Journal of Data Science **10**(2012), 441-464

A Data Mining Approach to Comparing American and Canadian Grade 10 Students' PISA Science Test Performance

Chong Ho Yu[1*], Charles Kaprolet[2], Angel Jannasch-Pennell[3] and Samuel DiGangi[4]
[1]*Azusa Pacific University,* [2]*Scottsdale Unified School District,*
[3]*KOI Education and* [4]*Arizona State University*

Abstract: According to 2006 Programme for International Student Assessment (PISA), sixteen Organization for Economic Cooperation and Development (OECD) countries had scores that were significantly higher than the US. The top three performers were Finland, Canada, and Japan. While Finland and Japan are vastly different from the US in terms of cultures and educational systems, the US and Canada are similar to each other in many aspects, thus their performance gap was investigated. In this study data mining was employed to identify factors regarding access to and use of resources, as well as student views on science for predicting PISA science scores among Grade 10 American and Canadian students. It was found that science enjoyment and frequent use of educational software play important roles in the academic achievement of Canadian students.

Key words: American students, Canadian students, data mining, enjoyment, PISA.

1. Introduction

According to the 2006 Programme for International Student Assessment (PISA) report on science performance, sixteen countries from the Organization for Economic Cooperation and Development (OECD) scored significantly higher than the US. The top three performers were Finland, Canada, and Japan. While Finland and Japan are vastly different from the US in terms of cultural and educational systems, the US and Canada are similar to each other in many aspects, thus their performance gap awaits explanation. In this study data mining was employed to identify factors for predicting PISA science scores among Grade 10 American and Canadian students. These factors include access to and use of resources as well as student views on science. It was found that Canadian high

*Corresponding author.

schools have better resources than their American counterparts and that science enjoyment, the number of books at home, intense interest in science, and frequent use of educational software play important roles in science achievement. In addition, for both American and Canadian students, science enjoyment, the number of books at home, intense interest in science, and frequent use of educational software play important roles in PISA science achievement, but Canadian students reported higher values for the first three variables. These findings imply that investment in computer resources might not be the immediate answer to improvement of science education. Due to the fact that Canadian students reported stronger science enjoyment and interest in science than their American counterparts, it is strongly suggested that American policymakers and educators should study Canadian students' motivation.

Although conventional procedures, such as regression analysis, were employed in this study, the primary tool for this research project is data mining, which is a cluster of techniques aiming to extract useful information and relationships from immense quantities of data (Larose, 2005). Data mining does not start from a strong preconception or a narrow hypothesis; rather it aims to detect patterns that are already present in the data. Following this line of reasoning, Luan (2002) views data mining as an extension of Exploratory Data Analysis (EDA). It is the conviction of the authors that this exploratory approach is suitable to a complex phenomenon endowed with large amount of data, such as student performance in an international context.

The subsequent discussion will be structured as follows. Section 2 is a brief introduction to the background of PISA, such as its assessment methods, and the similarities between the US and Canada in terms of their cultural and educational systems. Section 3 illustrates the research methodologies employed in this study, including t-tests, exact tests, data visualization, and data mining. Section 4 is a discussion of the findings whereas the final section points out the implications for education reform based upon the findings.

2. Background

PISA is a series of assessments, sponsored by OECD, and administered internationally to 15-year-olds, most recently representing 57 different countries (NCES, 2008). Since 2000, PISA has been administered every three years as a method of comparing national performance in science, mathematics, and reading, with "scientific literacy" receiving a heavy focus in 2006. The criteria for scientific literacy developed by PISA (OECD, 2003) are based on Bybee's (1997) model. According to Bybee, there are four levels of scientific literacy, namely, nominal, functional, conceptual/procedural, and multi-dimensional. At the nominal level, the learners cannot go beyond recalling terms and names as a result of rote learning. At the functional level, the learners are able to apply scientific knowledge in limited situations, but fail to make generalizations to broader contexts. PISA does not accept the preceding two as satisfactory outcomes of science education. The top level of Bybee's model requires insightful understanding of science, its historical background and its role in culture. In PISA's view, this advanced level is suitable to a few intellectual elites, but not to all citizens. To most educated citizens, the desirable level is conceptual and procedural scientific literacy, which is defined as the capacity of "using scientific knowledge to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world" (OECD, 2003, p.133). One of the goals of achieving a higher level of scientific literacy is pragmatic. OECD countries realize that today most manual or routine cognitive tasks can be performed by either cheaper laborers or computers, and thus it is expected that these types of jobs will continue to migrate from OECD countries to developing nations. In order for OECD citizens to fully participate in globalization, graduates must possess advanced problem-solving skills that go beyond following rules, and communication skills for illustrating complex scientific ideas in a user-friendly fashion. Based on this educational philosophy, PISA items are designed to test students' ability to apply science into various contexts rather than recalling names and terms (OECD, 2007).

The scientific literary assessment was broken down into three subscales: Identifying scientific issues, explaining phenomena scientifically, and using scientific evidence. PISA scores are measured on a scale ranging from zero to 1000, with a mean of 500 and a standard deviation of 100. In PISA 2006, about two-thirds of students scored between 400 and 600 points. On the exam, each student was awarded a score based on the item difficulty that matches the student's ability. In other words, a score can describe both student performance and item difficulty. For instance, a student with a score of 650 can be expected to complete a question with a difficulty rating of 650 (OECD, 2007). Although it is a psychometrically sound strategy to express student performance and item difficulty on a common scale, PISA adopts different scaling units than those used in conventional Item Response Theory (IRT) or Rasch modeling, in which the average is zero.[1]

Results from the 2006 PISA showed that the average score of science literacy for students in the United States was 489, lower than the mean of 500. US students scored lower than their peers in 22 other countries, 16 of which are fellow OECD jurisdictions (such as Canada, the United Kingdom, and Japan). By subscale, US students scored lower than average on "explaining phenomena

[1]Yu, C. H. and Osborn-Popp, S. (2005). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment Research and Evaluation* **10**, 1-19. Retrieved from http://pareonline.net/pdf/v10n4.pdf

scientifically" and "using scientific evidence".[2] Further, US students received an average score of 474 in mathematics literacy, lower than the OECD average of 498. Higher scores were obtained by 31 other countries, including Canada (527), the United Kingdom (495), France (496), and Japan (523) (NCES).

