

ARE GIRLS BETTER READERS THAN BOYS? WHICH BOYS? WHICH GIRLS?

Bozena White
Queen's University

Using data from the reading component of the Ontario Secondary School Literacy Test (N = 113,050), the effects of gender and curricular track for nine sub-scores of reading achievement were investigated. Only students indicating that they did not receive additional programming support were included in the analysis. Gender accounted for less than one per cent of variance in reading achievement. Gender differences for each curricular track were in the close-to-zero and small range. The results suggest that any observed differences may be of little practical consequences, and that the notion of under-achievement of boys' reading achievement has been greatly overstated.

Key words: reading achievement, gender, curricular track, Ontario Secondary School Literacy Test

Utilisant des données tirées du volet lecture du Test de compétences linguistiques de l'Ontario au secondaire (N = 113 050), l'auteure analyse les effets du genre et de la répartition en classes homogènes sur neuf sous-scores ayant trait à la lecture. Seuls les élèves ayant indiqué qu'ils n'avaient pas reçu un soutien supplémentaire ont été inclus dans l'analyse. Le sexe représentait moins de 1 % de l'écart dans les résultats. Les différences selon le sexe pour chaque groupement selon les aptitudes s'approchaient de zéro ou étaient très faibles. Les résultats semblent indiquer que toute différence observée peut n'avoir que peu de conséquences pratiques et que la notion de sous-performance des garçons en matière de lecture a été grandement exagérée.

Mots clés : rendement en lecture, genre, répartition en classes homogènes, écoles secondaires de l'Ontario

Reading is regarded as a fundamental skill necessary for personal learning and intellectual growth. In an increasingly interdependent global world, a literate population is essential not only for a nation's economic but also its social development. The need for government bodies to monitor and to encourage the development of this skill in the form of large-scale standardized assessments is increasingly evident at provincial, national, and international levels. Information obtained from these assessments should, in theory, provide data to both policymakers and educators as to how well their students read. Whether this information will, in turn, appropriately inform methods for improving literacy and reading achievement is perhaps less certain.

BACKGROUND

Based on recent results from large-scale reading assessments, the present researcher's concern relates to the consistent observation that girls, on average, surpass boys in their reading abilities. At the international level, girls have been reported to have surpassed boys in both the 1991 International Association's Evaluation of Educational Achievement (IEA) Reading Literacy Study of 9- and 14-year olds (Elley, 1992), and in the 2001 Programme for International Student Assessment (PISA) of 15-year olds (OECD, 2001). In the United States, a comparison of gender differences in the 2002 and 2003 National Assessment in Educational Progress (NAEP, 2004) indicated that at grade 8, the average score for boys declined while girls' scores increased. At national levels, the Canadian Council of Ministers of Education (CMEC, 1999) reported gender differences in literacy at two age levels: 13-year-old and 16-year-old girls consistently outperformed boys in reading test scores. In Ontario (Education and Quality and Accountability Office [EQAO], 2003), the results indicated that boys not only have an overall lower mean than do girls, but also have a higher chance of failing the reading component of the grade-10 literacy test than do girls. Given the importance of reading with regard to educational and individual development, both within school and in later in life (OECD, 2001), it is not surprising that concern regarding the purported gender gap in reading achievement, what might explain it, and how best to respond to it, appear to be widespread. Indeed, fuelled by media attention, the

current status of boys' under-achievement has been likened to a kind of globalized moral panic (e.g., Epstein, Elwood, Hey, & Maw, 1998).

Anxiety regarding the purported gender gap in reading achievement has not been limited to the general public as evidenced by media headlines, or by the growth in pop psychology books (e.g., CBC, 2005; Smith & Wilhelm, 2002). Reports from researchers who have analyzed data from international large-scale assessments have not only suggested that a closer examination of the overall lower reading achievement of boys is merited (Elley, 1992), but have gone so far as to suggest that "special intervention targeted to males is indicated" (Topping, Valtin, Roller, Brozo, & Dionisio, 2003, p. 11). In Ontario, a recent Ministry resource (Ontario Ministry of Education, 2005) for teachers entitled *Me Read? No Way! A Practical Guide to Improving Boys Literacy* includes an appeal to all educators to

share the common goal of providing equitable learning opportunities for all students, and that while providing equitable opportunities for girls is a familiar topic, providing them for boys is a relatively recent issue, but one that is appearing with increasing urgency on education agendas around the world. (p. 4)

SUGGESTED EXPLANATIONS AND PROPOSED STRATEGIES

Numerous populist explanations (biological and socio-cultural) have been offered for the proposed gender differences in reading. Considerable overlap occurs among these explanations and the solutions offered up. In many ways neither the explanations nor the proposed solutions are new (e.g., Ayres, 1909; Cohen, 1998; Maccoby & Jacklin, 1974). Recent populist explanations often draw on biological theories that emphasize that gender differences, in favour of girls, are rooted in the differential brain wiring, maturation rates, and chemistry of boys (For a review and critique of these theories see Alloway, Freebody, Gilbert, & Musprett, 2002). They are based on a belief that boys and girls are so biologically different that they require specific gender strategies to ameliorate the detrimental effects of what are considered feminized educational structures and practices (e.g., Sommers, 2000). A number of suggested strategies to mediate the gender gap have been

advanced (e.g., Gurian, Henley, & Trueman, 2001; Noble & Bradford, 2000) including the use of boy-friendly reading materials, the introduction of more male role models and teachers, adoption of technology-based programmes, and experimentation with single-gender schooling. Each of these is targeted at changing educational and professional practices to better meet what is perceived as the particular needs of boys. To illustrate, one Ontario superintendent of education in Ontario has suggested: "Our system has been based on passive learning that has suited girls more than boys...a focus on fiction engages girls more than boys. To engage boys we need more manuals and techie stuff" (Miller, 2003, n. p.).

