CANADIAN JOURNAL OF SCIENCE, MATHEMATICS AND TECHNOLOGY EDUCATION
2016, VOL. 16, NO. 1, 92–108
http://dx.doi.org/10.1080/14926156.2015.1119332

Routledge
Taylor & Francis Group

Canadian and United States Students' Performances on the OECD's PISA 2012 Problem-Solving Assessment

John A. Dossey [a] and Joachim Funke [b]

[a]Department of Mathematics, Illinois State University, Normal, Illinois; [b]Department of Psychology, University of Heidelberg, Heidelberg, Germany

ABSTRACT

This article presents an overview of the Organization for Economic Cooperation and Development's (OECD) Programme for International Student Assessment (PISA) 2012 Problem-Solving assessment. The assessment examined the capabilities of 15-year-olds in 40 nations and four large international cities, as well as the Canadian Provinces, to solve a set of 16 problem units presented in contextualized situations. These units each had two or three tasks for students to solve. Students' performances on these items served as a basis for developing the PISA score scale and the associated proficiency levels for each of the participating entities. Student performances on the PISA 2012 mathematics, reading, and science assessments provided an avenue to estimating student performances on the problem-solving assessment for students having similar score backgrounds in the three content areas. Student performances were also examined by item types and by differing problem situations. These latter two analyses provide interesting perspectives for comparing problem solving profiles across the participating entities. Finally, some conclusions are drawn concerning the study as a whole.

RÉSUMÉ

Cet article présente dans ses grandes lignes l'Évaluation PISA 2012 en résolution de problèmes, de l'OCDE. Cette évaluation analyse les compétences de jeunes de 15 ans dans 40 pays, 4 grandes villes internationales, et aussi dans les provinces canadiennes, pour résoudre une série de 16 unités de problèmes présentés en contexte. Chacune de ces unités comprenait 2 ou 3 tâches à faire résoudre par les élèves. Leurs performances dans chacune de ces tâches ont servi à mettre au point l'échelle de pointage du PISA et les niveaux de compétences liés à chacun des groupes participants. Les performances des étudiants à l'évaluation du PISA 2012 en mathématiques, en lecture, et en sciences ont ouvert une voie permettant d'estimer quelles seraient les performances en résolution de problèmes des élèves ayant obtenu des scores comparables dans les trois domaines de contenus. Les performances ont aussi été analysées par type d'élément et par type de situation problématique. Ces deux dernières analyses fournissent des perspectives intéressantes permettant de comparer différents profils de résolution de problèmes dans tous les groupes de participants. Enfin, nous tirons certaines conclusions au sujet de l'étude dans son ensemble.

CONTACT John A. Dossey jdossey@illinoisstate.edu 65111 East Crystal Ridge, SaddleBrooke, Tucson, AZ 85739.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ucjs.

What is the OECD and PISA and why are they interested in 15-year-olds' capabilities in problem solving?

The Organization for Economic Cooperation and Development (OECD) is an international organization, headquartered in Paris, France, consisting of 34 member democracies having market economies, including Canada and the United States. These countries share information and compare and contrast their policies, experiences, and goals with special emphases on economic growth, prosperity, and societal development. Central to these goals are inputs and outputs related to the education and resulting capabilities of each nation's youth.

As such, each of the OECD countries is interested in the future composition and qualifications of its workforce, as well as sustaining its resources while improving the living conditions of its citizens. In 2000, the OECD initiated a triennial assessment program focused on reading, mathematics, and science, known as PISA (Programme for International Student Assessment) and focused it on 15-year-olds' knowledge in the foregoing areas. These assessments, given to 15-year-old students near the end of compulsory education for many of the OECD countries, are less an achievement test of facts learned but, rather, an assessment of how prepared these students are to use their school and life acquired knowledge in real-world contextual situations (OECD, 2014a). In 2012 the OECD PISA program added a problem-solving assessment to its assessment program.

In 2003, PISA had an optional problem-solving assessment, but it was a paper-and-pencil test. The PISA 2012 Problem-Solving assessment was computer delivered and the tasks students faced were more challenging and dynamic in nature. The 2012 problem-solving assessment was also optional, because it was a new, but continuing, feature for the triennial assessment for participating countries. The computer delivery format allowed the presentation of visual and graphic information, along with some displays showing the movement of problem components, on a computer screen. Students, in turn, responded by entering their answers and marking significant parts of problem displays to illustrate their answers.

In all, 44 governmental entities participated in the 2012 PISA Problem-Solving assessment. The participants consisted of 40 countries, composed of 28 OECD countries and 12 non-OECD countries, and four economies representing large international cities. See Table 1 for a complete listing of the participants.

What is PISA problem solving?

The PISA 2012 definition of *problem solving* is grounded in the generally adopted notions of *problems* and *solving problems in context*. As such, PISA problem solving is defined as

> Problem solving competency is an individual's capability to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential as a constructive and reflective citizen. (OECD, 2013, p. 122)

In the PISA 2012 Problem-Solving assessment, over one half of the tasks are *interactive* (OECD, 2013). Examples of interactive problems encountered in life include discovering how to use an unfamiliar cellular telephone or automatic vending machine. These are examples of the situations the problem-solving research literature calls *complex problems* (Dörner, 1997; Frensch & Funke, 1995a; Sternberg & Frensch, 1991). Frensch and Funke (1995b) define complex problems through the relationship of givens, goals, and barriers as follows:

> The given state, goal state, and barriers between given state and goal state are complex, change dynamically during problem solving, and are intransparent. The exact properties of the given state, goal state, and barriers are unknown to the solver at the outset. (p. 18)

The problems in the assessment are divided by the nature of the problem situation into two classes as defined below. Each of the problems consists of an opening explanation of a context, or situation, followed by one or more tasks concerning the problem requiring a response from participating students. In all there were 16 problem situations divided into 42 tasks requiring a student response in the assessment. Six of the problems consisted of two tasks each and 10 of the problems consisted of three tasks each.

Table 1. Participating entities in the PISA 2012 Problem-Solving assessment.

OECD countries	Non-OECD countries	Large cities or regions
Australia	Brazil	Hong Kong-China
Austria	Bulgaria	Macao-China
Belgium	Chinese Taipei	Shanghai-China
Canada	Columbia	Singapore
Chile	Croatia	
Czech Republic	Cyprus[a]	
Denmark	Malaysia	
Estonia	Montenegro	
Finland	Russian Federation	
France	Serbia	
Germany	United Arab Emirates	
Hungary	Uruguay	
Ireland		
Israel		
Italy		
Japan		
Korea		
The Netherlands		
Norway		
Poland		
Portugal		
Slovak Republic		
Slovenia		
Spain		
Sweden		
Turkey		
England (UK)		
United States		

[a]The listing of Cyprus in this column is aware of the differences between Turkey and the United Nations and other countries concerning Cyprus. As such, the data in this report associated with Cyprus relate solely to the southern part of the island. In this article, when the data description refers to 44 participants, the Cyprus data are included. When the data description refers to 43 participants, the Cyprus data are omitted.