Some critics question whether it is a valid assertion that the US education system is failing on the ground that it is not meaningful to compare a highly populated and ethnically diverse nation like the United States with small and homogeneous city-states like Hong Kong and Singapore (Bracey, 2009). First, a study by OECD (2007) concluded that there is no significant relationship between performance in PISA and the size of the countries/regions. In other words, smaller regions or countries like Hong Kong, Macao, and Singapore do not have advantages due to their size. Nevertheless, in order to make a fair comparison, this study focused on countries that are similar in size, culture, and diversity. PISA (OECD, 2007) indicated that in terms of scientific literacy US performance was lower than many culturally similar countries, such as Canada, the United Kingdom, Australia, and New Zealand. Comparatively, culturally similar countries such as Canada and the United Kingdom received average scores of 534 and 515, respectively. In New Zealand, 3.9 percent of students reached the top level of scientific literacy, which is three times the OECD average. In the United Kingdom, Australia, and Canada, between 2 and 3 percent of their students reached the highest level. In the US, the figure was 1.5 percent.

Canada, the third highest scoring country on the PISA, is an interesting comparison due to its similarities and geographic proximity to the US. While the US population (over 300 million) is approximately 10 times that of Canada (over 30 million), data from Canada's 2006 census suggest its population is growing at a rate similar to that of the US. Culturally, both countries are predominantly Christian and English-speaking, and are each other's largest trading partners. In addition, the US and Canada are similar in their free market economic systems, patterns of production, and high standards of living.[3]

Demographically, both countries are predominantly Caucasian (approximately 80 percent), yet both of them have diverse minority populations. In the US, the

[2]National Center for Education Statistics (NCES). (2003). *Projections of Education Statistics to 2013.* Retrieved July 30, 2009, from http://nces.ed.eov/progratWproiections/index.asp.

[3]U.S. Census Bureau. (2009b). U.S. POPClock projection. Retrieved July 13, 2009, from http://www.census.gov/population/www/popclockus.html; Beauchesne, E. (2007, March 13). We are 31,612,897 [Electronic version]. National Post. Retrieved from http://www.nationalpost.com/story.html?id=73b94aac-08f0-477f-a72a-b8b640f6658f&k=90795; U.S. Census Bureau. (2007). Section 1: Population. Retrieved from http://www.census.gov/prod/2007pubs/08statab/pop.pdf; Statistics Canada. (2005). Population by mother tongue, by province and territory. Retrieved from http://www40.statcan.ca/l01/cst01/demo11a.htm; Central Intelligence Agency. (2009). The world factbook. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/CA.html

Hispanic and African American populations are the largest minority groups (at 15.4 and 12.8 percent, respectively) and represent a much larger proportion of the total population than any individual minority populations in Canada.[4] In Canada, non-Caucasian ethnic groups are generally equally represented, all ranging individually from 2.5 to 4 percent of the population. Both countries have undergone demographic changes over the last century, and educational policies have also changed to reflect the growing ethnic populations (Bibby, 1987; Breton, 1987; Day, 2000; Driedger and Halli, 2000; Gans, 2004; Glazer and Moynihan, 1970; Moodley, 1986; Skerrett, 2008).

Educationally, the US and Canada have similar systems, as well. Like the US, Canadian public education is funded by federal, provincial, and local governments, and each province has jurisdiction over its academic curriculum (Shaker, 2009; NCES, 2007). Structurally, the Canadian education system is set up similarly to the US, with students typically attending Elementary, Secondary, and Post-secondary schools. Education is compulsory until age 16, with the exception of Ontario and New Brunswick, where the compulsory age is 18 (Shaker, 2009; NCES, 2007).

While the governments of both countries fund public education, their roles in education have become increasingly controversial. This is particularly true in the United States, as concerns over the effectiveness of public education have arisen, but comparatively, a larger percentage of Canadian children attend public schools than their American counterparts, at 94 and 89 percent, respectively (OECD, 2003; NCES, 2003). Private and charter schools are available in both countries, and despite the criticism levied against the public school systems in both countries, enrollment in these schools represents only a minority of all children.

In terms of educational attainment, Canadian adults are more likely to complete higher education than their American peers. As of 2004, 45 percent of Canadians ages 25 to 64 had completed higher education, compared to 39 percent for the US (OECD, 2006a). When looking at high school completion rates for the same populations, only 40 percent of Canadians ended their education with high school, with 49 percent of Americans stopping with high school (OECD). The differences are greater when looking at a younger subset of the populations, with 53 percent of Canadians ages 25 to 34 completing higher education, compared to 39 percent for Americans (OECD).

[4] U.S. Census Bureau. (2009a). National characteristics: National sex, race, and Hispanic origin. Retrieved from http://www.census.gov/popest/national/asrh/NC-EST2008-srh.html; Statistics Canada. (2008). Ethnic origins, 2006 counts, for Canada, provinces and territories. Retrieved from http://www12.statcan.gc.ca/english/census06/data/highlights/ethnic/pages/Page.cfm?Lang=E&Geo=PR&Code=01&Data=Count&Table=2&StartRec=1&Sort=3&Display=All&CSDFilter=5000

Interestingly, despite the documented performance gaps and lower rates of college completion for the US, they spend more money on education than all other G8 countries. Education spending in the US in 2003 totaled seven percent of the national gross domestic product (GDP), compared to 5.2 percent for Canada.[5] However, while the US outspends other countries on education in general, these statistics are skewed by the US' emphasis on higher education, leaving spending rates for primary and secondary schooling lower than countries such as France and the United Kingdom (4.1 percent of the GDP for the US, compared to 4.2 and 4.6 for France and the United Kingdom, respectively) (OECD).

While many similarities exist between the US and Canada, both culturally and educationally, there exists a large gap in performance on international academic assessments. The goal of this study was to use data from the 2006 PISA to compare access to educational resources and views on science in relation to performance on the PISA science assessment.

3. Method

3.1 Data Source

This study utilized archival data downloadable from the PISA website.[6] One common criticism against the validity of PISA is that PISA recruited 15-year-old students, but different nations start formal education at different ages. As a result, PISA may be comparing apples with oranges (Bracey, 2009). It is true that PISA is age-based, and in this data set the grades of participants range from 9 to 11 while the vast majority is Grade 10 students. In order to perform a fair comparison, only Grade 10 observations were retained for this analysis. Hence, observations of 18,588 Grade 10 Canadian students and 4013 American students, as well as 896 Canadian schools 166 American schools, were extracted from the PISA datasets.