CRITICISMS OF PROPOSED STRATEGIES

A number of national and international authors have provided critiques of the proposed strategies (e.g., Gilbert & Gilbert, 1998; Kehler & Greig, 2005; Martino & Berrill, 2003; Mills, 2003). The proposed strategies have been characterized as quick-fix solutions that suggest simplistic strategies for extremely complex problems, that they are not based on sufficient empirical evidence as to their effectiveness, and that their implementation may lead to unintended negative consequences for boys, and/or for girls. Recommendations of increasing the reading materials that are better suited to the natural interests of boys have been criticized for encouraging a narrowly focused recovery effort for boys that relies on essentialist notions of what it means to be a boy (Anderson & Accomando, 2002).

Both the nature of the explanations and proposed strategies represent a curious situation because they appear to fly in the face of cautionary inclusions included in recent international reading study reports warning against using the results in their reports to make simple causal inferences between a particular factor (i.e., gender) and student achievement (e.g., National Assessment of Educational Progress, 2005). Results from a recent report provide some evidence that the crisis of boys' under-achievement in reading may simply be overstated, and that much of the pessimism about young males seems to derive from inadequate research, poor analysis, and discomfort with the relative position of the sexes (Mead, 2006).

AIM OF PRESENT STUDY

Given the current attention to the issues surrounding the gender debate in reading, it is important to investigate more fully the extent to which the truth value of the premise underlying much of the present discourse may be valid. In short, this study seeks to address the question: Are girls better readers than boys? I have addressed this question by utilizing reading achievement data derived from the province of Ontario's 2002 large-scale administration of the Ontario Secondary School Literacy Test (OSSLT). By examining in more detail the extent to which girls might be better readers than boys, I provide evidence to refute the explanations for the purported gender gap in reading achievement, as well as justification for the proposed gender specific strategies that follow from these explanations. In this manner, I have provided some assistance to allow educators to move beyond the existing parameters of gender-specific strategies, and to move towards more productive discussions regarding how reading achievement might be improved for all students.

METHODOLOGICAL CONSIDERATIONS

Part of the difficulty when reporting gender differences in reading achievement may rest with the methodology most commonly utilized to analyze large-scale assessment data. This method often utilizes a comparison of the means for the total population of each gender rather than on a comparison of differences within specific sub-groups (e.g., social economic variables, exposure to literate activities). In this respect, only one single background variable, namely gender, has been utilized and only one outcome measure (overall test score) has been most commonly reported. Less frequently, it appears, have researchers attempted to include several background variables. Although there are many challenges and pitfalls in comparing groups, what is learned by these comparisons depends, in no small way, on an adequate understanding of the degree and the context in which each group may, or may not, be unique. Without these understandings, data can easily be misinterpreted, and the generalization derived can be oversimplified.

Because researchers most commonly report standardized assessment findings based on the overall score achieved on a particular reading assessment, some researchers have attempted to investigate whether

gender differences exist in sub-sets of scores. A methodology that has been utilized to examine gender achievement in reading is to compare the results of girls and boys across the types of texts, or reading domains, in which a particular reading task occurs. Assessments differ not only in the names, and methods of classification they ascribe to these domains, but also in the weight they accord to each domain when calculating an overall reading score. In the Programme for International Student Assessment (PISA) reading study, for example, a distinction between non-continuous and continuous texts is made (OECD,1999). Continuous texts are formed of sentences and arranged in paragraphs, intended to be read from beginning to end (i.e., argumentative, descriptive, expository, narrative). Non-continuous texts are not defined by content or intention, but instead by structure (i.e., charts, forms, maps, schematics, and tables). In addition to written words, this type of text often includes spatial and numerical content – an area in which males have been often found to exceed (e.g., Halpern, 1992). Some evidence supports the notion that particular text types may accentuate or attenuate the purported gender differences in reading achievement. In Elley's (1992) analysis of the International Evaluation of Educational Achievement (IEA) Reading Literacy Study, girls overall were found to have a lesser advantage for documents compared to that of narrative and expository texts. Using data from the PISA (2000) reading study of Nordic countries, Lie, Linnakylä, and Roe (2003) found gender differences (favouring females) were much greater for narrative texts than they were for descriptive and expository texts. Although previous research has suggested that text type may be a factor to explain and/or to qualify gender differentials in reading performance, this suggestion may be dependent upon the specific population under consideration. Wagemaker (1996), for example, did not find that gender differences for sub-scores derived from different domains (e.g., narrative versus documents) were invariant when cross-cultural comparisons were made.

Current large-scale assessments of reading may also differentially emphasize what has been referred alternatively as reading processes, skills, strategies, and/or aspects (Mullis, Martin, & Gonzalez, 2004; OECD PISA, 2001). Each aspect is meant to represent a certain way of reading and responding to a text. Roe and Taube's (2003, p. 29) study of

Nordic countries involved in the PISA study indicated that boys are not outperformed to such an extent on the retrieve scale as they are on the reflect scale.

Rowe (2000, p. 14) points out that most of the apparent effort in relation to large-scale assessments has been focused on the measurement of students' achievements, rather than providing information regarding the sources of variability. Although there is documentation of gendered differences in reading achievement, as well as attitude, choice, and response for some boys (e.g., Millard, 1997), considerable observable evidence also suggests that such is not the case for all boys. Maccoby's (1990, p. 513) synthesis of decades of research on gender differences led her to claim that even when consistent differences between males and females were found, the amount of variance accounted for by sex was small, relative to the amount of variation within each sex. It has been repeatedly pointed out that boys are more different than alike, and that statistics lose sight of individual differences (e.g., Epstein et al., 1998).

Part of the difficulty may rest with treating either boys or girls as a uniform demographic group despite differences resulting from a variety of background characteristics. For example some studies have suggested that it is especially boys from low socio-economic groups, or from particular racial or ethnic groups (e.g., Luke, Freebody, & Land, 2000; Alloway & Gilbert, 1997), that are most at risk of literacy failure. Perhaps owing to the concerns expressed regarding the collection of racial information (Frank, 2005), these kinds of demographic information are not always collected. In addition, in a multi-cultural province such as Ontario, such efforts may be unwieldy. One demographic variable for which information has been collected in Ontario's reading assessment of high school students is the level of study in which students are enrolled. Curricular sub-groups of students have been most often thought of as tracks or streams where students are grouped by ability or achievement level for subjects (Oakes, 1985). Using this variable may not only assist in identifying which boys may be more at risk for poor reading achievement, but may also provide the opportunity to determine whether there concurrently exists a group of girls who may also be at risk (Teese, Davies, Charlton, & Polesel, 1995).