Using this vocabulary, PISA's interactive problems (27 tasks) are intransparent (i.e., there is undisclosed information) but not necessarily dynamic or highly complex. Most of the interactive problems in PISA 2012 were constructed using mathematical models whose parameters can be manipulated to achieve differing degrees of item difficulty. The two main models used consisted of contexts built around dynamical systems defined by a linear recurrence. In these problems, students were given or were directed to develop a model where the changes observed follow a fixed rule dependent on a starting point and where each new point is defined by the last known point and the action of the rule (Funke, 1993). The other model underlying PISA problems consisted of contexts focused on finite-state machines (Buchner & Funke, 1993). These problems are characterized by a set of states, a finite set of inputs, or actions that can be performed and a function f linking inputs and the current state by a rule to send the machine into a new state. An example of such a machine is a turnstile gate at a metro station operated by a token system. The turnstile has two states, *unlocked* and *locked*. There are two inputs that can be performed, one is *insert a coin* and the second is *push the turnstile*. Both the dynamical systems and finite-state machine–based models have been used extensively in problem solving research (Funke, 2001).

The *static* problems (15 tasks) making up the remainder of the PISA problem set consisted of tasks where all of the information necessary to solve a problem is disclosed to the problem solver at the outset; they are completely transparent by definition. These problems are more similar to the exercises and problems presented in a textbook.

All problems were delivered to students via computers in the problem-solving assessment. Thus, the information for both the context and the description of the related task were presented in verbal, tabular, graphical, or visual form for students on their computer screen. For the interactive items the computer's responses to the intermediate steps a student took in responding to a question in a task were moderated by the student's responses to previous steps in the same task.

Proficiency levels

Students' total scores for the 42 problem-solving tasks on the 2012 assessment allowed the placement of students into one of seven proficiency levels. Six of these levels, numbered 1 through 6, are based on the description of problem solving in the PISA 2012 Assessment Framework and student work observed in the 2012 assessment (OECD, 2014a). The seventh level, "below level 1," refers to work of students who performed below the lowest described level. The levels and their associated numerical scale score interval of proficiency are described in PISA 2012 Problem-Solving report. Because of space, we present here only the endpoints of level 6 and level 1, together with a medium level 3 description:

Level 6 (equal to or above 683 points)

At Level 6, students can develop complete, coherent mental models of diverse problem scenarios, enabling them to solve complex problems efficiently. They can explore a scenario in a highly strategic manner to understand all information pertaining to the problem. The information may be presented in different formats, requiring interpretation and integration of related parts. When confronted with very complex devices, such as home appliances that work in an unusual or unexpected manner, they quickly learn how to control the devices to achieve a goal in an optimal way. Level 6 problem-solvers can set up general hypotheses about a system and thoroughly test them. They can follow a premise through to a logical conclusion or recognize when there is not enough information available to reach one. In order to reach a solution, these highly proficient problem-solvers can create complex, flexible, multi-step plans that they continually monitor during execution. Where necessary, they modify their strategies, taking all constraints into account, both explicit and implicit. (OECD, 2014a, p. 57)

Level 3 (488 to less than 553 points)

At Level 3, students can handle information presented in several different formats. They can explore a problem scenario and infer simple relationships among its components. They can control simple digital devices, but have trouble with more complex devices. Problem-solvers at Level 3 can fully deal with one condition, for example, by generating several solutions and checking to see whether these satisfy the condition. When there are multiple conditions or inter-related features, they can hold one variable constant to see the effect of change on the other variables. They can devise and execute tests to confirm or refute a given hypothesis. They understand the need to plan ahead and monitor progress, and are able to try a different option if necessary. (OECD, 2014a, p. 57)

Level 1 (358 to less than 423 points)

At Level 1, students can explore a problem scenario only in a limited way, but tend to do so only when they have encountered very similar situations before. Based on their observations of familiar scenarios, these students are able only to partially describe the behavior of a simple, everyday device. In general, students at Level 1 can solve straightforward problems provided there is only a simple condition to be satisfied and there are only one or two steps to be performed to reach the goal. Level 1 students tend not to be able to plan ahead or set sub-goals. (OECD, 2014a, p. 57)

In addition to the above levels, there was a classification level containing those students whose performance fell below any of the above described proficiency levels found in level 1 through level 6.

The tasks are scaled to have a mean score among OECD countries of 500 points and a standard deviation of approximately 100 points. This results in approximately two thirds of the students across OECD countries having proficiency scores between 400 and 600 points. Students in the lowest classification level had proficiency scores below 358 points (OECD, 2014a). Though such descriptions of levels provide a glimpse of what students might do in hypothetical settings, a look at actual problems aids understanding.

Sample PISA problem-solving tasks

Very few tasks have been publicly released from the PISA 2012 Problem-Solving assessment. The following sample of tasks shows an example of a static task based on relations in a graphical network and an interactive task based on a finite-state machine paradigm. Further examples of tasks can be viewed in

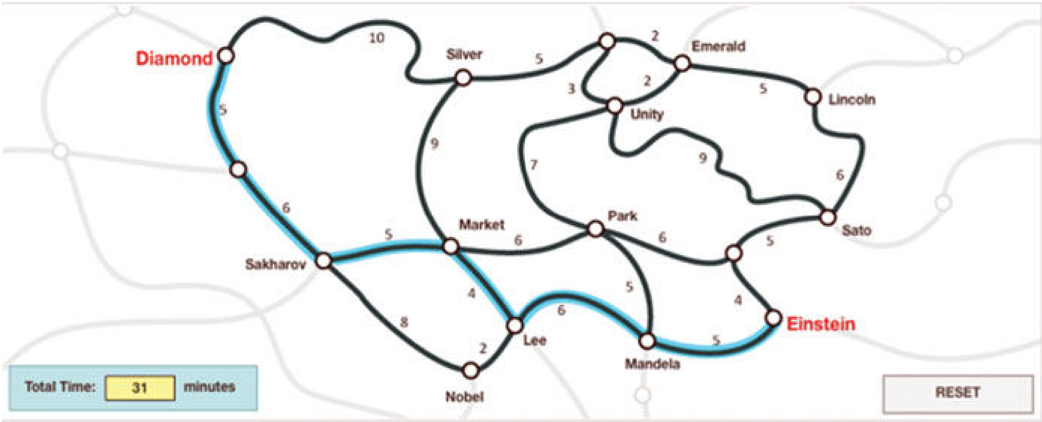

Figure 1. Second task of a three-task set centered on short travel times in traffic. Adapted from *Explore PISA 2012 Mathematics, Problem Solving and Financial Literacy Test Questions*, by Organization for Economic Cooperation and Development, 2014c, Paris, France: OECD Publishing.

a live computer simulation mode at the PISA website (OECD, 2014c). The complete report of the 2012 problem-solving study is available for free downloading at the PISA site (OECD, 2014a).

