PISA 2009

# Explaining the Gender Gap in Reading through Reading Engagement and Approaches to Learning

cmec
Council of Ministers of Education, Canada
Conseil des ministres de l'Éducation (Canada)

Canada

Research Paper

# PISA 2009: Explaining the Gender Gap in Reading through Reading Engagement and Approaches to Learning

**Authors**
**Maria Chuy,** Council of Ministers of Education, Canada
**Roy Nitulescu,** Human Resources and Skills Development Canada

cmec
Council of Ministers of Education, Canada
Conseil des ministres de l'Éducation (Canada)

Canada

# Summary

Although Canada has always ranked in the top quarter of the Programme for International Student Assessment (PISA) evaluations in reading, and Canadian students are still listed among the best readers in world, a significant gender gap has been identified. Girls outperform boys with a gap roughly equivalent to half a reading-proficiency level, or one full year of formal schooling. This paper uses the PISA 2009 data set to investigate and isolate the factors contributing to the gender gap in Canada.

The paper is divided into two parts. It begins with a short review of the available literature about gender differences in literacy and factors related to them. It pays special attention to socio-cognitive differences between girls and boys, notably reading habits and learning strategies, which could potentially be improved through appropriate pedagogies and educational policies. The next sections focus on the PISA 2009 measurement of student engagement in reading activities and approaches to learning. The results for Canada are presented, showing that girls (i) read more diversely and generally enjoyed reading to a greater extent, (ii) used control and memorization strategies more often, and (iii) were more aware of the most effective meta-cognition strategies compared to boys. The last section of part 1 examines the role of all these factors in reading-performance variation through a series of regression analyses.

Part 2 dissects and analyzes the potential for closing the gender gap by applying Oaxaca and Blinder's method. First, the method is defined and justified, then different regression models are outlined and described. Results indicate that besides enjoyment of reading, two reading strategies showed significant and important contributions to the gender differences in reading: *control* and *summarizing*. *Control* is a cognitive strategy focusing on understanding a task's purpose and its main concepts, while *summarizing* is a meta-cognitive strategy reflecting an awareness of the most efficient ways to condense information. Girls outperformed boys in the use of these important strategies. There was an interesting result regarding *memorization*, a cognitive strategy derived from the frequency with which students try to memorize the text (without special focus on understanding). It appears that *memorization*, being more frequently used by girls, has a negative effect on reading scores. Thus, if girls did not employ this technique as frequently, their reading performance would be even higher.

The conclusion summarizes the paper's findings and considers their policy implications.

# Introduction

Schools, universities, and ministries have widely investigated the inequalities between boys and girls in reading performance during the past ten years (Booth, Elliott-Johns, & Bruce, 2009). Despite several years of extensive research, the gender gap remains a major problem for most countries in the Organisation for Economic Co-operation and Development (OECD): girls continue to outperform boys in reading in a number of national and international assessments.

- According to the Programme for International Student Assessment (PISA), 15-year-old girls have been significantly ahead of boys since the first assessment in 2000 and the gap has remained significant ever since. More precisely, the difference between boys and girls increased from 32 points in 2000 to 39 points in 2009 (OECD, 2010a).[1]
- The latest cycle of the Progress in International Reading Literacy Study (PIRLS) in 2011 showed that grade 4 girls outperform boys in almost all participating countries. On average, girls scored 16 points higher than boys (Mullis, Martin, Foy, & Drucker, 2012).[2]
- US National Assessment of Education Progress (NAEP) results demonstrated that girls do better than boys in reading across grade levels (4, 8, and 12) and across several years of assessment (1992, 1994, 1998, 2000, 2002, and 2003). Overall, the gender gap gets larger as grade level increases (Klecker, 2005).

Considering such a worrisome trend in most of the available assessments, the gender gap in reading continues to concern the OECD countries and a number of research initiatives are seeking to find out how to efficiently reduce this gap.

This paper uses the PISA 2009 data set to investigate and isolate the factors contributing to the gender gap in Canada. Although Canada has always ranked in the first quarter of PISA evaluations in reading and Canadian students are still listed among the best readers in world, a significant gender gap has been identified. The difference in reading scores increased from 32 points in 2000 to 34 points in 2009, with girls consistently being ahead of boys (OECD, 2010a). Such a difference is roughly equivalent to half a PISA reading-proficiency level, or one full year of formal schooling. A gap of this magnitude cannot be ignored — a deep understanding of its contributing factors is needed to identify productive pedagogies and learning strategies for Canadian boys.

This paper consists of two parts. It begins with a short review of the available literature about gender differences in literacy and factors related to them. It pays special attention to socio-cognitive differences between girls and boys, notably reading habits and learning strategies, which could potentially be improved through appropriate pedagogies and educational policies. The PISA 2009 reading measures are next discussed in detail, because any investigation of the gender gap requires a good understanding of how it was assessed. The following two sections focus on the PISA 2009 measurement of student engagement-in-reading activities and approaches to learning. The results for Canada are presented and discussed for each measure of reading engagement (i.e., enjoyment of reading, diversity in reading, and on-line reading) and approaches to learning (i.e., cognitive and meta-cognitive strategies). The last section of part 1 examines the role of all these factors in reading-performance variation through a series of regression analyses.

[1] The mean score in PISA is set at 500 and standard deviation at 100.
[2] Similar to PISA, the mean test score was set at 500 and the standard deviation at 100.

Part 2 dissects and analyzes the potential for closing the gender gap by applying Oaxaca and Blinder's method. First, the method is defined and justified, then the preliminary regression models are outlined and described. Results from these preliminary models are then discussed. The paper moves on to question the role of reading enjoyment in reading performance and then describes and analyzes the final regression model. The conclusion summarizes the paper's findings while also considering their policy implications.

# Part 1

# Identifying Gender Differences

## Literature review: Why do girls outperform boys in reading?

In their summary of research on gender, Smith and Wilhelm (2002) show that boys take longer to learn to read, read less than girls, and experience more difficulties comprehending narrative and expository texts. A number of factors could explain boys' lower performance. Among the most discussed issues are:

- ***Brain-based/biological differences***

  Some researchers argue that boys and girls begin school with different developmental strengths and weaknesses (Gunzelmann & Connell, 2006). More precisely, girls' left hemispheres, which are responsible for auditory processing and verbal expression, develop before boys' do. This earlier maturation would allow girls to benefit from the traditional language approaches from the beginning of school (Gurian, Henley, & Trueman, 2001). It also appears that boys' fine motor skills are not as developed in the early years, leading to difficulties in mastering the "biomechanics" of reading and writing (i.e., holding a pen, turning a page; see Martino, 2008 for discussion). However, the OECD report that explores the brain and learning shows that despite the existence of functional and morphological sex differences, it is extremely difficult to determine the importance of these differences (OECD, 2007). Thus, there is no study to date that demonstrates the existence of gender-specific processes involved in building up the brain networks during learning (see also Ruble, Martin, & Berenbaum, 2006).

- ***Differences related to socialization***

  Another factor that could explain the gender gap in reading relates to students' identity and self-stereotypes. More precisely, a number of studies show that boys consider reading a "feminine" activity (Katz & Sokal, 2003; Ruble et al., 2006; Sokal et al., 2005; Wilhelm & Smith, 2009). Under pressure for masculine identities, boys try to conform to a "cool" masculine image and reject reading, considering it a "pastime" for girls (Younger & Warrington, 1996; Warrington, Younger, & Williams, 2000). Such disengagement is not only directed toward reading, but to schooling overall: boys spend less time on their homework, tend to break school rules, are less organized, and more distracted in the classroom (Arnot, David, & Weiner, 1999; Davies & Brember, 2001; Engels, Aelterman, Van Petegem, & Schepens, 2004). Boys come to believe that showing too much interest in school work, and especially language-related subjects, is inappropriate behaviour, and this belief contributes to their lack of motivation with reading. On the other hand, girls seem to place a higher value on academic effort and share social affiliation with one another by embracing reading activities (Van de gaer, Pustjens, van Damme, & de Munter, 2007).

- ***Differences related to engagement in reading***

  The lack of boys' engagement with literacy has been documented in the educational literature over the last two decades (Safford, O'Sullivan, & Barrs, 2004; Clark & Trafford, 1995; OFSTED, 1993, 2003). Considering that reading engagement is an even stronger predictor of literacy achievement than socioeconomic status (OECD, 2002), it seems to significantly influence boys' achievement in reading. Topping, Samuels, and Paul's data (2008) show that when girls and boys have similar levels of reading

engagement (measured both in terms of quantity of books read and quality of comprehension), they achieve similar gains.

There are also researchers who argue that boys' lack of engagement in reading is not generic, and concerns only certain types of readings. For instance, Smith and Wilhelm (2002, 2006) showed that boys prefer to read for utilitarian purposes rather than for leisure. They choose reading materials that give immediate feedback on their competence and enable them to make something (e.g., instruction manuals). However, because most school materials are not utilitarian, boys do not see the same purpose or value in them. Similarly, Oakhill and Petrides (2007) demonstrated that reading comprehension is significantly affected by the content of reading passages for boys, but not for girls. Because boys are more influenced by the level of their interest, literacy curricula should reconsider its choice of texts and content to include boys' out-of-school literacy experiences and interests (Brozo, 2002; Pirie, 2002; Marsh & Millard, 2000).

- ***Differences related to approaches to learning: cognitive and meta-cognitive strategies***

  Neither boys nor girls passively receive information — they are actively involved in the learning process and the construction of meaning, using both their prior knowledge and information from the text (Kintsch, 2004). Being able to manage one's own cognitive processes and to choose appropriate goals and strategies is a prerequisite of successful lifelong learning (Boekaerts, 2009; Ryan & Deci, 2009).

  Reading strategies can be classified into cognitive and meta-cognitive ones (Griva, Alevriadou, & Semoglou, 2011). Cognitive reading strategies relate to the application of specific techniques facilitating comprehension — they generally involve direct interaction with the text. Meta-cognitive strategies are generally defined as "thinking about thinking." They play the role of self-regulator, enabling a reader to monitor cognitive strategies and to use appropriate reading techniques for learning from reading.

  The use of reading strategies is mediated by gender. Girls use a wider range of strategies, show greater awareness of strategic meta-knowledge, and are more flexible (Chandler, Lizotte, & Rowe, 1998; Green & Oxford, 1995; Griva, Alevriadou & Semoglou, 2011). However, boys are better at information retrieval and can sacrifice deep understanding for correct responses and rapidity (Arnot et al. 1999; Smith & Wilhelm, 2002).

It is important to note that the differences described here do not act independently, but in relation to each other. Being closely intertwined, they make the gender gap difficult to address. The degree to which each factor contributes still needs to be investigated. This study tries to shed light on socio-cognitive factors tackled in PISA 2009, namely:

- *engagement-in-reading activities*, including enjoyment of reading, diversity in reading, and on-line reading
- *approaches to learning*, including both cognitive strategies (memorizing, elaborating, and controlling) and meta-cognitive ones (understanding and remembering, summarizing).