It is important to note that PISA employed a two-stage stratified sampling design. At the first stage schools were purposefully selected based upon a number of factors, such as location (state/province) and type (private or public). At the second stage students were randomly chosen from each of the sampled schools. In PISA 2006, there was a wide variation of sample sizes, ranging from 3,789 students in Iceland to more than 30,000 students in Mexico. Countries with a large population were sampled both at national and regional/state levels. The United

[5]UNESCO Institute for Statistics. (2008). Time series data: Total enrolment, school life expec-tancy and expenditure on education. Retrieved from http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx

[6]OECD. (2006b). Database: PISA 2006. Retrieved from http://pisa2006.acer.edu.au/downloads.php

States was an exception. In PISA 2006 only country-level results were provided by the United States (Xie, 2009, personal communication). In Canada, the sample size was considerably larger that than of the US because it needed to provide detailed information at the provincial level (OECD, 2009b). Additionally, the response rate of the US schools has been consistently low for many years (about 40 percent). American school administrators cited various reasons for declining participation in PISA. Common reported reasons include increased testing requirements at the national, state, and local levels, concerns about the timing of the PISA assessment, and loss of learning time. As a remedy, for 2006 data collection the PISA consortium employed a new sampling strategy by rescheduling the exam from September to November, which is near the end of the semester. In PISA 2006, the US response rate significantly improved to 68.95 percent, but it is still lower than that of Canada (83.2%) (NCES, 2008; OECD, 2009b).

3.2 Procedures and Instruments

For PISA tests, all students take pencil-and-paper tests within a two-hour time limit. Each student receives a different combination of items drawn from an item bank. The questions include both multiple-choice and open-ended items. These items appear in the form of testlet, also known as item parcel, in which several items are grouped together based on a common passage related to a real-life scenario. Each participated country is required to administer the test by following the same protocols so that students receive identical information before and during the test. In addition to cognitive items, students are required to answer a questionnaire, which takes 20-30 minutes to complete, providing background information about themselves and their homes. On the other hand, school principals take a 20-minute questionnaire about their schools (PISA, 2009b).

Four of the five PISA instruments were used for this study, namely, the cognitive item test, the school questionnaire, the student questionnaire, and the ICT familiarity component for students questionnaire. From the cognitive item test, a subset of 102 science-related item responses was selected. As mentioned before, although PISA uses a common scale for both test performance and item difficulty by centering both to 500, this scaling is different from the convention of item response theory or Rasch modeling. To make the results more interpretable, a one-parameter logistic IRT model was run to estimate the ability of examinees. This ability estimate, in which the average is set to zero, was treated as the dependent measure in the subsequent analysis.

A subset of items from the school survey relevant to school resources, especially technology resources, was extracted for this study. The items were classified into two groups: School resources and perception of shortage of resources (Table 1). It might be misleading if the raw numbers of various resources are used, because the seemingly abundant resources will be thinly spread in a large school. As a remedy, two new variables named "Instructional computer to student ratio" (ICSR) and "Computers with Web to student ratio" (CWSR) were created by dividing the number of computers for instructional purposes at school and computers with Internet connection by the total enrollment, respectively.

Table 1: A subset of items from the school questionnaire

School resources
About how many computers are available for instruction? About how many computers in the school are connected to the Internet/World Wide Web?
Perception of shortage
Is your school's capacity to provide instruction hindered by any of the following? Shortage or inadequacy of science laboratory equipment Shortage or inadequacy of instructional materials (e.g., textbooks) Shortage or inadequacy of computers for instruction Lack or inadequacy of Internet connectivity Shortage or inadequacy of computer software for instruction Shortage or inadequacy of audio-visual resources

Similarly, a subset of items was selected from the student survey, as shown in Table 2. The data type of responses to "which of the following are in your home" is binary, in which the answer is either "possessing it" or not. For those "how many" items, the data type is ordinal. OECD (2007) believes that an individual's scientific literacy is influenced by student attitudes and their awareness of the life opportunities open by possessing science competencies. In the section "views on science", the first five questions were combined into a construct named "science enjoyment" whereas the second set of "how much do you agree ⋯" questions were loaded into a construct called "science value". Similarly, the subset "how much interest in ⋯" was collapsed into a factor entitled "science interest". "Science enjoyment" is an indirect indicator of tendencies to engage with the material and continued investment in scientific endeavors. "Science value" is a measure of their perceived importance of scientific and technological advances on nearly everyone's life. "Science interest" is similar to "science enjoyment", but the focus is on specific disciplines (e.g., physics, chemistry) instead of their implications for society. These items all utilize a four-point Likert-scale, ranging from "strongly agree" (1) to "strongly disagree" (4). All three scales have high standardized Cronbach coefficient Alpha (Science enjoyment = 0.935250, Science value = 0.884392, Science interest = 0.843107) and factor analysis with scree plots indicates unidimensionality. Hence, their composite scores instead of individual item scores were used in the analysis.

Table 2: A subset of items from student questionnaire

Home Resources
Which of the following are in your home?
Computer for school work
Educational software
Link to Internet
Own calculator
Books to help with school work
Dictionary
DVD/VCR
How many
cell phones at home
TVs at home
computers at home
books at home
View on Science
Science enjoyment: How much do you agree?
Fun when learning about science
Like reading about science
Happy doing science problems
Enjoy acquiring science knowledge
Interested in learning science
Science value: How much do you agree?
Science improves living
Science helps to understand the world
Science helps relate to others
Science helps the economy
Use science as adult
Science is valuable to society
Science is relevant to me
Science helps me understand surroundings
Science brings social benefits
Opportunities to use science after school
Science interest: How much Interest in
Physics
Chemistry
plant biology
human biology
Astronomy
Geology
experimental design
scientific explanations

A subset of questions from the ICT familiarity component for students questionnaire, which is about how often students use computer resources, was selected for this study. For each item there are five response categories: 1. Almost every day, 2. Once or twice a week, 3. Few times a month, 4. Once a month or less, and 5 Never (Table 3).

Table 3: A subset of items from ICT familiarity component for students questionnaire

How often do you use computers for the following reasons?
Browse Internet for information
Play games
Write documents
Use Internet to collaboratev
Use spreadsheets
Download software
Draw, paint, or use graphics programs
Use educational software
Download music
Writing computer programs
For online communication

3.3 Analysis

While conventional procedures were employed in this study, data mining was the primary analytical tool for this project. Using data mining is appropriate to this study. First, using an extremely large sample size will cause the statistical power for any parametric procedures to be 100 percent. On the contrary, data mining techniques are specifically designed for large datasets without being threatened by Type I error. Second, conventional procedures usually require parametric assumptions, such as multivariate normality and homogeneity of variance, but these parametric assumptions are often violated. Data mining procedures are non-parametric in essence and thus do not rely on these assumptions. Third, the elements in this dataset represent multiple data types, including binary, ordinal, and interval scales. Data mining can handle this kind of complexity of data types in one single analysis without any data transformation. Moreover, data mining techniques, such as recursive partition trees, are robust against outliers and missing values (Fielding, 2007; Streifer and Shumann, 2005). Together, the preceding advantages led to the selection of data mining techniques for this analysis. In the following paragraphs the role of conventional and data mining procedures used in the analysis will be explained.