A static task

The first problem we examine is the second task in a set of three tasks in a problem associated with a map, or weighted graph, whose edges are labeled to show the travel time between adjacent suburbs. Students are asked to find the route associated with the shortest travel time between two suburbs and are asked to click on the edges, representing segments of roads, to show the route. They are given the minimum travel time and a computer app on the screen that calculates the time associated with the edges they select. The task is a static task in that all information is disclosed at the outset and the stem even provides the correct answer against which they can check their response. Figure 1 shows the task as presented to students with the exception that the correct route has been highlighted.

The mean of the 28 OECD country-level averages for students receiving full credit on this task was 70%. Canada and the United States were the only two North American countries participating in the PISA 2012 assessment. Seventy-seven percent of Canadian students correctly answering received full credit on this task, whereas 73% of U.S. students did the same. The OECD scale score for a full-credit response to this traffic task was 446, placing the OECD performance on the item midway through the level 2 proficiency interval.

An interactive task

The second example of a task is situated in a train ticket kiosk context. Field-testing of the item indicated no significant task discrimination bias favoring one country over another that might have been expected due to differential student familiarity with such ticket kiosks. This task asks students to purchase two individual trip tickets. Students are alerted that they qualify for a concession ticket, but if and when they attempted to select that alternative, the computer responds: “There are no tickets of this type available. Please press CANCEL and buy a different ticket.” Students then have to regroup and move back to the city subway choice and make a choice between two alternative ticketing choices within this category. The staging of the task is shown in Figure 2.

The mean country performance for the 28 OECD participating countries on the train task was 48% correct.

Fifty-five percent of Canadian students received full credit, whereas 51% of U.S. students received the same percentage correct. The OECD scale score for full credit for this ticket task was 638, placing the OECD performance one third of the way up into the interval associated with level 5 proficiency. Central

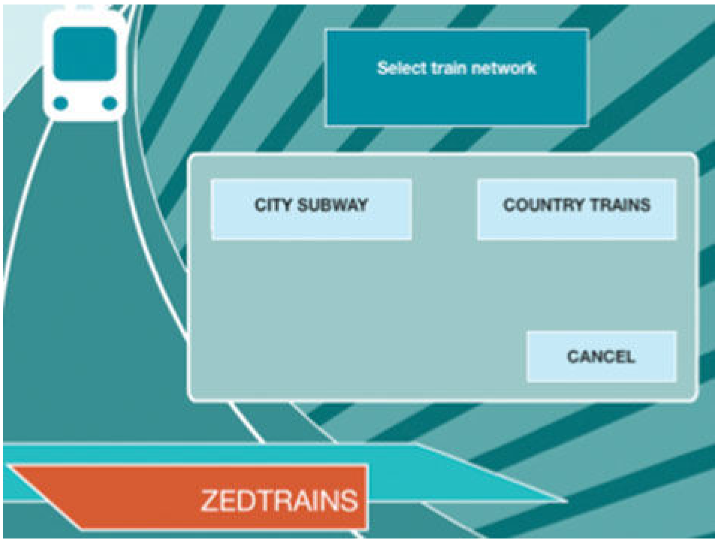

Figure 2. Second task of a three-task set centered on the purchase of tickets in TRAIN. Adapted from *Explore PISA 2012 Mathematics, Problem Solving and Financial Literacy Test Questions*, by Organization for Economic Cooperation and Development, 2014c, Paris, France: OECD Publishing.

to the students' performance was the capability to recover midway into a solution to the new added information, switch to a new approach, and launch another solution strategy and then, within this new strategy, compare two alternative prices, selecting the more economical solution.

Methodology

In order to ensure international comparability of the national populations of 15-year-olds, PISA employs an age-based definition targeting eligible participants as students who have completed 15 years and 3 months of age and are short of completing 16 years and 3 months of age. In the PISA Problem-Solving study, the mean age of the students in the sample was 15 years and 9 months (OECD, 2014a). Because PISA is not a direct achievement test of content that has been explicitly taught, sampling by age bands assures that the assessment is aimed at comparing like groups of students age-wise.

The participating nations and economies were urged to maximize the inclusion of all students in their student samples. All but eight of the participating countries and economies achieved a standard of excluding 5% or fewer of their students following the PISA-defined sampling standards. Among these eight failing to meet the criterion were Canada (6.38%) and the United States (5.35%). The United States did meet the 5% goal once language adjustments were made for all nations. Canada's adjusted exclusion rate remained just over the goal and was a result of accommodating special education and language issues (P. Brochu, personal communication, August 28, 2014).

For the actual PISA Problem-Solving study, the 16 problem situations were divided into four clusters, or blocks, of problems. Students were allotted 20 minutes to complete each block of problem-solving situations they were assigned. A given problem-solving situation and its tasks in the problem-solving assessment was administered to about one half of the sample of students assessed in a country. This process assured that each cluster appeared in each position possible in the varied multiblock assessment forms and that each cluster also appeared in the overall PISA assessment forms with each of the other mathematics, reading, and science clusters as well. This balanced incomplete block design made it possible to construct a single scale for problem solving and to link student performances in problem solving with performances in mathematics, reading, and science (OECD, 2014b).

This balanced, spiraled, incomplete block design made it possible to impute estimated scores for each student for each of the 42 tasks contained in the assessment, using processes similar to those used in the international Trends in Mathematics and Science Studies and the U.S. National Assessment of Educational Progress.

Table 2. Percentage of students at each proficiency level in PISA 2012 Problem-Solving.

Jurisdiction	< Level 1	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Canada	5.0	9.6	19.0	25.8	22.9	12.4	5.1
United States	5.7	12.5	22.8	27.0	20.4	8.9	2.7
OECD average	8.2	13.2	22.0	25.6	19.6	8.9	2.5

Findings

Distribution of students by proficiency

Table 2 contains the distributions of observed problem-solving proficiency by percentage for Canadian, the United States, and OECD student performances relative to each of the problem solving proficiency levels. Analysis of the data indicates that Canadian students performed significantly higher than U.S. students on the PISA 2012 Problem-Solving assessment and that both Canadian (526) and U.S. (508) students performed significantly higher than the OECD average scale score (500). The OECD mean is the average of the national mean scores of the participating OECD countries scaled to a mean score of 500 with a standard deviation of 100 points.

An examination of the data in Table 2 shows that the difference in performance between Canadian students and U.S. students is rooted in the smaller percentage of Canadian students in the groups in below level 1 through level 3 and a greater percentage of Canadian students in levels 4 through 6. The U.S. student performance showed that a greater percentage of U.S. students were in each proficiency level at or below level 3 and a greater percentage of Canadian students were found performing at levels 4 through 6. Further, the U.S. performance profile was very similar to that of the OECD countries but with fewer students in the lower two levels and moderately more students in levels 2 through 4.