Before discussing how these factors contribute to the gender gap, we survey measures and results of reading achievement used in PISA 2009.

## How reading was assessed in PISA 2009

PISA 2009 defines reading literacy as "the capacity of an individual to understand, use, reflect on and engage with written texts in order to achieve his/her goals, to develop his/her knowledge and potential, and to participate in society" (OECD, 2010a, p. 23). Thus, it goes beyond simple decoding and literal comprehension, and focuses

"on reading to learn," rather than "learning to read," by focusing on interpretation, reflection, and the ability to use reading to fulfill students' goals in life.

PISA questions are constructed so that test tasks are as close as possible to what students experience in the real world. Questions are organized around scenarios and varied in format. Around half of the questions are multiple choice and the other half are constructed-response questions, requiring both short and long responses. Reading materials vary from continuous texts (e.g., narration) and noncontinuous texts (e.g., graphs, lists) to mixed and multiple texts. Reading questions appear in four different contexts (i.e., personal, educational, occupational, public) and are meant to assess the following competencies: *access and retrieve*, *integrate and interpret*, *reflect,* and *evaluate*.

In PISA 2009, the assessment was organized in 13 linked testing booklets, with each student randomly assigned one of the booklets. Each individual was tested for 120 minutes, in which 54 per cent of her/his time was devoted to reading, 23 per cent to mathematics, and 23 per cent to science. Students were between 15 years 3 months and 16 years 2 months old at the time of the assessment.

## Reading results for Canada: Girls outperformed boys

PISA 2009 results showed that girls outperformed boys in reading literacy in every participating country. Among OECD countries, where the average reading score was 493 points, the gender gap was 39 points in favour of girls. In Canada, where the average reading score was 524 points, the gap was slightly lower at 34 points. Table 1.1 summarizes the descriptive statistics pertaining to gender differences in Canada.[3]

**TABLE 1.1 Overall reading scores in Canada, by gender**

| | Minimum | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|
| **Male** | 99.64 | 828.56 | 507.18 | 92.35 |
| **Female** | 79.50 | 843.46 | 541.53 | 84.86 |

Source: PISA 2009 data for Canada, calculations by author.

With an average score of 507 and 542 points respectively, both boys and girls in Canada performed above the OECD's average level (493 points) in reading literacy. However, analysis of standard deviations shows that there was more variation in the scores for boys than for girls in Canada. This suggests that girls not only performed better than boys on average, but they also had a more consistent performance distribution than boys. Moreover, with a standard deviation of 85 points (compared to 93 points in the OECD), the score variation for girls in Canada was smaller than the overall OECD score variation.

## How engagement in reading activities was measured in PISA 2009

In addition to responding to a two-hour subject assessment on reading, mathematics, and science, PISA students also completed a 30-minute questionnaire on their backgrounds. Among other items, this questionnaire includes a set of questions on student engagement, notably enjoyment of reading, diversity in reading, and on-line reading.(Besides containing questions on student engagement, the 30-minute questionnaire also asks students questions about their approaches to learning. These measures are discussed later.)

[3] It is worth noting that the PISA scale ranges from 0 to 1000 and that a score of 407 points is considered the minimum baseline level of proficiency that a student must have to be a productive member of society (OECD, 2010a). In the results for an overall reading literacy scale, the metric is based on a mean of 500 and a standard deviation of 100.

The level of *enjoyment of reading* was calculated by asking students the extent of their agreement with the following 11 statements:[4]

- I read only if I have to.
- Reading is one of my favorite hobbies.
- I like talking about books with other people.
- I find it hard to finish books.
- I feel happy if I receive a book as a present.
- For me reading is a waste of time.
- I enjoy going to a bookstore or a library.
- I read only to get information that I need.
- I cannot sit still and read for more than a few minutes.
- I like to express my opinions about books I have read.
- I like to exchange books with my friends.

The level of *diversity in reading* was calculated by asking students how frequently they read different types of reading materials, including: magazines, comic books, fiction books, nonfiction books, and newspapers.

*On-line reading* activity was measured by asking students how often they engaged in the following: reading e-mails, chatting on-line, reading on-line news, using an on-line dictionary or encyclopedia, searching on-line information to learn about a particular topic, taking part in on-line group discussions or forums, and searching on-line for practical information such as schedules, events, tips, and recipes.

The information from these three sets of questions was used to produce three standardized indexes, meant to reflect student engagement in reading. All three indexes had mean values of 0 and standard deviations of 1 at the OECD level, with positive values indicating greater engagement in reading.[5]

## Results for engagement in reading in Canada: Girls read more diversely and generally enjoyed reading to a greater extent than boys did

Table 1.2 gives index values and standard errors for reading diversity, on-line reading, and reading enjoyment for boys and girls, as well as their differences. Bold values for indexes of boys and girls mean that these values are significantly different from the OECD mean (with positive values being above the OECD average and negative values below it).[6] Bold values for male-female differences mean that these differences are statistically significant (no reference to OECD). Negative values in gender difference represent female advantages.

Results show that Canadian 15-year-old boys and girls exhibit substantial and statistically significant differences in their engagement in reading. More precisely, boys were below the OECD average for all three engagement indexes, although this difference was not significant for the on-line reading index. Girls enjoyed reading to a far greater extent than the OECD students on average.

[4] All negatively phrased items were reverse-scored.

[5] With the OECD mean set at 0, any positive Canadian index could be considered as being above the OECD average while any negative one could be considered below it.

[6] The OECD mean was calculated for a total of students, including both boys and girls.

**TABLE 1.2 Indexes and standard errors for engagement-in-reading activities in Canada, by gender**

| | | Reading enjoyment | On-line reading | Reading diversity |
|---|---|---|---|---|
| **Male** | *Average index* | **-0.28** | -0.03 | **-0.24** |
| | *Standard error* | (0.02) | (0.02) | (0.02) |
| **Female** | *Average index* | **0.55** | **-0.04** | 0.01 |
| | *Standard error* | (0.02) | (0.02) | (0.01) |
| **Difference (M/ F)** | *Average index* | **-0.83** | 0.00 | **-0.25** |
| | *Standard error* | (0.02) | (0.02) | (0.02) |

Source: PISA 2009 data for Canada, calculations by author.
Note: Bold values are significant at the 5% level.

Analysis of male-female differences in Canada shows that girls scored statistically higher than boys in reading diversity and reading enjoyment. (The female advantage was particularly large in reading enjoyment). However, both genders were equally engaged in on-line reading activities.

## Approaches to learning

PISA 2009 also measures approaches to learning — another set of factors that could possibly contribute to the gender gap in reading. Approaches to learning were measured through a series of questions related to cognitive and meta-cognitive strategies. The main difference between these two categories was that the first one (cognitive) focused on the *frequency* of a strategy use, and the second one (meta-cognitive) on students' *awareness* of strategy usefulness.

## How cognitive strategies were measured in PISA 2009

Cognitive strategies included *memorization, elaboration,* and *control strategies*.

The index of *memorization* was derived from the frequency with which students did the following: tried to memorize everything that is covered in the text; tried to memorize as many details as possible; read the text so many times that they can recite it; and read the text over and over again.

The *elaboration* strategies relied on the connection between new information and prior knowledge, out-of-school contexts, and personal experiences. To construct the elaboration index, PISA asked students how often they: tried to relate new information to prior knowledge acquired in other subjects; figured out how the information might be useful outside school; tried to understand the material better by relating it to personal experiences; and figured out how the text information fits in with what happens in real life.

The *control* strategies focused on understanding the purpose of a task and its main concepts. The control strategies index was based on students' reports of how often they: figured out what exactly they need to learn; checked if they understood what they have read; figured out which concepts they still haven't really understood; made sure that they remember the most important points of the text; and looked up additional information to clarify something they didn't understand.

The reported frequencies for these three sets of questions were combined into three standardized indexes, with mean values of 0 and standard deviations of 1 at the OECD level. Higher values on the indexes indicate more frequent use of the strategies.

## How meta-cognitive strategies were measured in PISA 2009

PISA assessment of meta-cognitive strategies focused on students' awareness of different strategies' usefulness. Such awareness constitutes a basis for self-regulation and helps students to efficiently manage their cognitive resources. Two meta-cognitive strategies were considered: *understanding and remembering*, and *summarizing*.

In order to calculate the index of *understanding and remembering*, PISA asked students to report on how useful they find the following strategies:

- I concentrate on the parts of the text that are easy to understand.
- I quickly read through the text twice.
- After reading the text, I discuss its content with other people.
- I underline important parts of the text.
- I summarize the text in my own words.
- I read the text aloud to another person.

In order to calculate the index of *summarizing*, students were asked to imagine themselves in a situation where they need to write a long and rather difficult two-page text about water fluctuations in a lake in Africa, and then to report on how useful they find the following strategies in this context:

- I write a summary. Then I check that each paragraph is covered in the summary, because the content of each paragraph should be included.
- I try to copy out accurately as many sentences as possible.
- Before writing the summary, I read the text as many times as possible.
- I carefully check whether the most important facts in the text are presented in the summary.
- I read through the text, underlining the most important sentences, then I write them in my own words as a summary.

The method for calculating indexes for meta-cognitive strategies was different than the one used for cognitive strategies — in this case a rater-scoring system. More precisely, PISA researchers compared students' reported ranks with "optimal" ranks determined by experts in cognitive processing. Based on the agreement with experts, two standardized indexes were calculated, for *understanding and remembering* and *summarizing* respectively.[7] Higher scores on indexes indicated higher agreement with the experts' rankings and, therefore, greater likelihood that students will efficiently self-regulate their own learning.

## Results for cognitive and meta-cognitive strategies in Canada

Table 1.3 gives index values and standard errors for five reading strategies for boys and girls, as well as their differences.[8]

- *Cognitive strategies: girls used control and memorization strategies more often than boys did.* For the control and memorization indexes, girls did significantly better than boys, with boys being below the OECD average and girls above it (see table 1.3). Regarding the elaboration index, both groups of students were below the OECD average, but boys were significantly above the girls.

[7] Like other measures, the two meta-cognitive indexes were calculated to have a mean of 0 and standard deviation of 1 across OECD countries, so that the results could be interpreted in a similar way.

[8] Data in Table 1.3 should be interpreted in the same way as previous data.

- *Meta-cognitive strategies: girls were more aware of the most effective meta-cognition strategies compared to boys.* Girls had statistically significant advantages in both meta-cognition indexes (see table 1.3), with boys being below the OECD average and girls above it. The female advantage was particularly large for the index of summarizing strategies.