To compare actual resources and perceived lack of resources between Canadian and American high schools, data visualization, t-tests and nonparametric

(Wilcoxon) exact tests were employed. While t-tests were used for examining the actual resources in terms of computer to student ratio, nonparametric tests were conducted for comparing perceived shortage of resources between the two samples. The scale of measuring the perception is ordinal in nature. Since each item was analyzed individually rather than being collapsed into a single inventory, nonparametric tests are the best match for the data. For multivariate analysis that included nominal, ordinal, and interval measures, data mining procedures such as recursive partition trees, also known as classification trees were employed.

All statistical procedures regarding this research study were performed in SAS 9.2 (SAS Institute, 2009), JMP 8 (SAS Institute, 2008) and Spotfire Miner 8.1 (TIBCO, 2009). It is noteworthy that different software packages implement classification trees differently. JMP 8 accepts a dependent variable as interval, ordinal, categorical, and even binary, but Spotfire Miner allows a binary variable only. On the other hand, Spotfire Miner enables the researcher to compare different models side by side, such as classification trees and logistic regression, but JMP 8 does not have this capability.

When the dependent measure is a continuous scale, JMP's partition tree selects the cutoff for partitioning based on its built-in algorithms. However, this cutoff might not be logical or meaningful. As mentioned before, item response theory modeling provides psychometricians with a logical cutoff by centering the scores to zero, which means that all ability estimates above zero are considered high performance whereas all ability estimates below or equal to zero are regarded as weaker performance. Following the IRT approach, a new variable called theta classification was created, in which "1" denotes high performance while "0" represents weaker performance. Using the ability estimates as a continuous scale could yield better precision, but transforming the variable into a binary one could help in interpretation. Thus, after performing the partition in JMP based on the continuous scale, a second classification tree using the dependent measure as a dichotomous variable was run in SpotFire Miner (TIBCO, 2009). For triangulation purposes, Spotfire's classification tree was compared against a logistic regression model. Classification agreement and receiver operating characteristic (ROC) curves were used to determine the predictive power of these models. Classification agreement is concerned with the percentage of matching between the predicted and the observed data points, whereas ROC is a graphical plot of the sensitivity (true positive rate) vs. $1 -$ specificity (false positive rate).

Last, after important predictors were identified, both t-tests and nonparametric tests were utilized to investigate the differences between US and Canadian students in terms of those relevant variables. Not only did this strategy narrow down the scope of comparison, but also reduced Type I error rate by not testing all variables.

4. Results

4.1 Actual Resources and Perceived Shortages

Some Canadian schools have instructional computer to student ratios (ICSR) of 1 or above, while most American schools cannot afford such a ratio. However, when school type is taken into account, a different story emerges. All Canadian schools that have ICSR above 1 are government-dependent schools (Figure 1). The same pattern could be observed in Computers with Web to student ratio (CWSR) (Figure 2). T-tests confirm that Canadian schools have significantly higher ICSR than Americans ($t(871) = -2.80$, $p = 0.0053$), as well as CWSR ($t(907) = 1.98$, $p = 0.0476$). However, the strength of correlation between ICSR and PISA science performance measured by ability estimate is insignificant ($r = 0.00371$, $p = 0.5909$), as is the pair of CWSR and ability estimate ($r = 0.00085$, $p = 0.9036$).

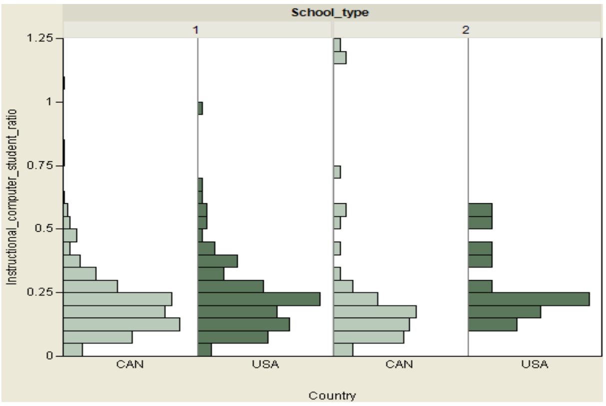

Figure 1: School type and instructional computer to student ratio by country (1 = Private independent, 2 = Government dependent)

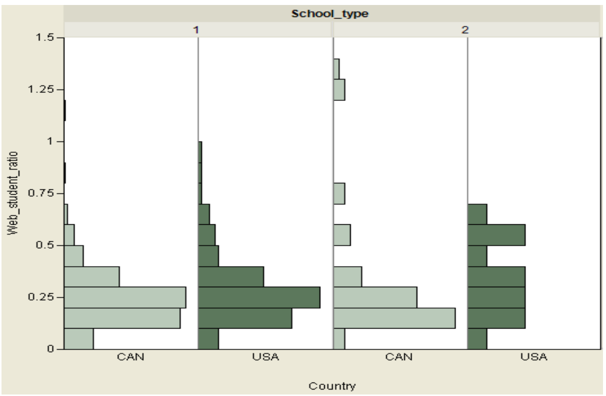

Figure 2: School type and computer with Web access to student ratio by country (1 = Private independent, 2 = Government dependent)

Nonparametric exact tests indicated that while answering the question regarding shortages of technological resources, there are some significant differences between Americans and Canadians in three out of six questions, as indicated in Table 4.

Table 4: Wilcoxon two-sample test

	n (Canada)	n (American)	t-approximation	Exact p value
Shortage of science lab equipment	855	161	0.6465	0.6460
Shortage of instructional materials	857	162	$^* < 0.0001$	*0.00001609
Shortage of computer for instruction	850	162	0.1053	0.1052
Lack of Internet connectivity	852	161	0.9580	0.9577
Shortage of computer software	853	162	*0.0027	*0.0025
Shortage of audio visual resources	855	162	*0.0002	*0.000217

*Significant at 0.05 level, two-sided test

With regard to perceived shortages of instructional materials, the obvious differences between Americans and Canadians could be found in Category 3 and 4 in relations to ICSR. While just a few Canadian schools that have a high ICRS expressed a lot of concerns (Category 4) with lack of instructional materials, more Americans have this concern. Also, more Americans that have a high ICSR rated their concern as "to some extent" (Category 3) than their Canadian counterparts (see Figure 3).