Both Canada and the United States performed statistically better than the OECD mean score on the problem-solving assessment, with Canada ranking third among the OECD nations participating and eighth among the 44 participating entities. The U.S. ranked 13th among the group of 28 OECD nations and 18th with respect to the full set of 44 participating nations or economies. The average scale score for Canadian students was 526, whereas the average scale score for U.S. students was 508. These mean scale scores placed each country's overall performance in the level 3 proficiency interval (488 to less than 553 points) on the PISA 2012 problem-solving scale. The difference in mean performance of Canadian and U.S. students was statistically significantly (OECD, 2014a).

Canada's 10 provinces each participated in the PISA 2012 Problem-Solving study through an OECD program that allowed nations or subregions of a nation to enhance the national sample to develop PISA performance reports for a subregion or subregions. All 10 Canadian provinces elected to participate in the PISA 2012 assessment, including the problem-solving study. The results of the provincial data for problem-solving proficiency scale scores are shown in Figure 3 and Table 3. The OECD and U.S. averages have been added to both Figure 3 and Table 3 as points of comparison. In Figure 3, the different shading of neighboring columns of the provincial averages, the total Canadian average, the U.S. average, and the OECD average indicate that the corresponding scale scores for differently shaded problem solving are significantly different.

An analysis of the data indicates that British Columbia's mean performance level was significantly higher than all other provincial groups, as well as the performance of Canada as a whole. Alberta, Ontario, and Quebec do not significantly differ from each other or from Canada but are significantly higher than the groups to the right in the figure. New Brunswick, Saskatchewan, and Nova Scotia's performances are significantly lower than the Canadian average but do not significantly differ from each other or from the performance of U.S. students. Manitoba and Newfoundland perform significantly lower than the United States but significantly above the OECD mean score. Prince Edward Island's performance was significantly beneath that of the mean OECD performance (Council of Ministers of Education, Canada [CMEC], 2014).

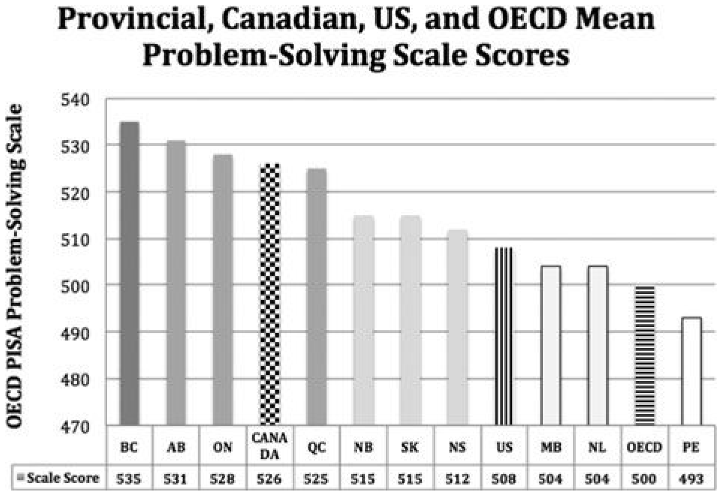

Figure 3. Performance of Canada, the United States, and Canadian Provinces on the OECD PISA or CMEC Problem-Solving assessment. Adapted from *Assessment Matters! Issue 6—How Good Are Canadian 15-Year-Olds at Solving Problems? Further Results From PISA 2012* (p. 3), by Council of Ministers of Education, Canada, 2014, Toronto, Canada: Author; and *PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems* (p. 52), by Organization for Economic Cooperation and Development, 2014a, Paris, France: OECD Publishing.

U.S., Canada, and broader global comparisons than the OECD

Table 3 lists the mean PISA 2012 scale scores for problem solving on the PISA 2012 Problem-Solving assessment for nations, economies, and the Canadian provinces. The OECD nations are listed on the left in the first column, the partner entities (non-OECD nations and the economies) are listed centered in the middle of the first column, and the provinces of Canada are listed to the right of the first column.

An examination of the data in Table 3 allows one to view the OECD countries, partner nations and economies, and the Canadian provinces in numerical scale score order while maintaining track of their individual group status in the overall PISA Problem-Solving assessment.

Correlation of problem-solving proficiencies with those from other domains

Because students in each of the participating entities in the PISA 2012 Problem-Solving study also had PISA 2012 mathematics, reading, and science scale scores as well, latent correlations among the participating students' scale scores in the three core domains and problem solving were calculated.

Table 4 contains three latent correlation tables, the first one, Table 4A, for all students from the participating entities having performances in the three content domains as well as in PISA 2012 problem solving. The following two tables, Tables 4B and 4C, provide similar latent correlation tables for the Canadian students and then for the U.S. students, respectively. All of these latent correlations were positive and significantly different from 0.

An examination of the latent correlations show that (a) the correlations among content domains and problem solving were large; (b) different content domains correlated with problem solving in different ways, descending from with mathematics, to science, and then to reading; and (c) as countries, Canada and the United States had similar patterns of latent correlations with the correlation between problem solving and reading being the lowest and the correlations between problem solving and mathematics the greatest.

An analysis of the joint latent correlations of reading, mathematics, and science indicated that 68% of the variance in problem-solving scores was explainable from skills and concepts in mathematics, reading, science, or some combination of these domains. The remaining 32% of the variability in the problem-solving score results from other concepts, knowledge, and skills required by the problem-solving assessment.

Using this finding, an expected problem-solving score was computed using multiple regression where student data from the three core content domains were used as predictor variables. Then, finding the differences between students' actual problem-solving scores and their expected problem-solving scores

Table 3. Scale scores for the 44 participating entities in PISA 2012 Problem-Solving.

Participating nations and governmental jurisdiction for PISA 2012 Problem-Solving and Canadian Provincial PISA 2012 Problem-Solving, all on the PISA 2012 Problem-Solving scale

OECD, *Jurisdiction*, or ***Province***	Scale score on OECD PISA Problem-Solving scale
Singapore	562
Korea	561
Japan	552
Macao-China	540
Hong Kong-China	540
Shanghai-China	536
British Columbia	535
Chinese Taipei	534
Alberta	531
Ontario	528
Canada	526
Quebec	525
Australia	523
Finland	523
England (UK)	517
New Brunswick	515
Estonia	515
Saskatchewan	515
Nova Scotia	512
France	511
Netherlands	511
Italy	510
Czech Republic	509
Germany	509
United States	508
Belgium	508
Austria	506
Newfoundland & Labrador	504
Manitoba	504
Norway	503
OECD AVERAGE	500
Ireland	498
Denmark	497
Portugal	494
Prince Edward Island	493
Sweden	491
Russian Federation	489
Slovak Republic	483
Poland	481
Spain	477
Slovenia	476
Serbia	473
Croatia	466
Hungary	459
Turkey	454
Israel	454
Chile	448
Cyprus	445
Brazil	428
Malaysia	422
United Arab Emirates	411
Montenegro	407
Uruguay	403
Bulgaria	402

provided a base to calculate participant estimates reflecting the degree to which a nation or economy's problem-solving performance exceeds what might be predicted by their students' mathematics, reading, and science performances. In essence, this proxy difference's measures can be considered as describing the performance of each nation's or economy's students when measured on problem-solving capabilities

Table 4. Correlations among PISA 2012 Problem-Solving scores for the 44 participating entities having scores in each of mathematics, reading, science, and problem solving.