**TABLE 1.3 Indexes and standard errors for cognitive and meta-cognitive strategies in Canada, by gender**

| | | Cognitive strategies | | | Meta-cognitive strategies | |
|---|---|---|---|---|---|---|
| | | Memorize | Elaborate | Control | Understand | Summarize |
| **Male** | *Average index* | **-0.16** | **-0.16** | **-0.09** | **-0.17** | **-0.19** |
| | *Standard error* | (0.02) | (0.02) | (0.02) | (0.02) | (0.01) |
| **Female** | *Average index* | **0.12** | **-0.25** | **0.30** | **0.12** | **0.24** |
| | *Standard error* | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) |
| **Difference (M/ F)** | *Average index* | **-0.28** | **0.09** | **-0.39** | **-0.29** | **-0.43** |
| | *Standard error* | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |

Source: PISA 2009 data for Canada, calculations by author.
Note: Bold figures are significant at the 5% level.

- *Strategies are correlated with each other:* The three general cognitive strategies and two meta-cognitive strategies should not be thought of as exclusive. In fact, they tend to complement each other as is evident in the positive correlations between most of the indexes (see table 1.4).

**TABLE 1.4 Correlations between reading strategies**

| Correlation | Memorize | Elaborate | Control | Understand | Summarize |
|---|---|---|---|---|---|
| **Memorize** | 1 | - | - | - | - |
| **Elaborate** | 0.36 | 1 | - | - | - |
| **Control** | 0.58 | 0.50 | 1 | - | - |
| **Understand** | 0.05 | 0.06 | 0.24 | 1 | - |
| **Summarize** | 0.05 | 0.02 | 0.26 | 0.44 | 1 |

Source: PISA 2009 data for Canada, calculations by author.
Note: All correlations were significant up to the 1% level, even when using the Bonferroni correction.

## Role of engagement-in-reading activities and approaches to learning in reading performance

To discuss the impact on reading performance of engagement-in-reading activities and approaches to learning, this section considers the results of the following regression models: (1) simple linear regression models, allowing us to examine the relationship between a single predictor and the outcome without controlling for other variables (as reported by Brochu, Gluszynski, & Cartwright, 2011), and (2) a multiple regression model, which allows one to calculate the effects of a particular predictor while controlling for all other predictors in the equation (this model represents an additional analysis to the results reported by Brochu, Gluszynski, & Cartwright, 2011).

### *Effects of engagement-in-reading activities*

- *Results of Simple Linear Regression Models:* Table 1.5 represents simple linear models' results for the engagement-in-reading factors. Two statistics are reported: marginal effect (defined as the change in reading scores that is produced by a one-unit change in a particular factor), and explained variance of each factor in reading performance (defined as $R^2$ x 100%). Reading enjoyment yielded almost 36 extra points in reading

proficiency, and explained 20 per cent of variation in reading scores. The factors of reading diversity and on-line reading factors had significant effects, but their explanatory powers were far below the 20 per cent captured by the reading enjoyment index. (Care must be taken when interpreting the regression effects because they do not always allow for inferring causal relationships. Thus, an increase in reading enjoyment does not necessarily cause a direct gain in reading scores).

**TABLE 1.5 Engagement-in-reading activities – Results of simple linear regressions models**

| | Change in reading score per unit increase of the index | | Explained variance in student performance ($R^2$ x 100%) | |
|---|---|---|---|---|
| | Effect | Standard Error | % | Standard Error |
| **Reading enjoyment** | **35.70** | (0.80) | 20.11 | (0.8) |
| **On-line reading** | **14.07** | (1.25) | 2.17 | (0.4) |
| **Reading diversity** | **18.41** | (0.98) | 4.32 | (0.4) |

Source: Brochu, Gluszynski, & Cartwright, 2011.
Note: Bold figures are significant at the 5% level.

## *Effects of approaches to learning: Cognitive and meta-cognitive strategies*

- *Results of Simple Linear Regression Models:* Among the cognitive strategies, only the control strategy showed a substantial effect. It yielded a relatively large 26-point gain in reading proficiency (see table 1.6). The memorization and elaboration strategies showed significant but small effects (around 3 points each) and quite low explanatory power at less than 1 per cent each.

  Both meta-cognition strategies were strong predictors of greater reading proficiency: the "understanding and remembering" strategy showed a marginal effect of 27 points, and summarizing had an effect of 35 points. Their explanatory power was also relatively strong, especially for the summarizing strategy which explained 16 per cent of the variation in reading scores alone.

  Together, the control strategy and both meta-cognition strategies may shed some light on how to improve overall reading proficiency in Canada.

**TABLE 1.6 Cognitive and meta-cognitive strategies – Results of simple linear regression models**

| | | Change in reading score per unit increase of the index | | Explained variance in student performance ($R^2$ x 100%) | |
|---|---|---|---|---|---|
| | | Effect | Standard Error | % | Standard Error |
| **Cognitive strategies** | *Memorize* | **3.30** | (0.99) | 0.16 | (0.1) |
| | *Elaborate* | **2.59** | (0.85) | 0.10 | (0.1) |
| | *Control* | **25.78** | (1.01) | 10.00 | (0.7) |
| **Meta-cognitive strategies** | *Understand* | **27.23** | (0.92) | 9.45 | (0.6) |
| | *Summarize* | **35.29** | (0.81) | 15.71 | (0.7) |

Source: Brochu, Gluszynski, & Cartwright, 2011.
Note: Bold figures are significant at the 5% level.

- *Results of a Multiple Regression Model:* Since most of the reading strategies are associated with each other (see correlations above), it is important to evaluate the effect of each strategy when controlling for the effects of others. In order to do this, all factors were included in a multiple regression model: memorization, elaboration, control, understanding and remembering, and summarizing. This model explained 24 per

cent of variance in reading scores. Table 1.7 contains adjusted effects and standard errors resulting from this model.

The new adjusted effects were quite different from the original effects reported in table 1.6. For the cognitive strategies, the control strategy's effect increased by almost 3 points, and the effects of the memorization and elaboration strategies decreased by 16 and 11 points respectively. For the meta-cognition strategies, the effect decreased by 16 points for the understanding-and-remembering index, and by 12 points for the summarizing index.

Overall, when cognitive and metacognitive factors were considered together, better reading performance was associated with better control, understanding, and summarizing skills, and negatively associated with memorizing and elaborating.

**TABLE 1.7 Cognitive and meta-cognitive strategies – Results of a multiple regression model**

| | | Change in reading score per unit increase of the index | |
|---|---|---|---|
| | | Effect | Standard Error |
| **Cognitive strategies** | *Memorize* | **-12.40** | (0.83) |
| | *Elaborate* | **-8.56** | (0.86) |
| | *Control* | **28.58** | (0.99) |
| **Meta-cognitive strategies** | *Understand* | **10.66** | (0.92) |
| | *Summarize* | **23.15** | (0.85) |

Source: PISA 2009 data for Canada, calculations by author.
Note: Bold figures are significant at the 5% level.

## Enjoyment of reading and summarizing as two main mediators of gender inequalities in reading

As regression analyses demonstrate, the enjoyment of reading and summarizing strategies seem to have a strong effect on reading performance. The PISA 2009 international report (OECD, 2010b, chap. 3) examines these two indexes in detail, considering them the most important mediators of gender inequalities in reading. The report states that, on average across OECD countries, almost 70 per cent of the gender gap in reading performance is the indirect result of the differences in how much boys and girls enjoy reading and how much they know about efficient ways to summarize information.

Canada is listed among the countries with the highest degree of mediation by reading enjoyment and summarizing (OECD, 2010b). According to the OECD report, the gender effect in Canada could be narrowed from 34.4 points to 0.5 by controlling for the indirect effect of these two factors. More precisely:

- if Canadian boys enjoyed reading as much as girls do, their reading score would increase by 29 points (see OECD, 2010b, table III.3.4);
- if Canadian boys were as aware of effective summarizing strategies as girls are, their reading score would increase by 15 points (see OECD, 2010b, table III.3.4).

When considering these findings, the influence of reading enjoyment on student performance cannot be interpreted univocally (we still do not know if children read better because they enjoy reading more, or if they enjoy reading because their reading skills are better). Thus, the role of other factors, such as reading strategies, should be better elucidated. The paper will address this question in the next part by means of a special statistical technique (the Oaxaca-Blinder method), that breaks the gender gap down into contributing factors to reveal how different cognitive and meta-cognitive strategies can explain gender inequalities in reading.

# Part 2

# Explaining the gender gap in reading: Oaxaca-Blinder decomposition

## How to measure the potential for closing the gender gap

As we've seen, boys and girls differ significantly in both their engagement in reading and their approaches to learning. These engagement and approach "factors" have been associated with greater reading performance due to the significance of their marginal effects. Although these marginal effects are relatively informative, it is possible to take the analysis even further to gain a deeper understanding of the factors contributing to the gender gap itself, and not only to greater reading performance. Ultimately, what needs to be measured is not just a marginal effect, but the *potential* for boys to *catch up* to girls and to close the gender gap in PISA reading measures. To achieve this, we applied the Oaxaca-Blinder decomposition (Jann, 2008) to several multiple regression models relating student background factors to reading performance.

## Group regressions and baseline reading scores

To perform the decomposition, identical regression models must first be estimated separately for boys and girls. Each one of these group regressions has an intercept term that is a lone coefficient unrelated to any of the variables in the regression model. To interpret this term, one must remember that a regression model contains a set of explanatory variables, which in this case helps predict PISA reading performance. So, if this set of predictors is held constant (i.e., controlled) for all students in the group, the resulting reading score would be represented by the intercept term. This intercept term can thus be defined as a baseline reading score. In this way, the intercept from the female-only regression model is the female baseline reading score, while the intercept from the male-only regression model is the male baseline reading score. The difference between the female and the male baseline score can be further defined as the *baseline gender gap*. This baseline gap is what persists after holding all the predictors of reading performance constant. Therefore, a significant baseline gap would indicate that important factors have been omitted from the regression model. On the other hand, a statistically insignificant baseline gap would indicate that the factors in the model explain the gender gap very well.

## Decomposing the gap into explainable and unexplainable portions

The decomposition method involves manipulating marginal effects from group regression models with the aim of separating explainable from unexplainable effects. Explainable effects are those that have been determined (by the decomposition) to be due to boys having different *endowments* than girls. For example, girls are more aware than boys of proper summarizing techniques; thus we can say that girls have been *endowed* with better summarizing skills than boys. The difference in endowments between the two thus translates into a marginal effect that we can explain, that is, girls perform better than boys in reading due to their approaches to summarizing texts. This endowment concept applies to all the factors that contribute to greater reading performance. The key feature of this model is that the explained marginal effect is an actual measure of the

potential that exists for bringing boys and girls to the same level of reading performance by matching their endowment levels.

