooperation and Development (OECD).
- Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, Y. (2001). *Measuring up: The performance of Canada's youth in reading, mathematics, and science*. Ottawa, ON: Statistics Canada.
- Dossey, J., Csapo, B., de Jong, T., Klieme, E., & Vosniadou, S. (2000). Cross-curricular competencies in PISA: Towards a framework for assessing problem-solving skills. In *The INES compendium: Contributions from INES networks and working groups*, prepared for Fourth General Assembly of the OECD Education Indicators Program, Tokyo, 11-13 September, 2000.
- Granheim, M., & Pettersson, S. (1995). Goals orientation and attainment in learning systems. In *Measuring what students learn* (pp.165-177). Paris: OECD.
- Grisay, A. (1995). Reflections on some unsuccessful indicators. In OECD. *Measuring what students learn* (pp. 63-80). Paris: OECD.
- Knutsen, A.E. (1995). The GOALS study: Analysis and implications. In OECD. *Measuring what students learn* (pp. 179-193). Paris: OECD.
- Lemke, M., Lippman, L., Bairu, G., Calsyn, C., Kruger, T., Jocelyn, L., et al. (2001). *Outcomes of learning: Results from the 2000 Program for International Student Assessment of 15-year-olds in reading, mathematics, and science literacy*. Washington, D.C.: U.S. Department of Education.
- Ministry of National Education (1995). *An evaluation of national assessments*. Paper for INES Network A Meeting, March 1995, Volterra, Italy: Author.
- Nohara, D. (2001). *A comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)*. Working Paper No. 2001-07. Washington, D.C.: National Center for Education Statistics.
- OECD. (1997a). *Intended goals in education systems: Final report*. Paris: Author.
- OECD. (1997b). *International indicators of education systems: Network A strategy for student achievement outcomes*. Paris: Author.
- OECD. (1997c). *Prepared for life*. Paris: Author.
- OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.
- OECD. (2001). *Knowledge and skills for life: First results from PISA*. Paris: Author.

- OECD. (2002a). *Education at a glance*. Paris: Author.
- OECD. (2002b). *PISA 2000 technical report*. Paris: Author.
- Owen, E.H., Hodgkinson, D., & Tuijnman, A. (1995). Towards a strategic approach for developing international indicators of student achievement. In *Measuring what students learn* (pp. 211-229). Paris: OECD.
- Owen, E.H., Moskowitz, J., Stephens, M., & Gil, G.A. (2003). Toward educational improvement: The future of international assessment. In J. Moskowitz & M. Stephens (Eds.), *Comparing learner outcomes*. London: Routledge Falmer. (publication pending).
- Rychen, D.S. (2001). Introduction. In D.S. Rychen & L.H. Salganik (Eds.), *Defining and selecting key competencies*. Bern, Switzerland: Hogrefe & Huber Publishers.
- Rychen, D.S. (2003). A holistic model of competence. In D.S. Rychen & L.H. Salganik (Eds.), *Key competencies for a Successful Life and a Well-Functioning Society* (pp. 41-62). Bern, Switzerland: Hogrefe & Huber Publishers.
- Rychen, D.S., & Salganik, L.H. (2002). *DeSeCo strategy paper on key competencies: A frame of reference for future assessment programs*. Neuchâtel, Switzerland: Swiss Federal Statistical Office.
- Salganik, L.H., & Stephens, M. (2003). Competence priorities in policy and practice. In D.S. Rychen & L.H. Salganik (Eds.), *Key Competencies for a Successful Life and a Well-Functioning Society* (pp. 13-40). Bern, Switzerland: Hogrefe & Huber Publishers.
- Trier, U.P. & Peschar, J.P. (1995). Cross-curricular competencies: Rationale and strategy for developing a new indicator. In *Measuring what students learn* (pp. 97-108). Paris: OECD.
- UNESCO. (1990). *World declaration on education for all: Meeting basic learning needs*. Statement adopted by the World Conference on Education for All in Jomtien, Thailand, March 1990.

Appendix: Notes on PISA Design

As with other international assessments, PISA uses a rotated booklet design, in which students answer different combinations of items from the total item pool. PISA 2000's seven hours of items were organized into nine booklets, each of which presented two hours of testing material for an individual student. In PISA 2000, every student answered reading literacy (major domain) items. Half of the students in each country also answered mathematical literacy items and the other half scientific literacy (minor domain) items.

Students answered a background questionnaire that took about 30 minutes to complete. Some also completed additional questions on learning and study practices (as part of the international option on self-regulated learning) and on their familiarity with computers. School principals also completed a questionnaire about their schools. Background information from these sources is being used to examine the factors that are associated with performance outcomes and how these factors relate similarly or differently across countries.

Correspondence concerning this article should be Eugene Owen,
National Center for Education Statistics, U.S. Department of Education
Email: Eugene.Owen@ed.gov