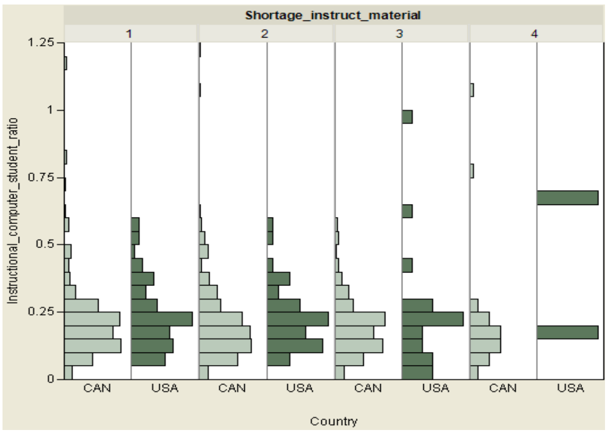

Figure 3: American and Canadian views to shortage of instructional materials (Shortage of instructional materials: 1 = Not at all, 2 = Very little, 3 = To some extent, 4 = A lot)

Pertaining to perceived shortage of computer software, major differences are found in Category 2 and 1 (Very little, Not at all) in relation to ICSR. With regard to perceived shortage of audio visual resources, again, the substantive difference is seen in Category 2 and 1 (Figures 4 and 5).

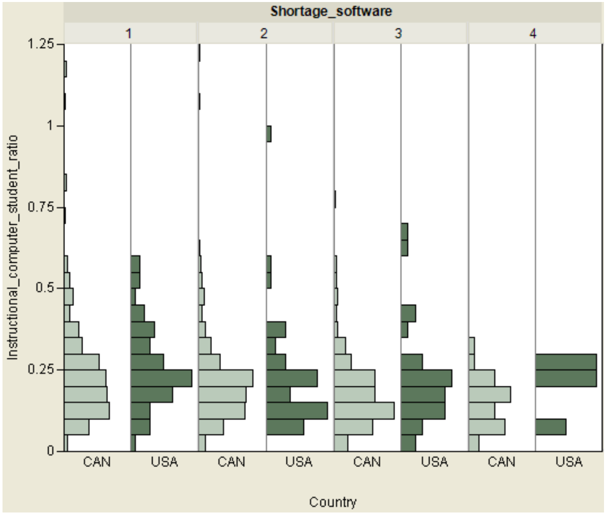

Figure 4: American and Canadian views to shortage of computer software (Shortage of computer software: 1 = Not at all, 2 = Very little, 3 = To some extent, 4 = A lot)

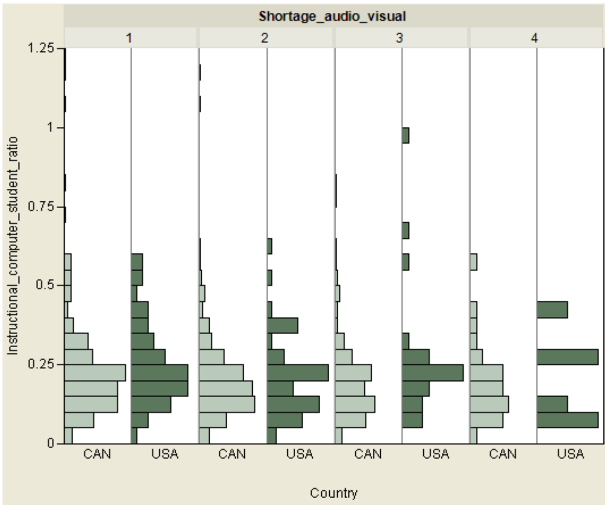

Figure 5: American and Canadian views to shortage of audio-visual resources (Shortage of audio visual resources: 1 = Not at all, 2 = Very little, 3 = To some extent, 4 = A lot)

4.2 Variables Predicting PISA Science Performance in Terms of Ability Estimate

Canadian students (mean = 0.084, SD = 0.72) outperform their American counterparts (mean = -0.1026, SD = 0.79) in terms of ability estimate, and the Satterthwaite t-test for unequal variances shows that the difference is statistically significant ($t(5537.5) = 13.79$, $p < 0.0001$). Figure 6 displays the spread of the two sets of scores. Based upon the rule of 1.5 * interquartile range, which is commonly used in boxplots, 111 extreme cases are identified in the Canadian data sets. After these outliers are removed, the difference between the American and Canadian ability estimates remains significant ($t(8275) = 13.21$, $p < 0.0001$).

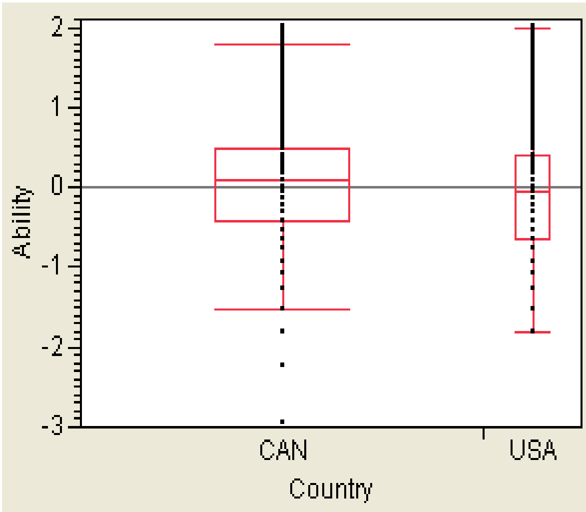

Figure 6: Boxplots of US and Canadian ability estimates

When all observations were included in the classification tree, it indicated that the most crucial factor in determining PISA science test scores was science enjoyment (Figure 7). Students who tend to enjoy learning science (indicated by smaller numbers, < 13) have an average ability estimate of 0.15, whereas the students who tend to have less enjoyment in learning science ($>= 13$) have a mean ability estimate of -0.14. For the group that enjoys science more, the best predictor of their performance is the number of books at home. From this study, possessing more books at home leads to better performance in PISA (1 = 0-10 books, 2 = 11-25 books, 3 = 26-100 books, 4 = 101-200 books, 5 = 201-500 books, 6 = More than 500 books). While it is not difficult to conjecture a causal link among science enjoyment, books at home, and science test scores, it is puzzling to see that the third important variable is the frequency of downloading music (1 = Almost every day, 2 = Once or twice a week, 3 = Few times a month, 4 = Once a month or less, 5 = Never). It is strange that students who download music once a month or less are more likely to earn more points on the PISA science test while all others are put into the lower performing group. Nonetheless, this partition could be result from pure chance fluctuations.