Unexplainable effects are those determined to be due to boys and girls not being equally *affected* by an equal level of endowments. In other words, boys and girls may not have the same marginal effects from engagement-in-reading factors or approaches-to-learning factors. Although such a difference is very difficult to explain, the decomposition can measure it. Significant unexplainable effects may imply that bringing boys and girls to the same level in terms of their endowments may not actually close the gap between the genders in reading performance. This unexplained effect has been typically used as a measure of discrimination (Jann, 2008). However, in this context such an interpretation is not relevant because boys and girls were assessed and evaluated identically. Effects from unobserved differences between boys and girls are also included in the unexplained contributions to the gap in the form of a constant term which is represented by the baseline gap defined earlier. This term is due to important factors being excluded from the regression model.

## Outline of models to be considered

Three multiple regression models, each with PISA reading performance as the *explained* variable, were decomposed via the method of Oaxaca and Blinder. The first model contains as *explanatory* variables the eight contributing factors: understanding and remembering, summarizing, control, memorization, elaboration, reading enjoyment, reading diversity, and on-line reading. These eight factors can be thought of as measuring the *academic* backgrounds of PISA students, since they effectively measure the amount of academic skill each student has been endowed with. The model: (1) introduces the reader to a simple example of how the Oaxaca-Blinder decomposition is interpreted, and (2) shows how the baseline gap becomes insignificant when the omitted variables are included in the second model.

The second model is similar to the first, however, it is augmented with a set of additional *control* variables (omitted from model 1), which measure the students' *nonacademic* backgrounds. These variables control for differences between students in their socioeconomic backgrounds, the language that they speak at home, whether they are native of Canada or first/second-generation immigrants, whether they go to school in an urban or a rural area, whether they have a positive or a negative relationship with their teachers, and finally whether they make frequent use of libraries in general.

These nonacademic control variables are meant to untangle indirect from direct effects. It is not necessarily clear whether a student's nonacademic background relates directly to their reading performance, or if instead it acts only indirectly. In other words, students may have a poor level of academic skill due to their socioeconomic situation, and thus perform poorly in PISA reading due to this two-step effect. Such a case would imply that the socioeconomic background of the student plays only a secondary role in PISA reading performance. Thus, the nonacademic control variables allow for the possibility of isolating the variables that play a primary role in explaining reading performance from the variables that play only a secondary role. Ultimately, this approach is

a more precise means of capturing the potential for closing the gender gap in reading literacy compared to the simpler approach of the first model.

The third and final model is described in the next section.

## Model 1: Academic factors

To get a rough idea of how much each of the eight academic factors contributes to the gender gap in reading performance, the explanatory variables in the first regression model were decomposed into explainable and unexplainable components. Table 2.1 summarizes this first model's Oaxaca-Blinder regression output.

**TABLE 2.1 Model 1: Oaxaca-Blinder decomposition with 8 factors, no controls**

| Model 1 | Summary Coefficients | |
|---|---|---|
| **Girls** | **543.83** | |
| **Boys** | **511.71** | |
| **Difference** | **32.12** | |
| **Explained** | **37.77** | |
| **Unexplained** | **-5.64** | |
| | **Explained / endowments** | **Unexplained / coefficients** |
| **Understand** | **2.35** | -0.03 |
| **Summarize** | **8.01** | **0.17** |
| **Control** | **7.94** | **-0.78** |
| **Memorize** | **-3.14** | -0.05 |
| **Elaborate** | **0.99** | **-1.31** |
| **Reading enjoyment** | **22.15** | -0.16 |
| **Reading diversity** | -0.44 | 0.08 |
| **On-line reading** | -0.11 | 0.08 |
| **Constant** | - | **-3.62** |

Source: PISA 2009 data for Canada, calculations by author. Bold figures are statistically significant at 5%.

The regression output is made up of three groups of coefficients. The first group at the top of table 2.1, labelled as *summary coefficients*, describes the overall decomposition of the gender gap. Here, the gap was measured to be 32 points in favour of girls (543.83 minus 511.71 points).[9] Each of the eight factors' individual contributions to the gender gap was decomposed into an explainable and an unexplainable portion.

The *explained contributions* are the marginal effects due to differences boys' and girls' endowments in reading engagement and approaches to learning. The sum of the explained contributions was 38 points. Among the eight indexes, the index of reading enjoyment contributed the most to the reading gap by far. Indeed, this model predicts that if boys and girls were to have the same level of reading enjoyment (specifically, if they both had the same value for the index of reading enjoyment as the average Canadian PISA student), the difference between their PISA reading scores would be narrowed by 22 points. The meta-cognition strategies (understanding and remembering, and summarizing) and the control strategy also contributed substantially to the gender gap. Together, the two meta-cognition strategies accounted for 10 of the 38 explained points. Similarly, the index of control strategy accounts for almost 8 of the 38 explained points. As for the memorization index, its explained contribution must be interpreted in the opposite way since it is negative. Indeed, girls reported

[9] The gap is different from the true one since the addition of explanatory variables distorts the gap's measurement.

using memorization strategies far more often than boys. However, this index was associated with lower reading scores, when other indexes were taken into account (see table 1.7). This implies that girls were, in a sense, penalized for relying on memorization as an approach to studying. Thus, if both boys and girls were to rely on memorization strategies to the same extent as the average Canadian PISA student, the gender gap would actually *expand* by three points. Differences between boys and girls in the use of elaboration strategies when studying explained very little of the gender gap in reading scores.

The *unexplained contributions*, on the other hand, were due to boys and girls not being equally affected by equal levels of endowments. This translates into an unexplainable source of gender differences in reading performance. Here the coefficients are all either statistically insignificant or very small. The sum of the unexplained contributions was -6 points. The bulk of these six points is due to the four-point baseline gap in favour of boys represented by the constant term in the decomposition. This baseline gap is due to the nonacademic factors being omitted from the model.

## Model 2: Academic and nonacademic factors

In order to untangle the indirect from the direct effects and get a more precise measure of how much each of the eight academic factors contributes to the gender gap, the two groups of explanatory variables in the second regression model were decomposed into explainable and unexplainable components. Table 2.2 summarizes the Oaxaca-Blinder regression output of this second model.

**TABLE 2.2 Model 2: Oaxaca-Blinder decomposition with 8 factors plus controls**

| Model 2 | Summary Coefficients | |
|---|---|---|
| **Girls** | **545.67** | |
| **Boys** | **514.49** | |
| **Difference** | **31.18** | |
| **Explained** | **32.89** | |
| **Unexplained** | -1.72 | |
| | **Explained / endowments** | **Unexplained / coefficients** |
| **Understand** | **2.25** | -0.02 |
| **Summarize** | **7.17** | **0.28** |
| **Control** | **6.36** | **-0.83** |
| **Memorize** | **-2.63** | -0.03 |
| **Elaborate** | **0.78** | **-1.26** |
| **Reading enjoyment** | **22.28** | -0.29 |
| **Reading diversity** | 0.20 | 0.00 |
| **On-line reading** | -0.10 | 0.15 |
| **Student/teacher relationship** | **0.31** | -0.17 |
| **Student's library use** | **-3.43** | 0.06 |
| **Socioeconomic status** | -0.56 | **2.20** |
| **2nd-generation immigrant** | 0.00 | **1.42** |
| **1st-generation immigrant** | 0.06 | 1.08 |
| **Home language** | **0.26** | -0.03 |
| **Urban** | -0.06 | -0.27 |
| **Constant** | - | -3.99 |

Source: PISA 2009 data for Canada, calculations by author. Bold figures are statistically significant at 5%.

**Note:** Annex 2 presents estimates of this model at the provincial level and for selected countries. Due to sample-size issues at the provincial level, the model was simplified to include fewer variables.

The summary coefficients at the top of table 2.2 suggest that the entire gender gap was due to explainable effects, since the unexplainable contributions as a whole were measured to be *statistically insignificant*. According to this model, girls outperformed boys in PISA reading only because they were more favourably endowed academically than boys. This can be explained by the significant effect sizes identified in table 1.5.

The individual explained contributions to the gender gap of each of the eight academic factors appear in the left-hand column of table 2.2. These coefficients are very similar in magnitude to the equivalent ones in table 2.1. The consistency of these results is especially encouraging since model 2 contains so many additional variables compared to model 1. Indeed, this fact supports two conclusions. First, the explained contributions coming from the eight *academic factors* are practically unaffected by the addition of the nonacademic control variables in the regression model. It is thus possible to conclude with considerable confidence that girls outperformed boys in PISA reading because they were more engaged in reading than boys and because they were more knowledgeable than boys about appropriate learning approaches. Such a strong statement could not be made based on model 1 alone.

Second, the explained contributions coming from the *nonacademic factors* are either statistically insignificant or not particularly substantial, with the exception of the factor of student library use. This suggests that these

control variables *do not directly relate* to reading performance. Instead they play only a secondary role (at best) in explaining the gender gap. This is important since it implies that policies aimed at bridging the gap between boys and girls in their engagement-in-reading and their approaches to learning would actually also bridge the gap in PISA reading performance, regardless of these students' nonacademic backgrounds. In other words, the eight academic factors play a primary role in determining the gender gap.

Looking closely at the individual explained contributions, this time it is possible to infer with considerable confidence that the bulk of the PISA reading performance gap between the genders was due to the fact that girls had a far greater appreciation for reading-related activities than boys did— that is, their index of reading enjoyment was much more pronounced than that of boys. Specifically, the index of reading enjoyment explained 22 points of the roughly 31-point gender gap. This coefficient was statistically significant at the 1 per cent level and corresponds to over 70 per cent of the gap. All the other coefficients contributed far less to the gap, by comparison. For example, besides reading enjoyment, the variables that contributed the most to the gap were the index of summarizing strategies (7 points), and the index of control strategies (6 points).

It is important to note that the baseline gap in this model (represented by the constant term -3.99) is statistically *insignificant*. This is due to the addition of the nonacademic factors that now explain a part of the actual gender gap that model 1 could not explain. This result implies that model 2 explains the gender gap as fully as is possible, given the data available. This is true because: (1) the insignificance of the baseline gap suggests that there are no longer any more unobserved variables creating a gap between boys and girls in PISA reading scores, (2) the unexplained contributions are also insignificant as a whole, at -1.72 points, and (3) academic factors play a primary role in determining the gender gap. Thus, our model seems to fully explain the gender gap.

The analysis has established that the index of reading enjoyment accounts for the larger part of the gender gap. This is not by any means a new result. On the contrary, the significance of this measure has been well documented ever since the first PISA round in 2000 (OECD, 2002, 2010b), as the literature review mentions. The result has also been confirmed from the regression analysis. But this paper endeavours to refine these findings by looking beyond the impressive explanatory strength of the index of reading enjoyment.