A: OECD total latent correlation between			
Mathematics	Reading	Science	and
0.81	0.75	0.78	Problem solving
	0.85	0.90	Mathematics
		0.88	Reading
B: Canadian latent correlation between			
Mathematics	Reading	Science	and
0.76	0.71	0.75	Problem solving
	0.82	0.87	Mathematics
		0.87	Reading
C: U.S. latent correlation between			
Mathematics	Reading	Science	and
0.86	0.80	0.83	Problem solving
	0.89	0.93	Mathematics
		0.91	Reading

Note. Table A is for all such participants in PISA 2012; Table B is for all such participants from Canada; and Table C is for all such participants from the United States. From OECD, 2014a.

when the contributions of the other core PISA domains are removed from the mix. This approach to viewing the data leads to the next comparison.

Comparing nations' students problem-solving performance relative to students with similar mathematics, reading, and science skills

Figure 4 illustrates the degree to which each of the participating nations' or economies' problem-solving performances exceeded, were equal to, or fell short of these predicted "problem-solving proxy" performance levels. The horizontal axis provides a marker of zero difference and the darker shaded bars indicate participating entities having performances that differ statistically, either larger or smaller, from the average difference in similar nations' or economies' performances predicted by the other three domains.

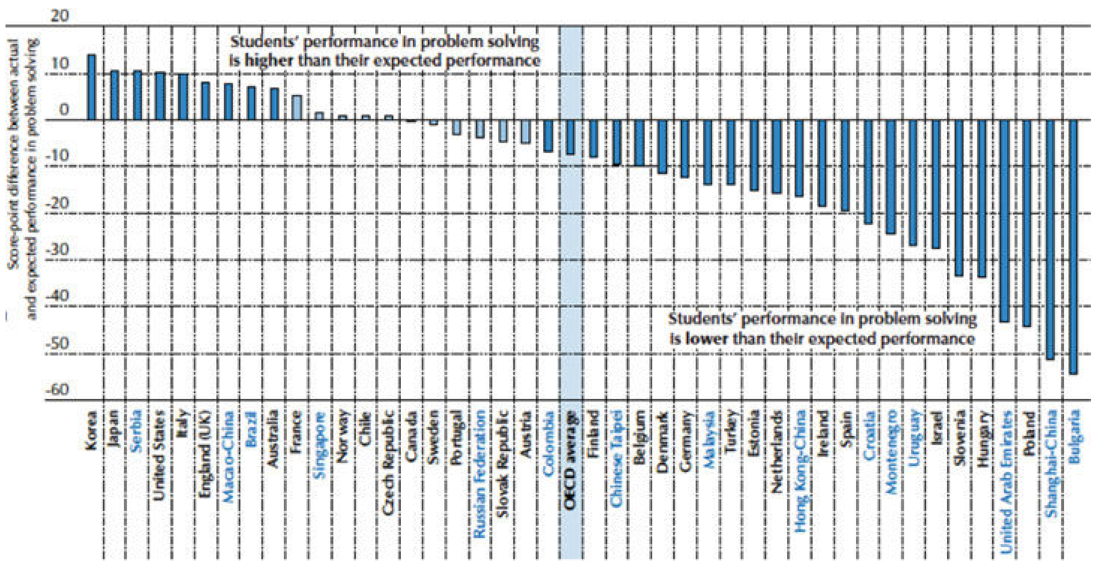

Figure 4. Relative performances in PISA 2012 Problem-Solving for students by country relative to their performances compared to students in other countries having similar skills in mathematics, reading, and science. Graphic taken from *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems* (p. 69), by Organization for Economic Cooperation and Development, 2014a, Paris, France: OECD Publishing, doi:10.1787/9789264208070-en.

The lightly colored bars in the middle of the distribution reflect countries with performances not statistically different from the average difference value as predicted by performances in the other three domains.

Canada's difference score for actual performance minus expected performance in problem solving is −0.2 points with a standard error of 1.9 points. It was judged as not being statistically different from the average difference value. The U.S. score for the same difference comparison is 10.4 points with a standard error of 2.1 points and was judged as being significantly greater than the average expectation of nations or economies having average difference scores. The OECD average difference score comparison was −7.5 points with a standard error of 0.5 points.

A graphical comparison with the total set of participating nations and economies shown in Figure 4 shows that Canada's problem expectation was essentially at the point of having a performance that equaled its expectation score. Even with this finding, Canada still ranked 10th in the overall performance as predicted by the problem-solving difference score. With Canada's difference score of −0.02 points from the prediction based on the other three PISA domains, it appears that problem-solving instruction is almost totally predicted by contributions of the domain areas of mathematics, reading, and science.

Though Canada's score was not significantly different than its expected performance, it was still significantly higher that the OECD average difference score. The United States ranked fourth overall in the difference point comparison with a difference score of 10.4 points and a standard error of 2.1 points. This performance was significantly above both its expected average and the OECD average difference. When one considers all 44 participating entities, the nations and economies whose scores exceed expectation are, in descending order: Korea (14.0), Japan (10.6), Serbia (10.6), the United States (10.4), Italy (9.9), England (8.0), China–Macao (7.7), Brazil (7.1), Australia (6.9), France (5.2), Singapore (1.5), Norway (1.1), Chile (1.0), and the Czech Republic (0.9). Canada (−0.2) was the next country in performance, but its difference did not significantly exceed average of comparable countries relative to the other three domains. However, Canada did exceed the OECD average difference of −7.5 points from the expected predicted score. Singapore, the overall leader in the unadjusted problem-solving scores rankings, was ranked 11th on the adjusted scores scale with a performance of 1.5 points above its expected score with a standard error of 1.0. These differences among the predicted scores and the actual obtained scores suggest differences in the ways in which differing nations' or economies' students approach, process, and describe solutions to the problem tasks in the assessment.

How well do 15-year-olds engage in cognitive processing to understand and solve problems in context?

The foregoing analyses of student performances for participating entities were based on the scores derived from the total set of problem-solving tasks across the assessment. We now shift to examining student work using the dimensions of problem solving built into the development of the problem-solving item set. For example, what differences appear when comparing responses from static and interactive problems? This type of information provides a basis for generalizations about cognitive processing patterns important to the approach to, decision making within, and communication of results of solving PISA-like problems.