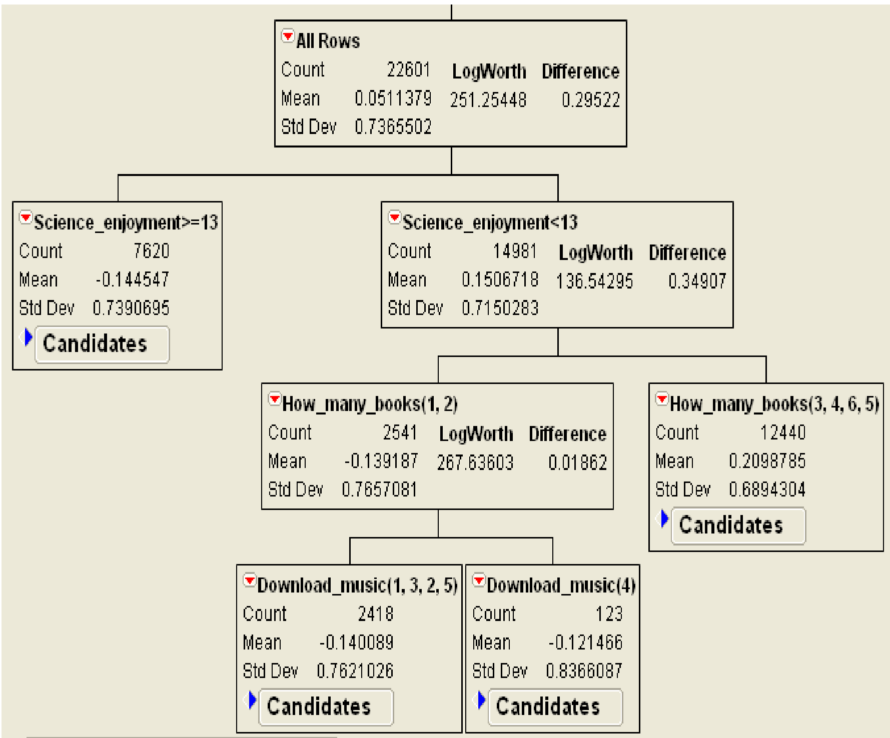

Figure 7: Partition tree using ability estimate

4.3 Variables Predicting PISA Science Performance in Terms of Theta Classification

After the ability estimate was transformed into a binary variable, it was found that 58.8 percent of Canadian students were considered high performers while 48.17% of US students were in this category. The Chi-square test shows a significant difference ($X = 148.711$, $p < 0.0001$) and the Fisher's exact test confirms this finding ($p = 7.140\text{E}-34$). Spotfire's classification tree was run with the theta classification. In Figure 8, the lighter portion of each rectangle depicts high performers while the darker portion signifies weaker performers. The classification tree identified science enjoyment, the number of books at home, frequent use of educational software, and science interest as the most important predictor to performance. This model is considered optimal because when the tree grows by further partitioning, these four variables keep recurring, as shown in Figure 9. In other words, increasing complexity does not yield additionally useful information. Thus, a trimmed tree with four variables was adopted.

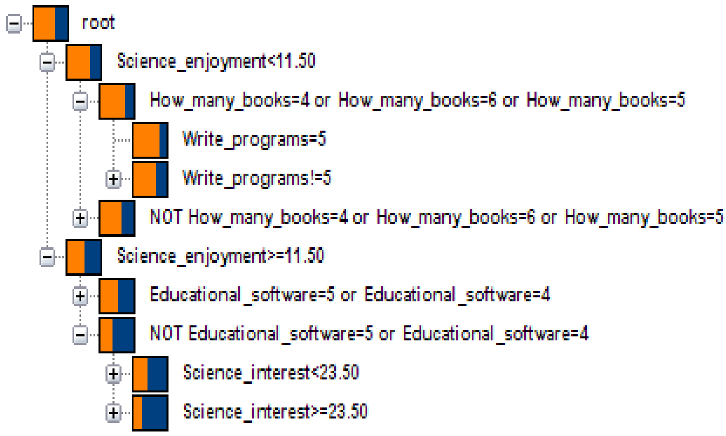

Figure 8: Partition tree using theta classification (trimmed)

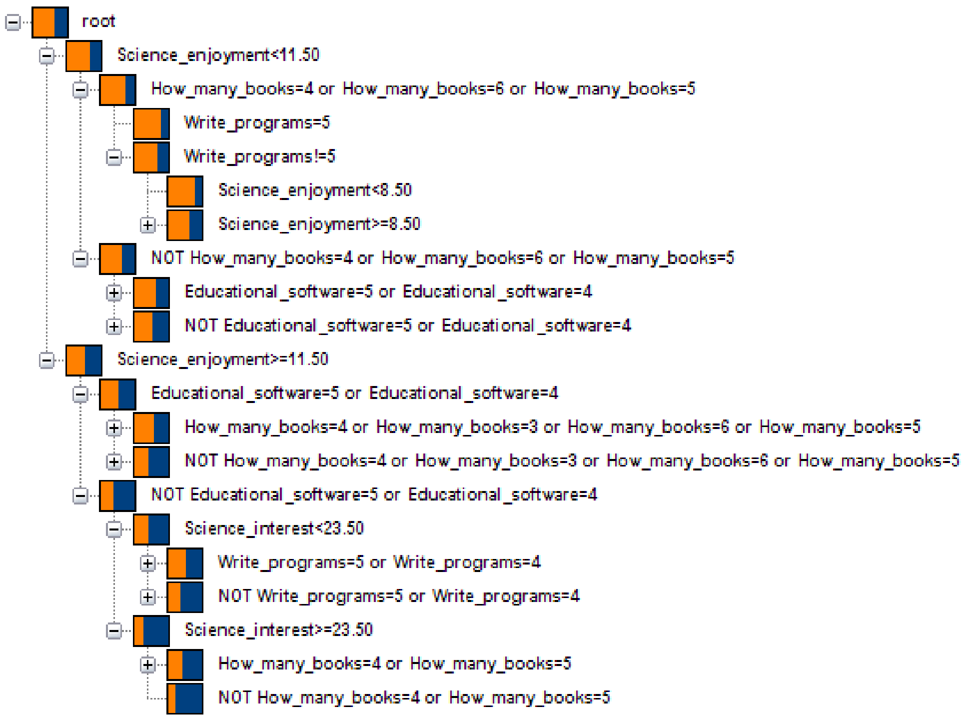

Figure 9: Partition tree using theta classification (untrimmed)

A logistic regression model was run side by side with the preceding classification tree. Unlike its classification tree counterpart, the logistic regression model suggests a longer list of important predictors, as shown in Table 5. Even if the list is filtered by using the criterion of the odds ratio above or equal to 1, it still retains seven variables. Nonetheless, it is the conviction of the authors that a simpler model tends to be more useful for guiding decision-making and future research.