## Looking beyond reading enjoyment

One may suspect that the index of reading enjoyment is crowding out other more subtle findings by taking up too much "statistical space." As table 2.3 shows, this index is significantly, although moderately, correlated with some of the other indexes included in the regression models presented earlier. More precisely, significant correlation coefficients above 0.21 are found for all three of the other indexes that best explained the gender gap: index of understanding and remembering, index of summarizing, and index of control. Most noteworthy is that the index of reading enjoyment and the index of reading diversity have a correlation coefficient of almost 0.50, which is relatively large in this context. The index of reading enjoyment also serves as a very good proxy for the reading scores themselves.[10] In fact, variations in reading enjoyment alone explain 20 per cent of the variations in reading scores.[11] Considering that multiple regression models involving several PISA variables rarely explain even 40 per cent of the variation in PISA scores, one variable explaining 20 per cent of score variations all by itself is an interesting fact. However, it is difficult to disentangle these measures. We cannot infer whether motivated readers are the ones performing very well in PISA reading or whether, on the contrary, it is high-performing students who tend to be motivated readers. Due to this challenge, the

[10] The correlation between the index of reading enjoyment and PISA reading scores is 0.45. This may be more than just moderate given that even the correlation coefficients between plausible values of the reading scores are only about 0.91.

[11] When regressing reading enjoyment on reading performance, the $R^2$ is 0.20.

remainder of this paper's analysis focuses on abstracting from the index of reading enjoyment to reach more nuanced conclusions.

**TABLE 2.3 Correlation coefficients with respect to reading enjoyment index**

| Correlation | Understand | Summarize | Control | Memorize | Elaborate | Reading diversity | On-line reading |
|---|---|---|---|---|---|---|---|
| **Reading enjoyment** | 0.21 | 0.26 | 0.32 | 0.16 | 0.17 | 0.47 | 0.17 |

Source: PISA 2009 data for Canada, calculations by author. All correlations were significant at the 1% level.

## Alternate version of model 2

Before attempting a more advanced type of analysis, it is worth returning to the second model we considered. The simplest way to abstract from the index of reading enjoyment is to completely drop it from the regression model (model 2b). It is of some interest to measure what effect this would have on the other coefficients in the otherwise unchanged model. The summary of the Oaxaca-Blinder regression output, after dropping the index of reading enjoyment, is presented in tables A1.1 and A1.2 of annex 1.

Whereas the second model managed to account for practically the entire gap, this third model explains just over half (54%) of the difference in PISA reading performance between boys and girls. This is to be expected given the explanatory strength of the index of reading enjoyment which has been dropped from this model. Otherwise, the coefficients are very similar to those estimated in the two preliminary models. The indexes that stand out the most are still the index of summarizing strategies as well as the index of control strategies. However, the index of reading diversity is now statistically significant. This index alone accounts for over 8 per cent of the gender gap in this model. Thus, it is evident that the index of reading enjoyment was crowding out the effect of reading diversity. Regarding the index of on-line reading activities, as in previous models, its coefficient is insignificant.

Perhaps the most interesting thing that occurs when reading enjoyment is dropped from model 2 is that the constant term becomes quite large and statistically significant. Indeed, model 2b has a baseline gap of 11 points in favour of girls, whereas this term was insignificant in model 2. This implies that the now unobserved effect of reading enjoyment (which was omitted from this model) is being captured by the constant term. Thus, omitting important variables creates a significant baseline gap that contains the omitted effects. There is no evidence of such an omission in model 2, and part of the purpose of model 2b was to show what kind of undesirable effect such an omission can have.

## Model 3: Adjusting for reading enjoyment

Unfortunately, it is incorrect to simply drop the index of reading enjoyment from the regression models. This is mainly due to the unwanted omitted variable effect. Such a modelling choice cannot be argued for because this reading-enjoyment factor plays an extremely strong role in explaining reading performance. More importantly, it has been identified as a primary role player in determining the gender gap itself. According to the analysis, it is indeed the single factor that holds the most potential for closing the gender gap. Therefore, another method was used to abstract from the index of reading enjoyment, this time without simply disregarding it. This method consists in isolating the part of the PISA reading score that is uncorrelated with the index of reading enjoyment and then using this new *adjusted measure* of reading performance to estimate and decompose the exact same regression models. First, we estimate a regression model consisting of only reading performance and the index of reading enjoyment. Next the model is used to *predict* reading scores based on reading enjoyment only. Finally, these predictions are subtracted from the true reading scores. This subtraction produces what we

call *residuals*. These residuals, by construction, are uncorrelated with reading enjoyment. Thus, the residuals are the new *adjusted measure* of reading performance.

One can consider that the new measure of reading performance has thus been adjusted for reading enjoyment in the following sense: the portion of the *adjusted* reading score that is accounted for by a student's reading enjoyment is now zero. In other words, reading enjoyment no longer plays a role in determining adjusted PISA reading performance. Thus, it is no longer necessary to include it in subsequent regression models. Indeed, the purpose of this adjustment procedure was to justify dropping the reading enjoyment factor from the regression model, but without completely ignoring it, as was the case with model 2b.

The regression in model 3 uses the adjusted reading scores as the explained variable and both academic and nonacademic factors as explanatory variables. Table 2.4 summarizes the Oaxaca-Blinder regression output for this model.

**TABLE 2.4 Model 3: Oaxaca-Blinder decomposition with adjusted reading scores**

| Model 3 | Summary Coefficients | |
|---|---|---|
| **Girls** | **4.70** | |
| **Boys** | 2.89 | |
| **Difference** | 1.81 | |
| **Explained** | **8.58** | |
| **Unexplained** | **-6.77** | |
| | **Explained / endowments** | **Unexplained / coefficients** |
| **Understand** | **2.08** | -0.03 |
| **Summarize** | **6.81** | **0.24** |
| **Control** | **5.98** | **-0.90** |
| **Memorize** | **-2.50** | -0.04 |
| **Elaborate** | **0.79** | **-1.30** |
| **Reading diversity** | **-0.59** | 0.13 |
| **On-line reading** | -0.10 | 0.14 |
| **Student/teacher relationship** | **0.23** | -0.20 |
| **Student's library use** | **-3.81** | -0.05 |
| **Socioeconomic status** | -0.55 | **2.01** |
| **2nd-generation immigrant** | -0.01 | **1.47** |
| **1st-generation immigrant** | 0.05 | 1.07 |
| **Home language** | **0.25** | 0.12 |
| **Urban** | -0.06 | -0.32 |
| **Constant** | - | **-9.10** |

Source: PISA 2009 data for Canada, calculations by author. Bold figures are statistically significant at 5%.
Note: Reading scores were adjusted for reading enjoyment.

According to this model, the gap in reading performance between boys and girls is not statistically significant (1.81 points). This is, of course, due to the new adjusted measure. Specifically, in the absence of the reading enjoyment factor, the model cannot distinguish a significant gap between boys and girls in adjusted PISA reading.

Now, despite the gender gap being measured as insignificant, the model still offers some important interpretive value. Indeed, the model predicts that girls will outperform boys by almost nine points in reading literacy given the differences in reading engagement and approaches to learning between themselves and boys. Since this

prediction is much larger than what the model actually measures (1.81 point), there remains an *unexplained* portion of the gender gap of almost -7 points in this case. The significance of this unexplained portion is mostly due to the statistically significant constant term, which represents the baseline gap. Here, the baseline gap is nine points in favour of boys. Thus, according to this model, when all factors are held constant, after abstracting from reading enjoyment, boys have an unexplained advantage over girls in PISA reading due to unobserved variables outside of the model. This implies that adjusting reading scores for reading enjoyment allows for the possibility of new variables to play a role in determining the gender gap. In a sense, it is because these new variables are not inside the model that the baseline gap is so pronounced.

The *explained* contributions from this model are very similar to the ones estimated by the other models we discussed. If boys and girls were equally aware of the most effective strategies to understand and remember, and summarize what they read,[12] their gap in reading performance would be narrowed by almost nine points (2.08 plus 6.81 points). Furthermore, if they were both to use goal-setting strategies to control their studying activities as often as the average Canadian PISA student does, the gap between boys and girls in PISA reading would narrow by another 5.98 points. Finally, if boys and girls relied equally on memorization strategies as the average Canadian PISA student does, the gender gap would widen by 2.5 points. As in the preliminary models, the index of elaboration strategies plays a negligible role in explaining the gender gap.

The *unexplained* coefficients are all very small, and they do not add up to a very strong effect. Among the seven coefficients related to the indexes of interest, only three were statistically significant at all. As for the statistically significant constant term, it is about nine points in the negative (-9.10 points). This term represents the part of the gap that cannot be captured by any of the variables included in the model. It represents a baseline male advantage in reading that can be measured only once reading scores have been adjusted for reading enjoyment (as described above).

Together all of these three models have very similar results. This shows how consistent the findings are. Regardless of how the index of reading enjoyment is incorporated into the model (or even not incorporated at all), the key coefficients maintain their statistical significance as well as their magnitude. This suggests that as much as the index of reading enjoyment alone accounts for the bulk of the gender gap in reading performance, the gap between the genders' use and knowledge of appropriate approaches to learning accounts for a persistently significant portion of their difference. Thus, the gap cannot be narrowed without equal pedagogical emphasis on both approaches to learning and reading engagement.

## Main findings

Part 1 identified endowment advantages in favour of girls, showing that girls have significant advantages over boys in their engagement-in-reading and approaches to learning. Additionally, the PISA-measured factors associated with these advantages were found to be predictors of greater PISA reading proficiency. The components identified by the three decomposition models in part 2 represent a real potential to close the gender gap in reading performance between boys and girls. This is because these explained effects directly relate to the real (PISA-measured) differences between boys and girls in their academic endowments. That said, the marginal effects resulting from simple linear and multiple regression models in part 1 could not be interpreted in this way.

The most persistent result of all is that the index of reading enjoyment contains the bulk of the numerical potential for improving the PISA reading performance of boys. In fact, even when accounting for a broad set of academic as well as nonacademic background variables, over 70 per cent of the gender gap is due to the fact

[12] That is, if boys and girls had the same index values as the average Canadian PISA student.

that girls have a greater index of reading enjoyment than boys do. That said, the paper's results show that it is very difficult to disentangle reading enjoyment from reading scores. It is therefore not obvious that policies narrowly aimed at encouraging boys to increase their enjoyment of reading activities (an abstract notion to begin with) will indeed narrow the gender gap in reading literacy by anything close to 70 per cent. Thus, policy efforts must be broad and focused on objectives.

Contrasting the results between model 2 and model 3 allowed for the conclusion that the factors *understand and remember*, *summarizing*, *control*, and especially *reading enjoyment*, play a *primary* role in explaining the potential for closing the gender gap while nonacademic and unobserved factors played only a secondary role at best. Recall that in model 2, there were no unobserved effects left over since the constant term (i.e., the baseline gap) was insignificant but also since the gap's unexplainable portion was insignificant as a whole. However, when reading enjoyment was abstracted from model 3, the baseline gap became significant due to unobserved effects now playing a role. But these unobserved effects can play only a *secondary* role in explaining the potential for closing the gender gap, since they took on significance only when reading enjoyment was abstracted from the framework. Thus, it is reading enjoyment that plays a *primary* role in explaining the gap, not the unobserved factors. Finally, since the three other factors (understanding and remembering, summarizing, and control) contributed persistently and significantly to the gap, they are considered to also play a *primary* role.