Based on his analysis of IEA data, Thorndike (1973) concluded that any gender differences that were observed were so small that they were not worthy of further consideration. Findings of statistical significance, therefore, represent only a first step in attempting to address the issue of gender differences in reading achievement. Cohen (1995) has emphasized that:

...the ritual of nil hypothesis testing has so dominated...research practice that it has inhibited our interest in the magnitude of the phenomena we study and the units in which they are measured, the basic stuff of which quantitative sciences are made. (p. 1103)

Statistical significance, therefore, does not automatically equate to substantive or practical effect – some statistically significant effects may be found to be meaningful while others may not. This observation is particularly relevant when using large data sets because findings of positive statistical findings can be found when even small numerical differences exist. As a result, measures of effect-size are recommended, allowing researchers to characterize "the magnitude of an effect or the strength of a relationship" (American Psychological Association [APA], 2001, p. 25). Many effect size indices address the magnitude of the difference between groups, or the relationship between variables. In the case of the former, differences are typically interpreted based on standard deviation units (e.g., one group's scores are 0.25 standard deviation units greater than those of the other group). In the case of the latter, differences are typically interpreted in terms of per cent of variance accounted for (e.g., variable X accounts for 25 per cent of the variance in variable Y). Although the interpretation of effect sizes remains a subjective endeavor, guidelines for interpreting the magnitude of effect sizes have been provided (Cohen, 1988). As a descriptive measure, the calculation of effect sizes would be helpful to address whether girls' and boys' reading achievement is more alike than different. This may assist in qualifying the extent to which the claims regarding boys' under-achievement in reading achievement may be overstated, and whether recommendations aimed at targeting resources towards improving their reading achievement are justified.

THE READING COMPONENT OF THE 2002 ONTARIO SECONDARY SCHOOL LITERACY TEST (OSSLT)

Study Objective

Using archival data from a large scale government mandated reading comprehension assessment, the objective of the present study addresses the question: Are girls better readers than boys? The methodology compares the effects of gender to that of level of study (or stream), and contrasts between-group differences to within-group differences. Level of study is used to address not only which boys, but also which girls may be at risk for poor reading achievement. I have used nine sub-scores of reading achievement to investigate the extent to which task characteristics assessed within particular text types might differentiate both between-group and within-group differences.

METHOD

Archival Data Set

I used a data set obtained from Ontario's Education Quality and Accountability Office (EQAO) in the analyses. EQAO, an independent agency of the Government of Ontario, has a mandate to evaluate and report on the quality of education in Ontario schools. EQAO develops and administers several province-wide tests (mathematics, reading and writing) at the primary, middle, and secondary school level. These tests are designed to measure student achievement against curriculum expectations. Results of these tests yield individual, school, school board, and provincial data on student achievement that help guide improvement planning (EQAO, 2006a).

All Ontario grade-10 students are required to complete the Ontario Secondary School Literacy Test (OSSLT), containing both a reading and writing component. There were a total of 146,539 students in grade 10 who were first-time eligible (FTE) to write the OSSLT; of these 1,637 were exempt (EQAO, 2003). The data file (N=132,234) included all grade-10 FTE students in the province of Ontario, and for whom complete 2002 OSSLT data for the reading component were available.

Subject Inclusion/Exclusion Criteria

A student questionnaire, included during the sitting of the OSSLT, required students to specify the level of study in which they were registered for the purposes of attaining their English credit. In 2002, English as a compulsory credit in Ontario was offered at the Academic and Applied levels where students were segregated into separate classes, and curriculum delivery and content were differentiated. Students, although enrolled in one of these tracks, but who indicated on the questionnaire that they were receiving additional programming, or differentiated support (e.g., Special Education Identification, English as a Second Language) were excluded. The importance of addressing questions relating to the performance of these particular groups of students is not considered to be trivial. However, it is not the question of interest in this study. A total of 113,050 students were retained in the final data set (Academic N = 90,185; Applied N = 22,865) representing 77.2 per cent of all FTE students.

OSSLT Instrument Measures, Materials, and Procedures

The OSSLT reading component required students to respond to 100 questions comprising three formats (multiple choice, short answer, and short answer with explanation).¹ Three reading skills were assessed: Skill 1 – directly stated ideas and information; Skill 2 – indirectly stated ideas and information; Skill 3 – connections between personal experiences and the ideas and information found in a selection. Each skill was assessed within the context of three text types: Text 1 – Informative (explanation, opinion); Text 2 – Graphic² (graph, schedule, instructions); Text 3 – Narrative (story, dialogue). The data set provided information that enabled the researcher to calculate a sub-score for each of the nine variables used in the analysis (3 Texts x 3 Reading Skills). Owing to differential weightings that were attached to the types of Texts and Skills in calculating the final overall score, the sub-scores were converted into percentages and were used, as such, in the analysis that follows.³ The OSSLT was administered in monitored, test-like conditions over a two day period during the fall of 2002. Four booklets containing a total of nine texts varying in length from two paragraphs to two pages were included in the assessment. The data are used to determine successful

attainment of the literacy test (pass score 60%) which is considered a requirement⁴ of attaining a secondary school diploma (EQAO, 2006b).

Statistical Data Analysis

An examination of the data concluded that multivariate assumptions were sufficiently met (Tabachnick & Fidell, 1996).⁵ A 2 x 2 MANOVA was conducted to establish whether there were overall (interaction, main) effects present on the nine outcome measures (three reading skills (direct, indirect, connections) assessed within the context of three text types (Informative, Graphic, Narrative). Classical eta-square (η^2) was used to calculate the proportion of total variation attributable to the two factors (gender and level of study) (Pierce, Block, & Aguinis, 2004). Follow up t-tests were conducted and descriptive statistics were used to determine for which outcome variables the groups differed. Cohen's *d* was used to report effect sizes.