Using the odds ratio comparison from logistic regression, students' problem-solving performances can be expressed and compared relative to a baseline of the average OECD performance. A ratio value of 1.00 indicates that a student is performing at the average performance level. A ratio of 1.10 indicates that a student is performing at a level 1.10 times the OECD average level or 10% above the OECD average level. In a like manner, a student with a ratio of 0.85 is performing 15% below the OECD level. Performances below the 1.00 level for a nation or economy indicate that their students' performances fell below the OECD average level with respect to the skills and conceptual structure demanded by the item or test being compared.

Table 5. Distribution of the PISA 2012 Problem-Solving items by the two categories of nature of problem setting and the four cognitive problem-solving processes required.

Nature of problem setting	Cognitive problem-solving process			
	Exploring and Understanding (10 items)	Representing and Formulating (9 items)	Planning and Executing (16 items)	Monitoring and Reflecting (7 items)
Static (15 items)	5	2	6	2
Interactive (27 items)	5	7	10	5

Note. From OECD, 2014a.

Item types

Earlier distinctions were made between static and interactive items, and the following analyses were made on the basis of students' responses to the entire set of items. Now, we shift to look through finer screens at students' performances across the static and interactive item problem situation classifications. Following this analysis, the PISA 2012 Problem-Solving items were reclassified according to which of four cognitive process categories, described in the *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (OECD, 2013; Philpot, Ramalingam, Dossey, & McCrae, forthcoming), played the main role in resolving each problem as posed. These four processes are each described by a dyad, or pair, of gerunds. These descriptors of the cognitive processes central to a solution can assist in linking the curricular importance of findings from the PISA assessment to the teaching and learning of problem solving.

The four descriptive dyads of cognitive processes are *exploring and understanding, representing and formulating, planning and executing*, and *monitoring and reflecting*. These particular cognitive activities have long been associated with problem solving (Funke, 2010; Mayer & Wittrock, 2006; Polya, 1945). The PISA Framework document for the 2012 Problem-Solving assessment can be consulted for a more in-depth description and discussion of the processes. It is available online through the PISA website (OECD, 2013).

Table 5 provides a categorization of the 42 problem-solving tasks by the nature of their problem situation crossed with the most likely cognitive processes required in reaching a solution. In order to assess the degree of independence of these classifications, a Fisher's exact test of independence was performed. The result ($p = .69$) indicates that there is little evidence to suggest that the classifications of problem situation and cognitive processes are not independent for the PISA tasks (Ferguson, 1966). Thus, we continue to examine the nature of the problem-solving situations crossed with the different cognitive processes.

Nature of problem situation analysis

An overall analysis of the problem situation effects on student performance showed distinct differences in participating entities' performances on interactive items relative to static items after effects associated with performances at the participant level and booklet administration were adjusted out. Table 6 provides a listing of the 43 participating entities, Cyprus is omitted as described earlier, having a better than expected performance on one of the natures of problem situation success categories. The left- and right-hand columns contain the countries that excelled in performance on interactive or static items, respectively. Those countries listed in the central column are those whose students did not show a better than expected performance on either item situation type. Note that the right-hand column's odd ratios can be interpreted as the percentage better for static settings when one makes the appropriate subtraction from 1.

The data in Table 6 indicate that U.S. and Canadian students are on their way to meeting the expectations held by current reform efforts that expect students to problem-solve and work with modeling skills in approaching problem-solving situations. Students from both nations showed performances that were significantly closer to higher performances on interactive tasks than on static tests. The data for 15-year-olds indicated that U.S. students (fourth) were about 13% higher than the OECD average, whereas Canadian students (seventh) were about 5% higher. Ireland leads this group with a 1.16 odds ratio, or

Table 6. Analysis of the effect of nature of item situation (static versus interactive) on performance.

Analysis of problem situation effects					
Better than expected performance on interactive items	Odds ratio	No significant difference on item types	Odds ratio	Better than expected performance on static items	Odds ratio
Ireland	1.16	France	1.06	Serbia	0.95
Korea	1.14	Italy	1.04	Croatia	0.94
Brazil	1.13	Spain	1.04	The Netherlands	0.94
The United States	1.13	England (UK)	1.03	Austria	0.93
Portugal	1.07	Czech Republic	1.02	Slovak Republic	0.92
Singapore	1.06	Belgium	1.02	Finland	0.92
Canada	1.05	Australia	1.02	Chinese Taipei	0.92
Japan	1.04	United Arab Emirates	1.02	Shanghai-China	0.92
		Germany	1.02	Denmark	0.91
		Columbia	1.02	Sweden	0.91
		Chile	1.01	Montenegro	0.85
		Hong Kong-China	1.00	Bulgaria	0.82
		Malaysia	0.98		
		Russian Federation	0.98		
		Israel	0.98		
		Uruguay	0.97		
		Estonia	0.97		
		Poland	0.97		
		Turkey	0.96		
		Hungary	0.96		
		Macao-China	0.95		
		Norway	0.94		

16% higher. Korea (second), Singapore (sixth), and Japan (eighth) lent an Asian influence to the top interactive item performers.

At the other end of the continuum were the nations or economies whose performances indicated significantly higher performance on static tests compared to interactive ones. Leading this group was Bulgaria with an 18% more likely odds ratio to perform better on static items. Bulgaria was joined by other Balkan and Baltic nations as high performers on static items.

These results suggest that U.S. and Canadian students were adept at gauging a problem, identifying critical features, and collecting the appropriate information needed for a solution. Such skills do not develop in a vacuum; teachers need to make decisions about presenting interactive problem situations that call for the application of these dynamic approaches to defining and ferreting out the solutions to interactive problems, problems where the solver has to make decisions about acquiring and using appropriate data and molding solution methods fitting the problem situation. Both problem-solving situation analyses indicate that Canadian students may not be getting as many opportunities for such problem solving outside the three core content areas of the curriculum as U. S. students may be.

Nature of cognitive problem-solving process analysis

The 42-item pool of PISA 2012 Problem-Solving tasks contained 16 planning and executing items and seven monitoring and reflecting items. Using an analysis system similar to that employed in examining the differences associated with the interactive and static problem settings, the performances of participating entities across the 2012 PISA Problem-Solving items with reference to the cognitive processes were determined. Here the comparisons are structured on a scale indicating where the performance with respect to each cognitive process category is examined along a continuum relative to expectations. The performance of individual entities was then reported using odds ratios to distinguish the observed performance from higher than expected to lower than expected. The exploring and understanding and the monitoring and reflecting processes had less stark differences in performance across entities, perhaps because the processes making up each of these categories are used throughout the course of developing a problem solution in looking ahead, cycling back and forth through information and relationships, and, finally, reviewing the solution.