Thus, it is possible to conclude that the performance gap was primarily due to the fact that girls were more favourably endowed than boys in academic predisposition and motivation. There is no strong evidence that it was due to any differences that may exist between boys and girls in how they were *affected* by their respective endowments.

The determinants of the gender gap have been dissected with the utmost care and in great detail, and the results stand out as clearly as they possibly could within a relatively simple framework. This is particularly evident in the fact that the key coefficients maintained their statistical significance and relative magnitude regardless of the regression model specification. Indeed, whether the model took into account the nonacademic background of students, whether the index of reading enjoyment was completely dropped, or reading scores were adjusted for it, the results remained quite robust throughout. This in-depth analysis and its robust findings can thus be used as evidence in support of policy recommendations.

# Conclusion and Policy Implications

PISA 2009 shows quite important inequalities in reading between boys and girls for most OECD countries, and Canada in particular. Once the inequalities are measured, a possible next step would be to explain them. What enables girls to outperform boys? Why do inequalities persist despite educational systems' continuous efforts to eliminate them? Is there anything that girls do differently than boys that makes them successful in reading? Those were some of the questions this paper tried to answer.

The first part of this paper studied in detail two sets of factors that could potentially explain the gender gap in reading for Canada: engagement-in-reading activities and approaches to learning. The analysis shows that enjoyment of reading has the strongest association with reading ability and dominates other factors. However, this association cannot be interpreted univocally. Do students read better because they enjoy reading more, or do they enjoy reading more because their reading skills are better? Analyses do not provide us with clear insight into this relationship. It is obvious that enjoyment of reading cannot be a sufficient precondition for high reading-achievement scores among students, and there should be other factors that are more "approachable" and could be directly addressed by teachers and students in the classroom (e.g., reading strategies). For this reason, part 2 of the paper focuses on a special statistical technique (the Oaxaca-Blinder method) that allows researchers to decompose the gender gap into contributing factors and to reveal how different cognitive and meta-cognitive strategies might explain gender inequalities in reading.

The results obtained in this analysis were consistent across several regression models. Besides enjoyment of reading, two reading strategies showed significant and important contributions to the gender differences in reading: *control* and *summarizing*. *Control* is a cognitive strategy focusing on understanding a task's purpose and its main concepts, while *summarizing* is a meta-cognitive strategy reflecting an awareness of the most efficient ways to condense information. Another meta-cognitive strategy with a significant but less-powerful contribution was *understanding and remembering*. Girls outperformed boys in the use of these three important strategies. An interesting result was revealed for *memorization*, a cognitive strategy derived from the frequency with which students try to memorize the text (without special focus on understanding). It appears that *memorization*, being more frequently used by girls, has a negative effect on reading scores. Thus, if girls did not employ this technique as frequently, their reading performance would be even higher and, therefore, the gender gap would be even larger. As for *elaboration*, a cognitive strategy relying on connecting new information with prior knowledge, its contribution to differences in reading performance was rather negligible. This was also the only strategy that boys seemed to employ more often than girls.

The results presented here should be interpreted carefully because the data are based on students' self-reports. The factors' contribution to the gender gap may look different if students' strategies were measured more directly. With this reservation in mind, policy-makers could consider that:

- Nurturing an enjoyment of reading will not be enough for improving boys' skills in reading, if it is not accompanied by the understanding of how different reading strategies can be used to learn effectively (OECD, 2010b). Thus, educators may consider initiatives for increasing the attention paid to meta-cognition and control strategies, which are proven to be the most effective ones in this study.
- Self-awareness is a prerequisite for successful learning, especially in adult life (Wolverton, 2008). Thus, it is important to re-evaluate how to stimulate boys' and girls' awareness of their reading strategies, since some of these strategies may turn out to be inefficient (e.g., *memorization*) and could hinder learning rather than help it.

# References


Arnot, M., David, M., & Weiner, G. (1999). *Closing the gender gap: Post-war education and social change*. Cambridge, UK: Polity Press.

Boekaerts, M. (2009), Goal-directed behaviors in the classroom. In K.R. Wentzel and A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 105–122). New York: Routledge.

Booth, D., Elliott-Johns, S., & Bruce, F. (2009). Boys' literacy attainment: Research and related practice. A report prepared for the Ontario Ministry of Education. Retrieved from http://www.edu.gov.on.ca/eng/research/boys_literacy.pdf

Brochu, P., Gluszynski, T., & Cartwright, F. (2011). *Second report from the 2009 Programme for International Student Assessment*. Toronto: Council of Ministers of Education, Canada (CMEC).

Brozo, W.G. (2002). *To be a boy, to be a reader: Engaging teen and preteen boys in active literacy*. Newark, DE: International Reading Association.

Chandler, J., Lizotte, R., & Rowe, M. (1998). Adapting teaching methods to learners' preferences, strategies, and needs. *College ESL*, *8,* 48–69.

Clark, A., & Trafford, J. (1995). Boys into modern languages: An investigation of the discrepancy in attitudes and performance between boys and girls in modern languages. *Gender and Education, 7*(3), 315–325.

Davies, J., & Brember, I. (2001). A decade of change? Monitoring reading and mathematical attainment in year 6 over the first ten years of the education reform act. *Research in Education*, *65,* 31–40.

Engels, N., Aelterman, A., Van Petegem, K., & Schepens, A. (2004) Factors which influence the well-being of pupils in Flemish secondary schools. *Educational Studies, 30*(2), 127–143

Green, J.M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly, 29*(2), 261–297.

Griva, E., Alevriadou, A., & Semoglou, K. (2011). Identifying gender differences in reading preferences and strategies employed by Greek students: A socio-cognitive perspective. In the unpublished proceedings of the *UNESCO International Conference "Mapping the Gender Equality: Research and practices—The national and international perspective*." Nicosia, Cyprus.

Gunzelmann, B., & Connell, D. (2006). The new gender gap: Psychological, neurobiological, and educational perspectives. *Educational Horizons*, *84*(2), 94–101.

Gurian, M., Henley, P., & Trueman, T. (2001). *Boys and girls learn differently: A guide for teachers and parents*. San Francisco: Jossey Bass.

Gurian, M., & Stevens, S. (2004). With boys and girls in mind. *Educational Leadership, 6*(3), 21–26.

Jann, B. (2008). A Stata implementation of the Blinder-Oaxaca decomposition. *The Stata Journal, 8*(4), 453–479. Available at http://www.ssc.wisc.edu/~jmuniz/jann_oaxaca.pdf

Katz, H., & Sokal, L. (2003). Masculine literacy: One size does not fit all. *Reading Manitoba*, *24*(1), 4–8.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed.) (pp. 1270–1328). Newark, DE: International Reading Association.

Klecker, B.M. (2005). *The "gender gap" in NAEP fourth-, eighth-, and twelfth-grade reading scores across years*. Paper presented at the annual meeting of the Midwestern Association of Educational Research, Columbus, OH.

Marsh, J., & Millard, E. (2000). *Literacy and popular culture*. London: Paul Chapman.

Martino, W. (2008). Boys' underachievement: Which boys are we talking about? In *What works? Research into practice*. Research Monograph # 12. Toronto: Ontario Ministry of Education, Literacy and Numeracy Secretariat. Retrieved from http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/Martino.pdf.

Mullis, I.V.S., Martin, M. O., Foy, P., & Drucker, K.T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

Oakhill, J.V., & Petrides, A. (2007). Sex differences in the effects of interest on boys' and girls' reading comprehension. *British Journal of Psychology, 98*, 223–235.

OECD. (2002). *Reading for change: Performance and engagement across countries: Results from PISA 2000.* Paris: OECD.

OECD. (2007). *Understanding the brain: The birth of a learning science*. Paris: OECD.

OECD. (2009). *Equally prepared for life? How 15-year-old boys and girls perform in school*. Paris: OECD.

OECD. (2010a). *PISA 2009 results: What students know and can do — Student performance in reading, mathematics and science* (Vol. 1). Paris: OECD.

OECD. (2010b). *PISA 2009 results: Learning to learn — Student engagement, strategies and practices* (Vol. 3). Paris: OECD.

OFSTED [Office for Standards in Education, Children's Services, and Skills]. (1993). *Boys and English*. London: HMSO.

OFSTED. (2003). *Yes he can: Schools where boys write well*. London: OFSTED Publications.

Pirie, B. (2002). *Teenage boys and high school English.* Portsmouth, NH: Boynton/Cook.

Ryan, R.M., & Deci, E.L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K.R. Wentzel and A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 171–196). New York: Taylor Francis.

Ruble, D.N., Martin, C.L., & Berenbaum, S.A. (2006). Gender development. In W. Damon (series Ed.) and N.E Isenberg (vol. Ed.), *Handbook of child psychology* (6th ed., Vol. 3) (pp. 858–932). New York: Wiley.

Safford, K., O'Sullivan, O., & Barrs, M. (2004). *Boys on the margin: Promoting boys' literacy and learning at key stage 2.* London: Centre for Literacy in Primary Education.

Smith, M.W. & Wilhelm, J.D. (2002). *Reading don't fix no chevys: Literacy in the lives of young men*. Portsmouth, NH: Heinemann.

Smith, M.W., & Wilhelm, J.D. (2005). Boys and reading: It's more complicated than you think. *The California Reader*, *38*(4), 5–13.

Smith, M., & Wilhelm, J. (2006). *Going with the flow*. Portsmouth, NH: Heinemann.

Sokal L., Katz H., Adkins M., Gladu, A,. Jackson-Davis, K., & Kussin, B. (2005). Boys will be "boys": Variability in boys' experiences of literacy. *The Alberta Journal of Education Research*, *51*(3), 216–30.

Topping, K.J., Samuels, J., & Paul, T. (2008). Independent reading: The relationship of challenge, non-fiction and achievement. *British Educational Research Journal, 34*, 505–524.

Van de gaer, E., Pustjens, H., van Damme, J., & de Munter, A. (2007). Impact of attitudes of peers on language achievement: Gender differences. *Journal of Educational Research, 101*(2), 78–93.

Warrington, M., Younger, M., & Williams, J. (2000) *S*tudent attitude, image and the gender gap. *British Educational Research Journal*, *26*(1), 393–407.

Wilhelm, J., & Smith, M. W. (2009). Boys and literacy: Complexity and multiplicity. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research* (pp. 360–371). New York: Guilford Press.

Wolverton, L.L. (2008). *Self-regulated learning strategies of successful developmental reading students*. Doctoral dissertation, Capella University, October 2008.

Younger, M., & Warrington, M. (1996). Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school. *British Journal of Sociology of Education, 17*(3), 299–313.