RESULTS

A 2 (Gender) x 2 (Level of Study) multivariate analysis with nine outcome variables was carried out. Predictors were gender (male and female) and level of study (Academic and Applied). Outcome variables were the nine sub-scores (three reading Skills assessed within the context of three Texts). Means and standard deviations are reported in Table 1. The preliminary results revealed significant multivariate effects for Gender, Wilks' Lambda = .994, $F(9, 113038) = 73.383$, $p < .001$, $\eta^2 = .006$, Level of Study, Wilks' Lambda = .771, $F(9, 113038) = 3739.292$, $p < .001$, $\eta^2 = .229$ and Gender x Level of Study, Wilks' Lambda = .998, $F(9, 113038) = 20.006$, $p < .001$, $\eta^2 = .002$. All three multivariate test statistics – Pillai's Trace, Hotelling's Trace, Roy's Largest Route – were found to be significant at the $p < .001$ level.

Results from the MANOVA indicated that gender accounted for less than 1 per cent of the reading achievement ($\eta^2 = .006$). Follow-up between-subjects effects for gender showed girls' average performance to be superior for all nine outcome variables (Table 1, Column 1). Girls were found to significantly outperform males ($p < .005$) for six outcome variables. The power of statistical tests with sample sizes as large as the present one is extremely high; yet, no significant gender differences were

found for Skill 1 – directly stated ($p = .118$) and Skill 3 – connections ($p = .087$) assessed within Informative Texts and Skill 2 – indirectly stated assessed within Graphic Texts ($p = .193$)

TABLE 1
Sub-Group Gender Means and Standard Deviations (in parentheses) for
Nine Dependent Variables (three reading skills assessed within three text types)

Text Type	Skill	Gender		Level of Study		Academic		Applied	
		All Males N = 54,556 (1)	All Females N = 58,494 (1)	Academic** n = 90,185 (2)	Applied n = 22,865 (2)	Males n = 40,883 (3)	Females n = 49,302 (3)	Males n = 13,673 (4)	Females n = 9,192 (4)
Informative	1	71.4 (13.4)	72.7 (13.4)	74.5 (12.3)	62.3 (13.0)	74.3 (12.2)	74.7** (12.4)	62.6** (13.1)	61.8 (13.0)
	2	68.7 (15.6)	69.8** (15.4)	72.5 (13.6)	56.6 (16.2)	72.4 (13.5)	72.5 (13.7)	57.5** (16.2)	55.4 (16.1)
	3	57.4 (17.3)	59.5 (17.2)	61.4 (16.0)	47.0 (17.6)	60.8 (15.8)	61.9** (16.0)	47.3* (17.7)	46.6 (17.4)
Narrative	1	80.9 (16.8)	82.5** (16.1)	84.3 (14.7)	71.7 (18.9)	84.0 (14.8)	84.5** (14.7)	71.6 (19.0)	71.7 (18.6)
	2	71.0 (17.0)	73.9** (15.8)	75.8 (14.3)	59.4 (17.8)	75.0 (14.6)	76.4** (14.0)	59.0 (18.1)	60.0** (17.3)
	3	59.2 (19.4)	62.0** (19.0)	64.1 (17.8)	47.1 (18.8)	63.3 (17.8)	64.8** (17.7)	47.0 (19.0)	47.2 (18.6)
Graphic	1	76.5 (18.3)	77.3* (17.8)	79.9 (16.2)	65.1 (20.0)	80.1* (16.2)	79.8 (16.2)	66.0** (19.9)	63.9 (20.0)
	2	71.6 (15.7)	73.6 (15.4)	75.5 (14.0)	61.6 (16.5)	74.9 (13.9)	75.9** (14.0)	61.8* (16.5)	61.1 (16.4)
	3	60.1 (20.2)	62.4** (20.5)	64.2 (19.4)	49.7 (20.4)	63.5 (19.1)	64.8** (19.6)	49.9 (20.2)	49.5 (20.6)

Location of asterisk indicates group for which performance was found to be significantly higher. * $p < .005$; ** $p < .001$.
Skill 1 directly stated ideas; Skill 2 indirectly stated ideas; Skill 3 making connections with personal experience and text

Level of study accounted for 22.8 per cent in reading achievement. Because MANOVAs are based on an optimal linear combination of all the dependent variables, effect sizes (classical eta squared) were calculated for each of the outcome variables. Effect size ranges ($.08 < \eta^2 < .17$) indicated a slightly reduced effect size for this factor with any one particular outcome variable. Follow-up between-subjects effects revealed that Academic students' average performance was consistently and significantly higher ($p < .001$) than that found for Applied students' for the nine outcome variables (Table 1, Column 2). Results for interaction effects, while found to be significant ($p < .001$), were close to zero ($\eta^2 = .002$), and were not analyzed further.

To investigate within-group gender differences more fully, two sets of t-tests, one for each Level of Study, were conducted. The adjusted p -value was $0.05/9 = 0.005$ (Hummel & Sligo, 1971). Academic girls' mean performance was found to be significantly superior to their male counterparts for seven of the outcome variables (Table 1, Column 3). No significant differences were found for Skill 2, indirectly stated ideas – Informative texts ($p = .302$). Boys' performance was significantly better than girls ($p < .005$) for Skill 1, directly stated ideas – Graphic texts. The pattern of relatively superior performance of Academic girls' means over Academic boys' was to some extent reversed in the analysis of the Applied stream (Table 1, Column 4). Applied males' mean performance was found to be significantly superior to their female counterparts for five of the outcome measures ($p < .005$) (all three reading skills assessed in Informative Texts, and Skill 1, directly stated, and 2, indirectly stated, assessed in Graphic Texts). The mean of Applied females was found to be significantly superior to their male counterparts for only one outcome measure (Skill 2, indirectly stated, in Narrative Texts). No significant gender differences in mean scores were observed for three of the outcome variables (Skill 3, connections, in Graphic and Narrative Texts; Skill 1, directly stated, in Narrative Texts; $p = .128, .539, .712$ respectively).