Table 7. Country leader rankings for high and low performance on expected cognitive processes usage across the PISA item set.

Cognitive process strengths and weaknesses in problem solving			
Five highest performing countries	Odds ratio	Five lowest performing countries	Odds ratio
	Exploring and Understanding		
Singapore	1.19	*Croatia*	0.79
Norway	1.19	Chile	0.77
Hong Kong-China	1.17	*Montenegro*	0.77
Korea	1.16	*Columbia*	0.77
Australia	1.14	Turkey	0.75
	Representing and Formulating		
Macao-China	1.38	*Croatia*	0.83
Chinese Taipei	1.36	*Montenegro*	0.82
Shanghai-China	1.33	*Uruguay*	0.80
Korea	1.32	*Columbia*	0.74
Singapore	1.23	*Bulgaria*	0.69
	Planning and Executing		
Bulgaria	1.35	*Chinese Taipei*	0.79
Montenegro	1.35	*Shanghai*	0.78
Croatia	1.30	*Hong Kong-China*	0.78
Columbia	1.29	Korea	0.71
Uruguay	1.28	*Singapore*	0.71
	Monitoring and Reflecting		
Columbia	1.29	*Chinese Taipei*	0.87
Chile	1.28	*Macao-China*	0.86
Turkey	1.15	Austria	0.85
Spain	1.15	Norway	0.84
Uruguay	1.15	Denmark	0.82

Note. Countries/Large cities with names in italicized type are non-OECD members. Those countries in regular text font are OECD members.

Table 7 contains a brief summary of the individual analyses by the five highest and five lowest performing nation/economy leaders for each process dyad of cognitive processes, along with their odds ratios.

Two features stand out in examining the listing of the five highest and five lowest nations/economies on performance against expected performance for each cognitive process. The first is relative to the focal groups for this article—Canada and the United States. Neither country was listed in the top or bottom five participating entities listing for any of the four cognitive process dyads.

In exploring and understanding, Canada's odds ratio was 1.02 whereas that of the United States was 1.01. These ratios ranked Canada and the United States in the 20th and 21st positions, respectively, out of the 43 participating entities with rankings. In representing and formulating, Canada's odds ratio was 1.12 and the United States' was 1.02. These ratios ranked them in the 8th and 18th positions, respectively, out of the 43 participating entities. In planning and executing, the odd's ratios of the two North American's countries' odds ratios were switched in order, with the United States at 0.94 and Canada's at 0.92. These performances placed the United States and Canada in rankings 31st and 33rd, respectively. In monitoring and reflecting, the odd's ratios were the United States at 1.08 and Canada at 0.97. These ratios resulted in the United States being ranked 11th and Canada being ranked 34th out of the 43 participating entities ranked.

Note that the process rankings for Canada and the United States changed from the first two process dyads to the last two process dyads. This indicated a higher performance than expected for the Canadian students on exploring and understanding and on representing and formulating. Though U.S. students had the advantage relative to expectations for planning and executing and for monitoring and reflecting, this same shifting of positions again was noted for perhaps less focus outside the three core content domains compared to U.S. students. Perhaps the most evident was in the changes from the lower than expected ratings for representing and formulating to the participating entities rated in the top five countries for planning and executing.

This "flipping" of performances was also observed by the OECD because they noted a "substantial overlap" between participants who were strong on the first two process dyads listed and in the finding

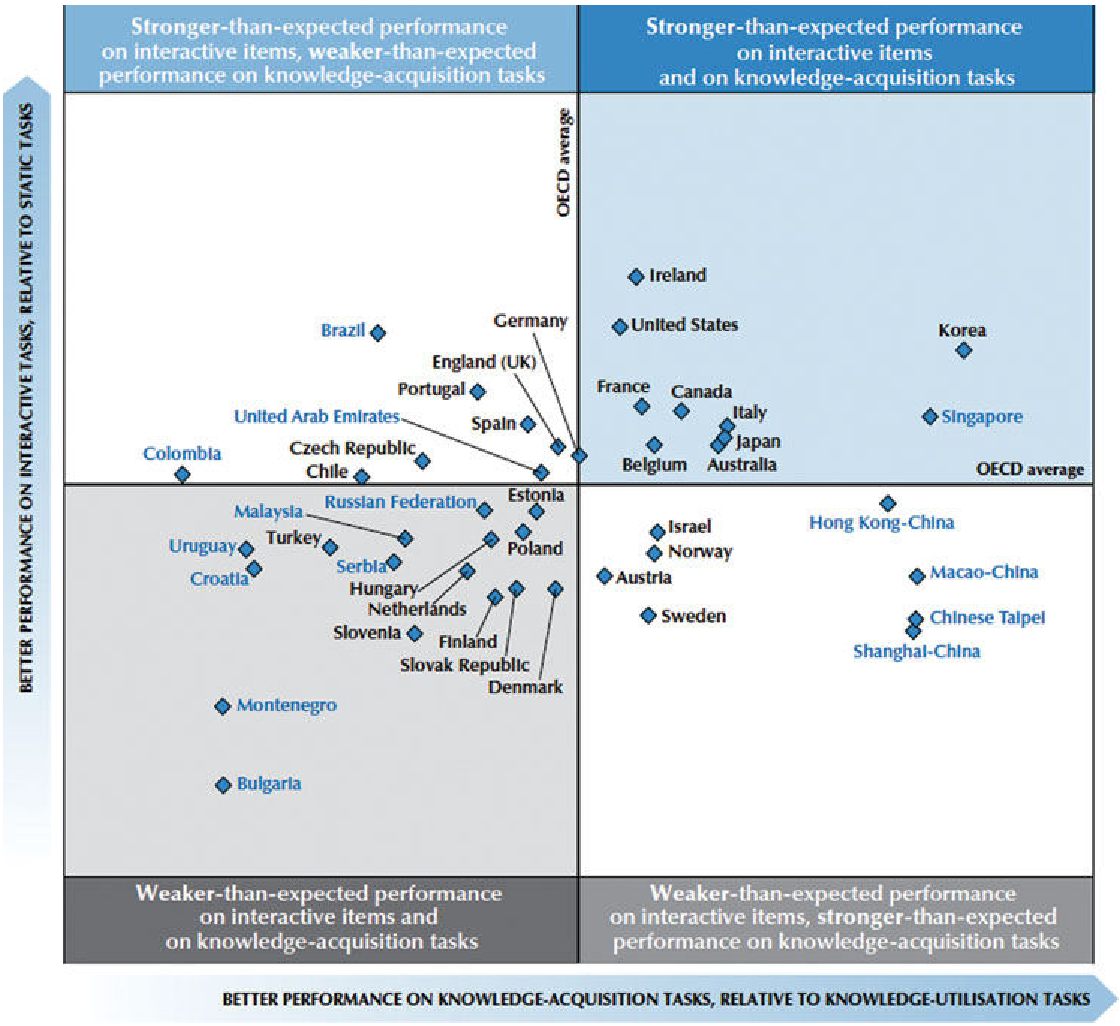

Figure 5. Joint analysis of strengths and weaknesses, by nature of the problem and by process. Graphic taken from *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems* (p. 90), by Organization for Economic Cooperation and Development, 2014a, Paris, France: OECD Publishing, doi:10.1787/9789264208070-en.