# Annex 1

# Detailed regression tables

The following two tables contain the details of all the Oaxaca-Blinder decomposition models that the paper estimated. Not all of the models were mentioned in the text, but they are included here for the reader interested in such details. The alternate models (models 2b, 2c, 3b) were estimated to assess the sensitivity of certain coefficients to the model specification.

## Model summaries:

**Model 1**: the PISA Reading score is the explained variable. The eight academic factors are the explanatory variables.

**Model 2** is the same as model 1, but augmented with nonacademic control variables.

Model 2b is the same as model 2, but with reading enjoyment dropped from the list of explanatory variables.

Model 2c is the same as model 2, but with reading enjoyment, reading diversity, and on-line reading factors all dropped from the list of explanatory variables.

**Model 3**: The adjusted PISA reading score is the explained variable. All academic and nonacademic factors are included in the list of variables, except for the index of reading enjoyment.

Model 3b is the same as model 3, but with reading scores being adjusted for the student's grade as well as reading enjoyment.

**TABLE A1.1 Explained contributions**

| | Model 1 | | | Model 2 | | | Model 2b | | | Model 2c | | | Model 3 | | | Model 3b | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e |
| **Girls** | 543.83 | *** | 1.69 | 545.67 | *** | 1.72 | 545.44 | *** | 1.71 | 545.44 | *** | 1.70 | 4.70 | *** | 1.50 | 4.12 | *** | 1.44 |
| **Boys** | 511.71 | *** | 1.78 | 514.49 | *** | 1.77 | 514.60 | *** | 1.76 | 514.46 | *** | 1.75 | 2.89 | * | 1.70 | 2.67 | * | 1.60 |
| **Difference** | 32.12 | *** | 1.94 | 31.18 | *** | 1.95 | 30.84 | *** | 1.94 | 30.98 | *** | 1.93 | 1.81 | | 1.80 | 1.45 | | 1.73 |
| **Understand** | 2.35 | *** | 0.29 | 2.25 | *** | 0.27 | 2.82 | *** | 0.31 | 2.81 | *** | 0.30 | 2.08 | *** | 0.26 | 2.34 | *** | 0.26 |
| **Summarize** | 8.01 | *** | 0.48 | 7.17 | *** | 0.45 | 8.32 | *** | 0.49 | 8.53 | *** | 0.50 | 6.81 | *** | 0.44 | 6.84 | *** | 0.44 |
| **Control** | 7.94 | *** | 0.62 | 6.36 | *** | 0.52 | 7.56 | *** | 0.62 | 8.04 | *** | 0.65 | 5.98 | *** | 0.49 | 5.39 | *** | 0.46 |
| **Memorize** | -3.14 | *** | 0.31 | -2.63 | *** | 0.26 | -3.14 | *** | 0.29 | -3.04 | *** | 0.28 | -2.50 | *** | 0.25 | -2.32 | *** | 0.24 |
| **Elaborate** | 0.99 | *** | 0.24 | 0.78 | *** | 0.20 | 0.77 | *** | 0.20 | 0.62 | *** | 0.17 | 0.79 | *** | 0.20 | 0.71 | *** | 0.17 |
| **Reading enjoyment** | 22.15 | *** | 0.87 | 22.28 | *** | 0.85 | - | - | - | - | - | - | adj† | - | - | adj‡ | - | - |
| **Reading diversity** | -0.44 | * | 0.24 | 0.20 | | 0.25 | 2.55 | *** | 0.34 | - | - | - | -0.59 | ** | 0.24 | -0.42 | * | 0.23 |
| **On-line reading** | -0.11 | | 0.16 | -0.10 | | 0.12 | -0.09 | | 0.12 | - | - | - | -0.10 | | 0.12 | -0.07 | | 0.08 |
| **Student/teacher relation** | - | - | - | 0.31 | *** | 0.12 | 0.62 | *** | 0.18 | 0.68 | *** | 0.20 | 0.23 | ** | 0.10 | 0.29 | ** | 0.12 |
| **Student's library use** | - | - | - | -3.43 | *** | 0.30 | -2.26 | *** | 0.26 | -1.14 | *** | 0.26 | -3.81 | *** | 0.32 | -3.66 | *** | 0.33 |
| **Socioeconomic status** | - | - | - | -0.56 | | 0.36 | -0.63 | | 0.39 | -0.70 | | 0.44 | -0.55 | | 0.35 | -0.45 | | 0.28 |
| **2nd-generation student** | - | - | - | 0.00 | | 0.03 | 0.02 | | 0.03 | -0.01 | | 0.03 | -0.01 | | 0.03 | 0.03 | | 0.04 |
| **1st-generation student** | - | - | - | 0.06 | | 0.06 | 0.06 | | 0.07 | 0.02 | | 0.06 | 0.05 | | 0.06 | 0.06 | | 0.06 |
| **Home language** | - | - | - | 0.26 | ** | 0.11 | 0.28 | ** | 0.11 | 0.29 | ** | 0.11 | 0.25 | ** | 0.10 | 0.21 | ** | 0.09 |
| **Rural/urban status** | - | - | - | -0.06 | | 0.07 | -0.08 | | 0.07 | -0.07 | | 0.07 | -0.06 | | 0.07 | -0.05 | | 0.06 |
| **Explained - Total** | 37.77 | *** | 1.30 | 32.89 | *** | 1.38 | 16.79 | *** | 1.31 | 16.03 | *** | 1.19 | 8.58 | *** | 1.04 | 8.90 | *** | 0.99 |

Source: PISA 2009 data for Canada, calculations by author. Explained portion of gender gap for several explanatory variables and models.

†: Reading scores were adjusted for reading enjoyment. ‡: Reading scores were adjusted for reading enjoyment and grade.

Significance levels of 1% are denoted by ***, 5% by **, and 10% by *. All other coefficients are statistically insignificant.

**TABLE A1.2 Unexplained contributions**

| | Model 1 | | | Model 2 | | | Model 2b | | | Model 2c | | | Model 3 | | | Model 3b | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e | coef | p | s.e |
| **Girls** | 543.83 | *** | 1.69 | 545.67 | *** | 1.72 | 545.44 | *** | 1.71 | 545.44 | *** | 1.70 | 4.70 | *** | 1.50 | 4.12 | *** | 1.44 |
| **Boys** | 511.71 | *** | 1.78 | 514.49 | *** | 1.77 | 514.60 | *** | 1.76 | 514.46 | *** | 1.75 | 2.89 | * | 1.70 | 2.67 | * | 1.60 |
| **Difference** | 32.12 | *** | 1.94 | 31.18 | *** | 1.95 | 30.84 | *** | 1.94 | 30.98 | *** | 1.93 | 1.81 | | 1.80 | 1.45 | | 1.73 |
| **Understand** | -0.03 | | 0.03 | -0.02 | | 0.03 | -0.02 | | 0.03 | -0.02 | | 0.03 | -0.03 | | 0.03 | -0.03 | | 0.03 |
| **Summarize** | 0.17 | ** | 0.08 | 0.28 | *** | 0.10 | 0.30 | *** | 0.11 | 0.27 | ** | 0.11 | 0.24 | ** | 0.10 | 0.19 | ** | 0.10 |
| **Control** | -0.78 | *** | 0.26 | -0.83 | *** | 0.28 | -0.74 | *** | 0.27 | -0.80 | *** | 0.28 | -0.90 | *** | 0.29 | -0.77 | *** | 0.28 |
| **Memorize** | -0.05 | | 0.08 | -0.03 | | 0.07 | -0.09 | | 0.11 | -0.13 | | 0.11 | -0.04 | | 0.07 | -0.04 | | 0.06 |
| **Elaborate** | -1.31 | *** | 0.35 | -1.26 | *** | 0.35 | -1.15 | *** | 0.36 | -0.99 | *** | 0.33 | -1.30 | *** | 0.36 | -0.98 | *** | 0.36 |
| **Reading enjoyment** | -0.16 | | 0.19 | -0.29 | | 0.21 | - | - | - | - | - | - | adj† | - | - | adj‡ | - | - |
| **Reading diversity** | 0.08 | | 0.13 | 0.00 | | 0.13 | -0.08 | | 0.13 | - | - | - | 0.13 | | 0.12 | 0.13 | | 0.12 |
| **On-line reading** | 0.08 | | 0.08 | 0.15 | | 0.11 | 0.18 | | 0.12 | - | - | - | 0.14 | | 0.10 | 0.12 | | 0.10 |
| **Student/teacher relation** | - | - | - | -0.17 | | 0.59 | -0.36 | | 0.60 | -0.54 | | 0.59 | -0.20 | | 0.57 | -0.17 | | 0.57 |
| **Student's library use** | - | - | - | 0.06 | | 0.30 | 0.18 | | 0.29 | 0.14 | | 0.27 | -0.05 | | 0.31 | -0.10 | | 0.29 |
| **Socioeconomic status** | - | - | - | 2.20 | ** | 1.00 | 2.55 | ** | 1.03 | 1.40 | | 1.02 | 2.01 | ** | 1.01 | 2.28 | ** | 1.00 |
| **2nd-generation student** | - | - | - | 1.42 | ** | 0.69 | 1.29 | * | 0.70 | 0.96 | | 0.69 | 1.47 | ** | 0.69 | 1.51 | ** | 0.69 |
| **1st-generation student** | - | - | - | 1.08 | | 0.79 | 1.14 | | 0.77 | 0.80 | | 0.75 | 1.07 | | 0.79 | 1.35 | * | 0.77 |
| **Home language** | - | - | - | -0.03 | | 0.63 | -0.54 | | 0.66 | -0.39 | | 0.65 | 0.12 | | 0.63 | 0.12 | | 0.63 |
| **Rural/urban status** | - | - | - | -0.27 | | 0.91 | 0.02 | | 0.90 | 0.12 | | 0.91 | -0.32 | | 0.92 | -0.10 | | 0.90 |
| **Constant** | -3.62 | *** | 1.62 | -3.99 | * | 2.30 | 11.36 | *** | 2.40 | 14.14 | *** | 2.39 | -9.10 | *** | 2.35 | -10.98 | *** | 2.31 |
| **Unexplained - Total** | -5.64 | *** | 1.56 | -1.72 | | 1.51 | 14.04 | *** | 1.54 | 14.96 | *** | 1.51 | -6.77 | *** | 1.47 | -7.45 | *** | 1.41 |

Source: PISA 2009 data for Canada, calculations by author. Unexplained portion of gender gap for several explanatory variables and models.

†: Reading scores were adjusted for reading enjoyment. ‡: Reading scores were adjusted for reading enjoyment and grade.

Significance levels of 1% are denoted by ***, 5% by **, and 10% by *. All other coefficients are statistically insignificant.