To address the magnitude of effects between and within groups Cohen's d (1988) was used. For between group effects, d was calculated by subtracting the female mean score from the male mean score and dividing this difference by the average standard deviation of males and females. Following the convention of Hyde (2005) negative values of d means that males scored lower than females on a dimension, and positive values of d indicate that males scored higher than females. This analysis was carried out separately for each Level of Study; specifically, Academic girls were compared to Academic boys, Applied girls were compared to Applied boys. For within group effects, d was calculated by subtracting the Academic mean score from the Applied mean score and dividing this difference by the average standard deviation of Academic and Applied students for each of the outcome variables. This analysis was carried out separately for each gender; specifically, Academic girls were compared to Applied girls, and Academic boys were compared to

Applied boys (Table 2, Columns 3 and 4). Hyde's (2005) category system for interpreting effect sizes is used: close-to-zero $d = 0.10$; small $0.11 < d < 0.35$, moderate $0.36 < d < 0.65$, large $d = 0.66$ to 1.00 , or very large $d > 1.00$.

Table 2: Within and Between Group Effect Sizes

Text Type	Skill Type	Academic Fe compared to Academic M (1) ^{***}	Applied Fe compared to Applied M (2) ^{***}	Academic Fe compared to Applied Fe (3) ^{****}	Academic M compared to Applied M (4) ^{****}
Informative	1	-0.03 ^{**}	0.06 ^{**}	1.02 ^{**}	0.92 ^{**}
	2	-0.01	0.13 ^{**}	1.14 ^{**}	1.00 ^{**}
	3	-0.07 ^{**}	0.04 ^{**}	0.92 ^{**}	0.80 ^{**}
Narrative	1	-0.03 ^{**}	-0.01	0.76 ^{**}	0.73 ^{**}
	2	-0.10 ^{**}	-0.06 ^{**}	1.04 ^{**}	0.97 ^{**}
	3	-0.08 ^{**}	-0.01	0.97 ^{**}	0.89 ^{**}
Graphic	1	0.02 [*]	0.11 ^{**}	0.87 ^{**}	0.78 ^{**}
	2	-0.07 ^{**}	0.04 ^{**}	0.97 ^{**}	0.86 ^{**}
	3	-0.07 ^{**}	0.02	0.76 ^{**}	0.69 ^{**}

* $p < .05$. ** $p < .001$.

*** Negative values of d indicate that males scored lower on a dimension, and positive values of d indicate that males scored higher.

**** Positive values indicate that Academic students scored higher
Skill 1 directly stated ideas; Skill 2 indirectly stated ideas; Skill 3 making connections with personal experience and text

Although Academic females' performance relative to that of their male peers' was found to be significantly superior for seven of the nine outcome variables, the magnitude of these differences was all found to be in the close-to-zero range ($0.02 < d < 0.10$). In the Applied group,

findings of significant differences in favour of boys translated into an effect size in the small range ($0.11 < d < 0.13$) for two outcome variables, while the remaining three variables, as was the one variable (Skill 2 Narrative Texts) where Applied girls' performance was superior relative to boys were found to be in the close-to-zero range ($0.01 < d < 0.06$). The magnitude of these between-group findings can be contrasted with those found for within-group differences that follow.

Two separate within-group comparisons were carried out – one for each Level of Study. In the first comparison for all reading achievement outcome variables, Applied girls were compared to Academic girls, and in the second comparison Applied boys were compared to Academic boys. Whereas effect sizes comparing Academic girls to Academic boys, as well as Applied girls to Applied boys, were found to be in the close-to-zero and small range ($0.01 < d < 0.13$) effect sizes comparing Academic girls to Applied girls, as well as Academic boys to Applied boys were found to be in the large ($0.73 < d < 1.0$) and very large range ($1.0 < d < 1.14$).

With regard to text type, overall the largest within-group differences were found for Informative Texts. With regard to skill type, the largest within-group differences were found for Skill 2 (indirectly stated) regardless of the text type in which this skill was assessed. For the remaining two Skills (directly stated and connections), no consistent pattern across Text type was found.

DISCUSSION

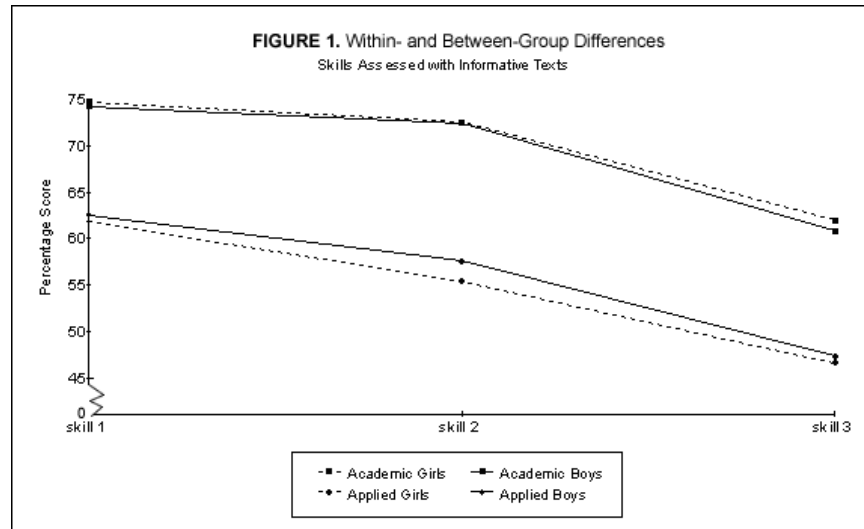
The results of this study, which sought to investigate the extent to which girls are better readers than boys, provided the following conclusions. Gender failed to account for even 1 per cent of the variance in reading achievement. Using data derived from international and national large-scale assessments of reading with a similar age group, this finding is consistent with those found by other researchers (Chiu & McBride-Chang, 2006; Hoglebe, Nist & Newman, 1985; Thorndike, 1973). This finding puts into question the generalizability of either biological and sociological explanations advanced to explain gender differences, and the effectiveness of proposed gender-specific strategies that follow from these explanations.