that these same entities were also ranked lower in performances on the latter two process dyads. The first two processes are focused on knowledge acquisition and arrangements. The latter two process dyads are focused on knowledge utilization. This overall process based analysis indicates that

> In general, what differentiates high-performing systems, and particularly East Asian education systems, …, from lower-performing ones, is their students' high level of proficiency on "exploring and understanding" and "representing and formulating" tasks. (OECD, 2014a, p. 88)

These first two process dyads relate to the actual cognition and modeling of the problem, whereas the latter two dyads deal with the performing of the analytics laid out in the formulation and the latter communication of the findings and their ramifications.

Figure 5 illustrates the relative positions of the 43 nations/economies ranked, with Cyprus omitted as explained earlier, according to their odds ratios for the relative to the nature of problem setting and cognitive processes odds ratios across the 42 PISA tasks. The figure provides a ranking by the odds ratios for success on interactive items, compared to static items, on the vertical axis and by the odds ratios for success on knowledge acquisition tasks (exploring and understanding or representing and formulating), compared to knowledge utilization tasks (planning and executing) on the horizontal axis. Both scales are presented with logarithmic scales. The odds ratios for monitoring and reflecting are not reflected in this figure, because of their overlap with the previous processes. Further, the patterns in the distribution and the differences involved for the pairing of monitoring and reflecting were not as pronounced as for the linkages on the first three cognitive scales.

An examination of the data in Figure 5 shows that the United States and Canada points are both in the upper right-hand quadrant of the graphic. This indicates a propensity, relative to the entire set of participating entities, for Canadian and U.S. students to be more apt at performing better than expectations on interactive problems than static problems and, at the same time, performing better on items related to cognitive processes than those reflecting knowledge acquisition. However, both countries are closer to the origin than some of the other participating entities in the first quadrant. This perhaps suggests a slightly more balanced position than some of the other entities, such as Korea and Singapore, with respect to an overall balance of the cognitive skills associated with this quadrant of performance. These latter two participant performances tended to flip from the higher to lower performance across the dyads as mentioned above, whereas Canadian and U.S. students had a more balanced level across the process dyads.

Summary

What do the PISA 2012 Problem-Solving findings say to Canadian and U.S. educators?

First, Canadian and U.S. students can problem solve. In particular, when the data are viewed from an average scale score perspective, both countries rank in the upper 40% of the 44 participating entities. From this perspective, Canadian students rank statistically higher than U.S. students in problem solving. Later, when Canadian and U.S. students' expected problem-solving scores are compared with students' actual problem scores, Canadian and U.S. students' difference scores are in the upper quarter of the entities participating in the study. This again shows that students from both countries are achieving well relative to many of their international peers in problem solving.

Second, in the examination of students' performances relative to problem situations, students in both countries placed higher for work on interactive tasks than on tasks situated in static problems. Further, examinations of data from the cognitive dyads most related to student work showed that students from both countries performed well on tasks that called for students to acquire and process new information, again ranking their performances ahead of the students in the majority of the participating entities. Students in other participating entities were more likely to resort to cognitive behaviors related to utilizing known knowledge, rather than carefully examining and understanding the situations at hand.

Third, relative to other participating entities, the Canadian and U.S. performances were consistently above OECD averages. The same can be said about the Canadian provinces' performances. Unfortunately, the United States did not have a state-level PISA study.

Fourth, in comparing the percentage of Canadian and U.S. students reaching the various levels of performance on the PISA Problem-Solving scale to the percentage of students in other participating entities, results show that both Canada and the United States should strive to move students performing at levels below level 1 through level 3 upward to higher levels of performance on problem solving. Movement up the proficiency scale at these levels will open doors for these students to the world of work, to increasing demands for participation in learning in career-related on-the-job education programs, and, eventually, to higher education. The current results show that both nations' students have a leg up on other participating entities by their performances above expectations in problem solving, but they have to capitalize on this with performance throughout their studies and in daily life settings.

ORCID

John A. Dossey http://orcid.org/0000-0002-3410-4831
Joachim Funke http://orcid.org/0000-0001-9129-2659

References

Buchner, A., & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem solving research. *The Quarterly Journal of Experimental Psychology*, *46A*(1), 83–118. doi:10.1080/14640749308401068

Council of Ministers of Education, Canada. (2014). *Assessment matters! Issue 6—How good are Canadian 15-year-olds at solving problems? Further results from PISA 2012*. Toronto, Canada: Author.

Dörner, D. (1997). *The logic of failure. Recognizing and avoiding error in complex situations*. New York, NY: Basic Books.

Ferguson, G. A. (1966). *Statistical analysis in psychology and education* (2nd ed.). New York, NY: McGraw-Hill.

Frensch, P. A., & Funke, J. (Eds.). (1995a). *Complex problem solving: The European perspective*. Hillsdale, NJ: Erlbaum.

Frensch, P. A., & Funke, J. (1995b). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube & K. F. Wender (Eds.), *The cognitive psychology of knowledge: The German Wissenspsychologie project* (pp. 313–330). Amsterdam, the Netherlands: Elsevier. doi:10.1016/S0166-4115(08)62663-1

Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking and Reasoning*, 7(1), 69–89. doi:10.1080/13546780042000046

Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, *11*, 133–142. doi:10.1007/s10339-009-0345-0

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 287–303). Mahwah, NJ: Erlbaum.

Organization for Economic Cooperation and Development. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris, France: OECD Publishing. doi:10.1787/9789264190511-en

Organization for Economic Cooperation and Development. (2014a). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems* (Vol. V). Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/pisa/keyfindings/PISA-2012-results-volume-V.pdf

Organization for Economic Cooperation and Development. (2014b). *PISA 2012 technical report*. Paris, France: OECD Publishing.

Organization for Economic Cooperation and Development. (2014c). *Explore PISA 2012 mathematics, problem solving and financial literacy test questions*. Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/pisa/test

Philpot, R., Ramalingam, D., Dossey, J., & McCrae, B. (forthcoming). Factors that influence the difficulty of problem solving items. In B. Csapó, J. Funke, & A. Schleicher (Eds.), *The nature of problem solving*. Paris, France: OECD Publishing.

Polya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.

Sternberg, R. J., & Frensch, P. A. (Eds.). (1991). *Complex problem solving: Principles and mechanisms*. Hillsdale, NJ: Erlbaum.