# Annex 2

**TABLE A2.1 Explaining the gender gap in Canada**

| | Summary Coefficients | |
|---|---|---|
| Girls | **545.35** | ... |
| Boys | **514.32** | ... |
| Difference | **31.03** | ... |
| Explained | **17.42** | ... |
| Unexplained | **13.61** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.85** | -0.02 |
| Summarize | **8.75** | **0.25** |
| Control | **8.61** | **-0.80** |
| Memorize | **-3.00** | -0.13 |
| Elaborate | **0.66** | **-1.03** |
| Socioeconomic status | -0.71 | 1.53 |
| 2nd generation | 0.00 | 1.02 |
| 1st generation | 0.04 | 0.73 |
| Home language | **0.28** | 1.70 |
| Urban | -0.06 | -0.77 |
| Constant | ... | **11.13** |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.2 Explaining the gender gap in Finland**

| | Summary Coefficients | |
|---|---|---|
| Girls | **564.53** | ... |
| Boys | **511.89** | ... |
| Difference | **52.64** | ... |
| Explained | **23.35** | ... |
| Unexplained | **29.29** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **7.60** | 0.48 |
| Summarize | **14.59** | 0.08 |
| Control | **3.41** | **3.25** |
| Memorize | **-2.17** | **-1.57** |
| Elaborate | -0.16 | -0.81 |
| Socioeconomic status | 0.54 | 0.02 |
| 2nd generation | 0.01 | -0.22 |
| 1st generation | -0.24 | 0.19 |
| Home language | -0.22 | -7.56 |
| Urban | 0.00 | -4.16 |
| Constant | ... | ... |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.3 Explaining the gender gap in the United States**

| | Summary Coefficients | |
|---|---|---|
| Girls | **518.10** | ... |
| Boys | **494.54** | ... |
| Difference | **23.56** | ... |
| Explained | **11.02** | ... |
| Unexplained | **12.54** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.61** | **-1.21** |
| Summarize | **6.82** | 0.19 |
| Control | **6.62** | 0.33 |
| Memorize | **-3.91** | -0.20 |
| Elaborate | 0.21 | **-0.60** |
| Socioeconomic status | -1.68 | -0.79 |
| 2nd generation | 0.05 | -0.65 |
| 1st generation | 0.22 | -0.11 |
| Home language | 0.02 | 4.23 |
| Urban | 0.06 | -0.47 |
| Constant | ... | 11.82 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.4 Explaining the gender gap in the OECD**

| | Summary Coefficients | |
|---|---|---|
| Girls | **514.21** | ... |
| Boys | **484.28** | ... |
| Difference | **29.93** | ... |
| Explained | **12.29** | ... |
| Unexplained | **17.64** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.97** | **-0.10** |
| Summarize | **7.23** | -0.05 |
| Control | **4.41** | **0.14** |
| Memorize | **-1.96** | **-0.16** |
| Elaborate | **0.33** | **-0.11** |
| Socioeconomic status | -0.73 | 0.16 |
| 2nd generation | 0.00 | -0.33 |
| 1st generation | -0.01 | -0.13 |
| Home language | **0.10** | -1.09 |
| Urban | -0.04 | -2.75 |
| Constant | ... | **22.05** |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.5 Explaining the gender gap in Newfoundland and Labrador**

| | Summary Coefficients | |
|---|---|---|
| Girls | **529.68** | ... |
| Boys | **487.06** | ... |
| Difference | **42.62** | ... |
| Explained | **15.22** | ... |
| Unexplained | **27.39** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | 2.05 | 0.33 |
| Summarize | **7.38** | 0.12 |
| Control | **8.51** | -0.86 |
| Memorize | **-3.72** | 0.61 |
| Elaborate | 0.42 | -0.42 |
| Socioeconomic status | -0.43 | 2.56 |
| 2nd generation | 0.01 | 0.13 |
| 1st generation | -0.13 | -0.20 |
| Home language | 0.24 | **-110.66** |
| Urban | 0.89 | 3.05 |
| Constant | ... | **132.73** |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.6 Explaining the gender gap in Prince Edward Island**

| | Summary Coefficients | |
|---|---|---|
| Girls | **517.09** | ... |
| Boys | **474.28** | ... |
| Difference | **42.81** | ... |
| Explained | **27.54** | ... |
| Unexplained | **15.26** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **5.28** | 0.24 |
| Summarize | **10.73** | 1.06 |
| Control | **13.16** | -0.34 |
| Memorize | -2.13 | -0.29 |
| Elaborate | -0.29 | -1.32 |
| Socioeconomic status | 0.33 | 3.12 |
| 2nd generation | -0.49 | -0.16 |
| 1st generation | -0.41 | -0.29 |
| Home language | 1.06 | -20.84 |
| Urban | 0.30 | -4.28 |
| Constant | ... | 38.35 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.7 Explaining the gender gap in Nova Scotia**

| | Summary Coefficients | |
|---|---|---|
| Girls | **534.59** | ... |
| Boys | **509.39** | ... |
| Difference | **25.20** | ... |
| Explained | **14.91** | ... |
| Unexplained | **10.29** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **3.30** | 0.52 |
| Summarize | **8.02** | -0.31 |
| Control | **8.90** | -0.40 |
| Memorize | **-3.76** | 0.20 |
| Elaborate | 0.06 | 0.24 |
| Socioeconomic status | -1.63 | 3.56 |
| 2nd generation | -0.08 | -0.65 |
| 1st generation | -0.18 | -0.15 |
| Home language | 0.52 | -12.89 |
| Urban | -0.23 | 4.30 |
| Constant | ... | 15.87 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.8 Explaining the gender gap in New Brunswick**

| | Summary Coefficients | |
|---|---|---|
| Girls | **520.54** | ... |
| Boys | **487.25** | ... |
| Difference | **33.29** | ... |
| Explained | **15.05** | ... |
| Unexplained | **18.24** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **3.55** | -0.02 |
| Summarize | **8.98** | -0.21 |
| Control | **10.83** | -0.08 |
| Memorize | **-6.39** | 0.04 |
| Elaborate | -0.25 | -1.20 |
| Socioeconomic status | -0.48 | **4.82** |
| 2nd generation | 0.00 | -0.61 |
| 1st generation | 0.21 | 0.16 |
| Home language | **-1.28** | -18.33 |
| Urban | -0.12 | -4.50 |
| Constant | ... | **38.15** |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.9 Explaining the gender gap in Quebec**

| | Summary Coefficients | |
|---|---|---|
| Girls | **542.19** | ... |
| Boys | **514.44** | ... |
| Difference | **27.75** | ... |
| Explained | **16.04** | ... |
| Unexplained | **11.71** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **3.30** | 2.09 |
| Summarize | **5.50** | -1.83 |
| Control | **8.48** | -0.20 |
| Memorize | **-2.24** | -0.55 |
| Elaborate | **1.45** | -0.61 |
| Socioeconomic status | -1.05 | 0.54 |
| 2nd generation | 0.13 | -0.43 |
| 1st generation | 0.26 | -0.55 |
| Home language | 0.19 | -5.32 |
| Urban | 0.00 | 6.31 |
| Constant | ... | 12.25 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.10 Explaining the gender gap in Ontario**

| | Summary Coefficients | |
|---|---|---|
| Girls | **551.86** | ... |
| Boys | **520.10** | ... |
| Difference | **31.76** | ... |
| Explained | **18.81** | ... |
| Unexplained | **12.94** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **3.05** | -0.05 |
| Summarize | **9.44** | 0.50 |
| Control | **8.81** | **-1.74** |
| Memorize | **-2.96** | 0.18 |
| Elaborate | 0.46 | **-1.29** |
| Socioeconomic status | -0.31 | 0.97 |
| 2nd generation | 0.05 | **3.23** |
| 1st generation | 0.02 | 1.55 |
| Home language | 0.38 | 7.47 |
| Urban | -0.12 | -4.96 |
| Constant | ... | 7.09 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.11 Explaining the gender gap in Manitoba**

| | Summary Coefficients | |
|---|---|---|
| Girls | **515.47** | ... |
| Boys | **488.81** | ... |
| Difference | **26.66** | ... |
| Explained | **15.76** | ... |
| Unexplained | **10.90** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.93** | -0.05 |
| Summarize | **7.59** | -0.79 |
| Control | **9.66** | -0.37 |
| Memorize | **-2.84** | -0.23 |
| Elaborate | 1.15 | -2.21 |
| Socioeconomic status | -2.54 | 0.45 |
| 2nd generation | 0.24 | 0.65 |
| 1st generation | -0.19 | -0.22 |
| Home language | -0.23 | -14.71 |
| Urban | -0.02 | -8.14 |
| Constant | ... | 36.54 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.12 Explaining the gender gap in Saskatchewan**

| | Summary Coefficients | |
|---|---|---|
| Girls | **527.97** | ... |
| Boys | **490.83** | ... |
| Difference | **37.15** | ... |
| Explained | **22.18** | ... |
| Unexplained | **14.97** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.72** | **-2.57** |
| Summarize | **9.97** | -0.37 |
| Control | **11.68** | 0.55 |
| Memorize | **-2.51** | 0.04 |
| Elaborate | 0.68 | -2.32 |
| Socioeconomic status | 0.29 | 0.49 |
| 2nd generation | -0.16 | 0.29 |
| 1st generation | 0.07 | 1.31 |
| Home language | -0.28 | 34.43 |
| Urban | -0.29 | -1.82 |
| Constant | ... | -15.06 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.13 Explaining the gender gap in Alberta**

| | Summary Coefficients | |
|---|---|---|
| Girls | **552.80** | ... |
| Boys | **521.32** | ... |
| Difference | **31.47** | ... |
| Explained | **15.79** | ... |
| Unexplained | **15.69** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **3.15** | 0.14 |
| Summarize | **9.01** | 0.04 |
| Control | **7.23** | -0.20 |
| Memorize | **-3.25** | -1.06 |
| Elaborate | **0.82** | -0.14 |
| Socioeconomic status | -0.67 | 6.09 |
| 2nd generation | -0.46 | 1.57 |
| 1st generation | 0.01 | 0.92 |
| Home language | 0.01 | -3.06 |
| Urban | -0.08 | 5.32 |
| Constant | ... | 6.06 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.

**TABLE A2.14 Explaining the gender gap in British Columbia**

| | Summary Coefficients | |
|---|---|---|
| Girls | **546.01** | ... |
| Boys | **513.71** | ... |
| Difference | **32.31** | ... |
| Explained | **19.38** | ... |
| Unexplained | **12.93** | ... |
| | **Explained / endowments** | **Unexplained / coefficients** |
| Understand | **2.33** | -0.41 |
| Summarize | **12.33** | -0.04 |
| Control | **6.28** | -0.52 |
| Memorize | **-1.90** | -1.18 |
| Elaborate | 0.25 | -0.54 |
| Socioeconomic status | -0.52 | 0.71 |
| 2nd generation | 0.00 | -2.30 |
| 1st generation | 0.21 | 1.39 |
| Home language | 0.56 | -4.55 |
| Urban | -0.16 | -3.83 |
| Constant | ... | 24.19 |

Source: PISA 2009 data; calculations by author. Bold figures are statistically significant at 5%.