To emphasize the weak relation of gender to the reading sub-scores, the results of the present study found that statistically significant gender differences in mean sub-scores were not found in one third of the cases. When positive statistical differences were found in favour of girls, the power afforded when using such large data bases was able to detect positive statistical differences for mean differences as small as 0.8 per cent of the sub-score. When students were grouped according to their level of study, statistically significant relations favouring boys were found in one third of the cases (6/18), and although in the Academic stream, performance significantly favoured girls on most outcome measures (7/9), this pattern was to a large extent reversed in the analysis with the Applied group (5/9).⁶

The findings of this study indicate that, regardless of gender, performance for Narrative texts was highest, followed by Graphic Texts, with the weakest performance on Informative texts. Consistent with results found by other researchers (e.g. Lie et al., 2003; Scheuneman & Gerritz, 1990), comparisons of mean differences revealed that girls consistently outperformed boys in Narrative texts. However, the close-to-zero and small effect sizes ($0.01 < d < 0.10$) suggest that these differences may not have practical consequences. Any presumed advantage that boys might have for Graphic texts produced mixed results, depending on level of study. Such inconsistency in the direction of this advantage is similar to that found by both Rosen (2001) and Wagemaker (1996).

When considering the reading skills assessed within each Text type, findings of invariance in the present study favouring either girls or boys were not consistently found (See Figure 1). When gender comparisons across the nine outcome variables were made across both the academic and applied streams, findings of significant differences in variance favouring girls were found on only two outcome measures (Skill 1, directly stated, Graphic Text and Skill 2, indirectly stated, Narrative Text). In both of these cases, the magnitude of the effect was found to be in the close-to-zero and small range ($.02 < d < .11$). In the 14 out of 18 cases (2 groups \times 9 outcome measures) where statistically gender differences were found, regardless of whom they favoured, the magnitude of those differences was again found to be in the close-to-zero

range ($0.02 < d < 0.10$) and small range ($0.11 < d < 0.13$). The magnitude of the effect sizes for either the total overall reading score ($d = 0.15$), or for any of the sub-scores whether they favoured either girls or boys suggest that neither Skill type or Text type is of substantive consequence. Both the inconsistency in terms of which gender was favoured on a particular reading achievement measure, and the effect size findings are similar to the findings of studies with similar age samples (e.g. Hedges & Nowell, 1995; Hyde, 2005). These findings suggest that the current concern regarding the under-achievement of boys in reading achievement appears to have been overstated. There appears to be little evidence that the observed gender differences in reading achievement have practical consequence.



Those advancing brain theory (e.g., Gurian & Stevens, 2004) to support differentiated gender strategies may be, therefore, telling only part of the story. Researchers utilizing brain imaging techniques (e.g., Jaeger et al., 1998; Shaywitz et al., 1995) have found that although some men and some women may activate different portions of the brain when carrying out some reading related language tasks (phonological and syntax), any of the observed differences exist in the absence of significant

behavioral differences. The close-to-zero and small effect sizes of gender differences found in this study appear to support this notion thereby bringing the relevancy of brain theory into question.

Additionally, the findings from this study provide some empirical evidence for addressing what appear to be internal inconsistencies with the explanations that have been advanced to explain the purported gender gap (i.e., why some boys are doing quite well; why within-group differences are greater than between-group differences). The comparison of within- to between-group differences carried out in this study provided the opportunity to quantify the magnitude of each of these differences. With regard to whether the reading achievement of girls can be characterized as more different or similar to boys, Cohen's (1988) *U* statistic quantifies the percentage of non-overlap of idealized distributions. In the case of a *d* equal to 0.20, the *U* equals 15 per cent; that is, approximately 85 per cent of the areas overlap (see also Figure 2). The finding of effect sizes for between-group gender differences in the close-to-zero and small range supports the notion advanced by Hyde's (2005) Gender Similarities Hypothesis. This hypothesis holds that boys and girls are similar on most, but not all, psychological variables. In terms of reading achievement for the 2002 OSSLT, the magnitude of the differences found in girls' and boys' reading achievement implies that that they are more alike than different across all outcome variables.

In contrast to the relatively small amount of variance in reading achievement accounted for by gender, level of study was found to account for 22.8 per cent of the variance in reading achievement. Effect sizes for within-group (level of study) differences for boys were found to be in the large range ($0.69 < d < 1.0$), for girls they were found to be in the large and very large range ($0.76 < d < 1.14$). A *d* value equal to 1 means that the two groups are separated by one standard deviation, translating into approximately 68 per cent of non-overlap in distributions. Figure 2 illustrates the degree to which boys are more different from one another than they are same; the within-group differences are almost ten times greater than those found for the between-group gender differences. The same case was found to apply to girls.

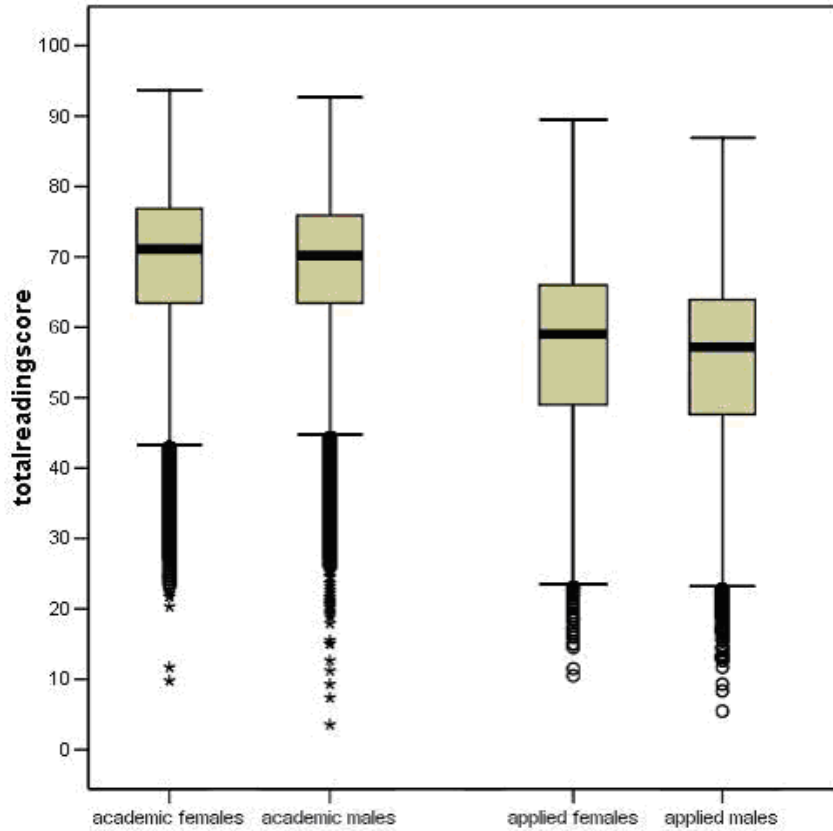


FIGURE 2. Boxplots for Total Overall Reading Score for Four Sub-Groups.