Table 5: Important predictors to PISA performance according to logistic regression

Variable	Odd ratio	Wald Statistic	DF	Pr
Science enjoyment	0.907	334.60	1	< 0.0001
How many books at home	1.182	149.90	5	< 0.0001
Frequent use of educational software	1.209	74.11	4	< 0.0001
Frequent use of computers for writing documents	0.884	57.03	4	< 0.0001
Frequent use of computers for writing programs	1.110	49.41	4	< 0.0001
Frequent use of computers for downloading music	1.088	26.07	4	< 0.0001
How many TV	0.873	23.46	3	< 0.0001
Frequent use of spreadsheets	1.068	21.99	4	< 0.0001
Frequent use of computers for playing games	0.962	19.04	4	< 0.0001
How many computers at home	1.103	18.73	3	< 0.0001
Frequent use of graphics programs	0.976	12.08	4	0.02
Frequent use of computers for online communication	0.931	10.71	4	0.03
Frequent use of computers for collaborating on the Internet	1.016	10.15	4	0.04

The predictive power of the two approaches was evaluated by both classification agreement and ROC curves. Table 6 indicates that although the logistic regression model outperforms the classification tree in predicting high performers (1), their positions are reversed in predicting weaker performers. However, the overall predictive power of the classification tree is stronger than the logistic regression model (67.1 percent vs. 65.6 percent). This assessment is bolstered by the overlaid ROC curves. ROC curves illustrate sensitivity (true positive rate) and $1 -$ specificity (false positive rate). The ideal prediction outcomes are 100 percent sensitivity (all true positives are found) and 100 percent specificity (no false positives are found). In the chart, the 45 degree diagonal gray line represents the baseline. When there is no modeling, the probability is 0.5. Thus, a good classifier should depict a ROC curve leaning towards the upper left of the graph. Figure 10 shows that in most cases the classification tree, shown by a lighter line, is superior to the logistic regression, presented by a darker line. No matter whether simplicity, classification agreement, or ROC curves was used as the criterion for determining the model choice, it is obvious that the classification

Table 6: Classification agreement between the predicted and observed for all students

	Predicted and observe matched (1)	Predicted and observe matched (0)	Overall
Classification tree	83.4%	42.9%	67.1%
Logistic regression	83.8%	38.7%	65.6%

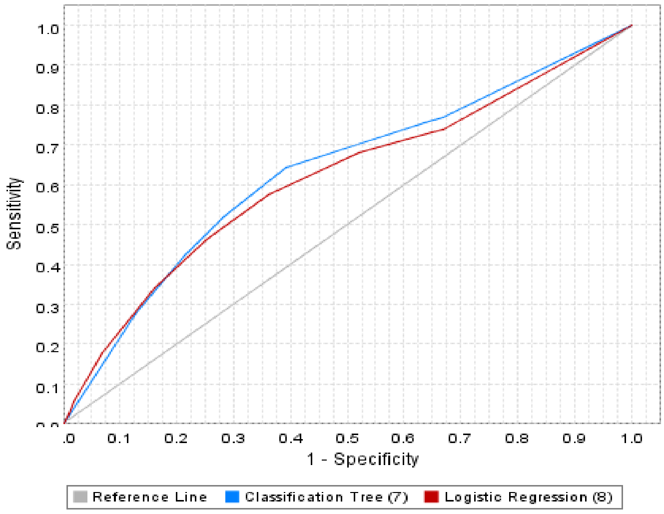

Figure 10: ROC comparing classification tree and logistic regression for all observations

tree approach is more advantageous than logistic regression modeling. Henceforth, in the subsequent analysis only the results of classification trees were discussed.

In both "science enjoyment" and "science interest", outliers were present among the US students. After outliers were removed, the Satterthwaite t-tests for unequal variances indicated that Canadian students tend to enjoy science more than their American peers ($t(5871.4) = -12.19$, $p < 0.0001$). On average, Canadian students also have more intense interests than American students ($t(5666.3) = -6.96$, $p < 0.0001$). Pertaining to frequent use of educational software, there are too many missing in the American group and thus no meaningful comparison could be performed, but the Wilcoxon Rank Sums test for the variable "How many books at home" (ordinal) shows that Canadian students have more print sources than Americans ($p < 0.0001$).

5. Discussion

In summary, Canadian high schools appear to have slightly more reported resources than American schools in terms of "instructional computer to student ratio" and "computer with Web to student ratio". However, there is no significant correlation between PISA science scores and the computer-student ratio. In addition, the perception of shortages of school resources by Canadians and that of their American counterparts substantively differ from each other with regard to shortages of instructional materials, educational software, and audio-visual resources. Nevertheless, the perception of shortages of resources might not necessarily equate to actual shortages. Availability of resources does not imply use of those resources and it is possible that students in schools with high computer-student ratios do not end up using those resources during the day, resulting in a perception that there are not enough of those resources available. The perception differences await further research.

With regard to the crucial factors contributing to the performance gap, namely, science enjoyment, science value, books at home, and use of educational software, the findings in this study partly concur with prior research. Using all PISA 2006 data across 57 countries, Perez and Cromley (2009) found that the direct effect of enjoyment on science scores was moderate. Nonetheless, the association between science scores and science enjoyment was strongest in higher performing, wealthier countries. At first glance, Perez and Cromley's (2009) finding seems to support Bracey's (2009) assertion that PISA tests favor affluent students whose homes and families have more resources. However, it is important to note that the unit of analysis in Cromley's study is "nation" (wealthy nations vs. developing nations) rather than socio-economic status within nations. OECD (2009a) found that among top performers, motivation, indicated by enjoyment and active engagement in science learning inside and outside school, is unrelated to socio-economic factors. Moreover, in another study OECD (2007) found that after adjusting the expenditure on educational institutions per student between the ages of 6 and 15 years using purchasing power parities (PPP), spending for educational resources can explain only 19% of the variance in PISA science scores among different countries.

Further, based on the data from German students who took part in PISA 2003, Wittwer and Senkbeil (2008) found that students' access to a computer was not linked with their performance in mathematics. In addition, it did not matter how often students used a computer at home. Although Wittwer and Senkbeil's study was concerned with mathematics skills rather than science knowledge, their study and this one together imply that investment in computer resources might not be the immediate answer to improvement of science and math education. In this study almost none of the computer-related variables, except frequent use of educational software, turned out to be significant predictors of PISA science performance. But, it is important to point out that the nature of the software packages is education-oriented, and thus perhaps the content rather than the medium is what matters. Indeed, previous studies show that the media content consumed by children, rather than just using the media regardless of the content, affect scholastic outcomes (Graber, Nichols, Lynne, Brooks-Gunn and Botvin, 2006; Gunter, Clifford and McAleer, 1997). Surprisingly, in the Internet era a conventional medium (books) was identified as the crucial factor contributing to better science scores. The implication for parents, students, teachers, and policy makers is that they should consider re-prioritizing deployment of different types of educational resources. Last but not least, Canadian students report higher science enjoyment and science interest than their American counterparts. Although many studies pertaining to the relationship between motivation and science education have been conducted (e.g., Abrahams, 2009; Boyer, Phillips, Wallis, Vouk and Lester, 2009; Dede, Ketelhut and Ruess, 2004; Giancola, 2001) and some of them are situated in cross-cultural contexts (e.g., Berger and Hanze, 2009), the argument of "comparing apples and oranges" is occasionally raised by American educators. As illustrated in the background section, Canada and the United States are very similar in terms of economy, culture, population composition, and education systems. Thus, it is strongly suggested that American policy makers and educators should take this finding seriously by studying how neighboring Canadian students are motivated. Developing programs to build a positive attitude toward science among American students might be the key to narrowing the performance gap between these two North American countries.

References

Abrahams, I. (2009). Does practical work really motivate? A study of the affective value of practical work in secondary school science. *International Journal of Science Education* **31**, 2335-2353.

Berger, R. and Hanze, M. (2009). Comparison of two small-group learning methods in 12th grade physics classes focusing on intrinsic motivation and academic performance. *International Journal of Science Education* **31**, 1511-1527.

Bibby, R. W. (1987). Bilingualism and multiculturalism: a national reading. In *Ethnic Canada* (Edited by L. Driedger), 158-169. Copp Clark Pitman Ltd., Toronto, Canada.

Boyer, K. E., Phillips, R., Wallis, M. D., Vouk, M. A. and Lester, J. C. (2009). Investigating the role of student motivation in computer science education through one-on-one tutoring. *Computer Science Education* **19**, 111-135.

Bracey, G. (2009). Big tests: what ends do they serve? *Educational Leadership* **67**, 32-37.

Breton, R. (1987). Symbolic dimensions of linguistic and ethnocultural realities. In *Ethnic Canada* (Edited by L. Driedger), 44-64. Copp Clark Pitman Ltd, Toronto, Canada.

Bybee, R. W. (1997). *Achieving Scientific Literacy: From Purposes to Practices.* Portsmouth, Heinemann, New Hampshire.

Day, R. J. F. (2000). *Multiculturalism and the History of Canadian Diversity.* University of Toronto Press, Toronto.

Dede, C., Ketelhut, D. and Ruess, K. (2004). Designing for motivation and usability in a museum-based multi-user virtual environment. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California.

Driedger, L. and Halli, S. S. (2000). Racial integration: theoretical options. In *Race and Racism: Canada's Challenge* (Edited by L. Driedger and S. S. Halli), 55-76. McGill-Queen's University Press, Montreal, Quebec.

Fielding, A. H. (2007). *Cluster and Classification Techniques for the Biosciences.* Cambridge University Press, New York.

Gans, H. J. (2004). The American kaleidoscope, then and now. In *Reinventing the Melting Pot: The New Immigrants and What It Means To Be American* (Edited by T. Jacoby), 33-46. Basic Books, New York.

Giancola, S. P. (2001). Technology programs ··· for all or for some? *Journal of Science Education and Technology* **10**, 369-384.

Glazer, N. and Moynihan P. (1970). *Beyond the Melting Pot.* MIT Press, Cambridge, Massachusetts.

Graber, J. A., Nichols, T. Lynne, S. D., Brooks-Gunn, J. and Botvin, G. J. (2006). A longitudinal examination of family, friend, and media influences on competent versus problem behaviors among urban minority youth. *Applied Developmental Science* **10**, 75-85.

Gunter, B., Clifford, B. R. and McAleer, J. L. (1997). Learning from multi-topic science programmes on mainstream television. *Zeitschrift für Medienpsychologie* **9**, 3-23.

Larose, D. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining.* Wiley-Interscience, New Jersey.

Luan, J. (2002). Data mining and its applications in higher education. In *Knowledge Management: Building a Competitive Advantage in Higher Education* (Edited by A. Serban and J. Luan), 17-36. Josey-Bass, Pennsylvania.

Moodley, K. A. (1986). Canadian multicultural education: promises and practices. In *Multicultural Education in Western Societies* (Edited by J. A. Banks and J. Lynch), 51-75. Praeger Publishers, New York.

NCES. (2007). *Comparative Indicators of Education in the United States and Other G-8 Countries*: *2006.* Jessup, Maryland: Author.

NCES. (2008). *Highlights from PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context.* Washington, District of Columbia: Author.

OECD. (2003). *The PISA 2003 Assessment Framework - Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris: Author.

OECD. (2006a). *Education at a Glance*: *OECD Indicators 2006.* Paris: Author.

OECD. (2007). *PISA 2006*: *Science Competencies for Tomorrow's World.* Paris: Author.

OECD. (2009a). *Education at a Glance 2009*: *OECD Indicators.* Paris: Author.

OECD. (2009b). *PISA 2006 Technical Report.* Paris: Author.

Perez, A. and Cromley, J. (2009). Predicting science achievement from science media use and science enjoyment across 57 countries using PISA 2006. Paper presented at the 2009 annual meeting of the American Educational Research Association. San Diego, California.

Shaker, P. (2009). Preserving Canadian exceptionalism. *Education Canada* **49**, 28-32.

Skerrett, A. (2008). Racializing educational change: melting pot and mosaic influences on educational policy and practice. *Journal of Educational Change* **9**, 261-280.

Streifer, P. A. and Schumann, J. A. (2005). Using data mining to identify actionable information: breaking new ground in data-driven decision making. *Journal of Education for Students Placed at Risk* **10**, 281-293.

Wittwer, J. and Senkbeil, M. (2008). Is students' computer use at home related to their mathematical performance at school? *Computers and Education* **50**, 1558-1571.

Received November 4, 2011; accepted February 1, 2012.

Chong Ho Yu
Associate Professor
Department of Psychology
Azusa Pacific University
901 E. Alosta Ave., PO Box 7000 Azusa, CA 91702-7000, USA
chonghoyu@gmail.com

Charles Kaprolet
School Psychology
Scottsdale Unified School District
8505 E. Valley View Rd., Scottsdale, AZ 85250, USA
ckaprolet@susd.org

Angel Jannasch-Pennell
President
KOI Education
P.O. Box 13626, Phoenix, AZ 85002-3626, USA
angel@koi-education.com

Samuel DiGangi
Associate Professor
Mary Lou Fulton Teachers College
Arizona State University
PO Box 37100, Phoenix, AZ 85069, USA
sam@asu.edu