The boxplots above are useful for examining the spread and overall range in the total OSSLT reading scores for each of the four groups. The box itself contains the middle inter-quartile range. The line in the box indicates the median value for each group. The ends of the whiskers indicate the minimum and maximum data values, unless outliers and extreme values are present in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range for outliers and 3 times the inter-quartile range for extreme cases (indicated by asterisks). The boxplots indicate that the overlap in the distribution of scores for within-group (level of study) gender comparisons are much greater than for between-group comparisons. As well, they indicate that not all boys are performing poorly, or that all girls are performing well in their reading achievement.

At the heart of the concern that boys' needs are not being met is the assumption that all boys are not doing well in reading achievement. Practice-oriented, gender specific strategies that are intended to be implemented wholesale for all boys do not appear to address this issue satisfactorily. Figure 2 indicates quite clearly that not all boys are doing

poorly. Fifty per cent of Academic boys achieved a total reading score greater than 73.5 per cent, and approximately fifty per cent of Applied boys attained a score of 60 per cent or greater (the OSSLT passing score). On the other hand, some boys are not doing as well. Approximately 11 per cent of Academic boys and 49 per cent of Applied boys failed the OSSLT. Moreover, strategies intended to address the under-achievement of boys fail to consider that there concurrently exists a group of girls who are also at risk for poor reading achievement. Treating all boys, or all girls, as the same masks the high level of risk that certain boys and certain girls have for failing the OSSLT. Some boys are at risk for failing the OSSLT, but virtually the same percentage of girls are also at risk (Academic girls' 10.3 per cent; Applied girls' 53.4 per cent).

Reasons for Over-Stating Gender Differences

It is difficult to trace the reasons why the under-achievement of boys, in at least reading achievement outcomes may have been overstated, and to some extent misrepresented. There is some suggestion that it is not so much that boys are doing worse in reading achievement, but rather that that girls have improved their performance faster, leading to the belief that boys are falling behind. In this sense, the concern regarding boys' under-achievement may be partly a matter of perspective. The notion that women may be surpassing men, in some areas, may be difficult for those who adhere to traditional stereotypical norms. As a result, any news that girls may be surpassing boys may be used to support, and promote particular stereotypical educational or ideological beliefs (Mead, 2006).

An additional source feeding the boy crisis might be found in the limited types of statistical analysis that are included in large-scale assessment reports. Media coverage of the findings of these reports is then limited to reporting differences using overall averages, based on findings of positive statistical significance. However, findings of positive statistical significance do not automatically equate to substantive or practical effect. It is doubtful whether a large portion of the public understands this distinction. Without adequate clarification, or evidence to the contrary, it is perhaps not surprising that the under-achievement of boys may be largely overstated in some areas. In short, the crucial

question of how large the differences should be for them to be important for decision making appears not to have figured prominently in the current gender debate.

Limitations of the Current Study

This investigation was a descriptive study that sought to investigate the direction and the magnitude of gender differences for the reading component of the 2002 OSSLT. The results indicated the presence of weak gender differences on the OSSLT measures on tasks that presumably tapped a range of tasks related to text processing assessed within the context of three different text domains. It did not address questions about the component skills and cognitive skills that have been found to be important in reading comprehension research (e.g., word reading and its predictors, vocabulary, strategies, listening comprehension), or the manner in which background knowledge of the Ontario curriculum objectives included in the OSST may have affected performance. Secondly, Level of Study was used as a means to contrast between- to within-group differences, and although the effect sizes for level of study were found to be in the large range, all variables included in this study were correlational. The students in the present study were in their second year of high school, and reasons for course level selection were not addressed. Future studies may wish to identify factors that may be associated with stream selection, and whether these factors might be differentiated based on gender. Thirdly, the Academic and Applied level groups established in this study were based on structural classroom distinctions within Ontario's educational jurisdiction; therefore, the findings of large effect sizes for reading achievement outcomes may not generalize to other educational jurisdictions. Finally, the present study limited itself to only those students who did not identify themselves as receiving additional or alternative programming support, and therefore did not address whether similar findings of small effect sizes for gender will generalize to other sub-groups (i.e., students who are learning disabled, gifted).

CONCLUSION

Findings from a number of recent reports from large-scale assessments of reading achievement have reported that on average girls surpass boys in their reading abilities. Using data from one educational jurisdiction's large-scale assessment of reading achievement (OSSLT), the present study investigated the extent to which girls might be better readers than boys. The data included in this study represented 77 per cent of all first time eligible students taking the 2002 OSSLT. The small effect size associated with gender (less than 1%) suggests that there is not a homogenous group of successful reading behaviors or processes that is clearly perpetuated in either sex across any of the Text types or Skills used in this assessment. As a result, there appears to be little support to confirm either biological or socio-cultural explanations of gender differences in reading achievement, or for the gender specific strategies that have been recommended to remediate the purported gender gap.

Using level of study as a structural programming distinction, the tests of statistical significance regarding gender differences found opposite findings in relation to the direction of gender differences. The incongruity between the statistical conclusions of the between-group comparison suggests that although the question of which sex demonstrates greater reading achievement at the high school level may still be left largely unresolved, the close-to-zero and small effect sizes suggest that any of the observed differences may be of little practical consequences. In sum, the findings of this study strongly suggest that the notion of under-achievement of boys in the area of reading achievement has been greatly overstated. Further studies supporting this finding may be required. If, it is found that large gender gaps in reading achievement do not appear to exist at various age levels, or using different measures of reading achievement, it seems appropriate for reading research to continue to focus on better understanding the skills, processes and knowledge underlying reading comprehension that can be found in, and taught to, either gender. For those interested in improving the reading comprehension abilities of students, it is hoped that further evidence supporting this study's finding may provide sufficient impetus for moving beyond the gender debate.

ACKNOWLEDGEMENTS

Special thanks are extended to my advisor, John Kirby, and advisory committee member, Elizabeth Lee as well as to Anne McKeough and Ron Holden for their support and encouragement.

NOTES

¹ Although recognizing that format may also differentially impact performance, insufficient information was contained in the data set to allow this type of investigation.

² Graphic Texts were often embedded in Informative Texts.

³ The total score was calculated out of 200. Weightings for Skills 1 to 3 and Texts 1 to 3 were 30, 45, 25 and 45, 27, 28 percent respectively.

⁴ Presently, if a student is unsuccessful, he or she cannot graduate until he or she passes the test or completes the newly created Ontario Secondary School Literacy Course (OSSLC).

⁵ Although many of the cell sizes were found to be unequal, following Tabachnick and Fidell (1996) forced equalizing sample size was not pursued.

⁶ Of interest is that in each stream, the numerically superior gender performed better; by chance, we might have expected them to perform worse.

REFERENCES

- Alloway, N., Freebody, P., Gilbert, P., & Musprett, P. (2002). *Boys, literacy and schooling: Expanding the repertoires of practice*. Canberra, ACT: Commonwealth Department of Education, Science & Training. Retrieved July 29, 2007 from http://www.dest.gov.au/sectors/school_education/publications_resources/profiles/boys_literacy_schooling.htm
- Alloway, N., & Gilbert, P. (1997). Boys and literacy: Lessons from Australia. *Gender and Education*, 9(1), 49-58.
- American Psychological Association. (APA). (2001). *Publication manual of the American psychological association* (5th ed.). Washington, DC: Author.
- Anderson, K. J., & Accomando, C. (2002). 'Real' boys? Manufacturing masculinity and erasing privilege in popular books on raising boys. *Feminism & Psychology*, 12(4), 491-516.
- Ayres, L. P. (1909). *Laggards in our schools*. New York: Russell Sage Foundation.

- CBC News. (2005). Scores up, but tests slammed. October 20, 2005. Retrieved December 15, 2006, from <http://www.cbc.ca/canada/toronto/story/2005/10/20/to-eqao-20051020.html>
- Chiu, M. M., & McBride-Chang, C. (2006). Gender, context, and reading: A comparison of students in 43 countries. *Scientific Studies of Reading*, 10(4), 331-362.
- Cohen, J. (1995). The earth is round ($p < .05$): Rejoinder. *American Psychologist*, 50(12), 1103-4.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, M. (1998). 'A habit of healthy idleness': Boys' underachievement in historical perspective. In D. Epstein, J. Elwood, V. Hey & J. Maw (Eds.), *Failing boys? Issues in gender and achievement* (pp. 13-34). Buckingham, UK: Open University Press.
- Council of Ministers of Education, Canada [CMEC]. (1999). School achievement indicators program (SAIP), 1998, Report on Reading and Writing. Toronto: Author. Retrieved July 29, 2007, from <http://www.cmec.ca/pcap/rw98le/pages/ResultsE.stm>
- Education and Quality and Accountability Office [EQAO]. (2003). *Ontario secondary school literacy test, October 2002: Report of provincial results English language schools*. Retrieved July 16, 2007, from http://www.eqao.com/pdf_e/03/03P006e.pdf
- Education and Quality and Accountability Office [EQAO]. (2006a). *Large-Scale Assessment: Supporting the Everyday Work of Schools*. Retrieved July 29, 2007, from <http://www.eqao.com/AboutEQAO/GeneralQuestions.aspx?Lang=E>
- Education and Quality and Accountability Office [EQAO]. (2006b). *With learning in mind*. Retrieved January 20, 2006, from <http://www.eqao.com/NR/ReleaseViewer.aspx?Lang=E&release=b06R005>
- Elley, W. (1992). *How in the world do students read?* The Hague: International Association for the Evaluation of Educational Achievement.
- Epstein, D., Elwood, J., Hey, V., & Maw, J. (Eds.). (1998). *Failing boys? Issues in gender and achievement*. Buckingham, UK: Open University Press.

- Frank, B. (2005). Addressing boys' achievement in school. *Voice*, 7(2), 18-20.
- Gilbert, R., & Gilbert, P. (1998). *Masculinity goes to school*. London UK: Routledge.
- Gurian, M., & Stevens, K. (2004). With boys and girls in mind. *Educational Leadership*, 62(3), 21-26.
- Gurian, M., Henley, P., & Trueman, T. (2001). *Boys and girls learn differently*. Hoboken, N.J.: Jossey-Bass.
- Halpern, D. K. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41-45.
- Hogrebe, M. C., Nist, S. L., & Newman, I. (1985). Are there gender differences in reading achievement? An investigation using the high school and beyond data. *Journal of Educational Psychology*, 77(6), 716-724.
- Hummell, T. J., & Sligo, J. R. (1971). Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 76, 49-57.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, September, 581-592.
- Jaeger, J. J., Lockwood, A. H., Van Valin, R. D., Kemmerer, D. L., Murphy, B. W., & Wack, D. S. et al. (1998). Sex differences in brain regions activated by grammatical and reading tasks. *Neuroreport*, 9, 2803-7.
- Kehler, M., & Greig, C. (2005). Reading masculinities: Exploring the socially literate practices of high school young men. *International Journal of Inclusive Education*, 9(4), 351-370.
- Lie, S., Linnakylä, P., & Roe, A. (Eds.). (2003). *Northern lights on PISA unity and diversity in the Nordic countries in PISA 2000*. Programme for International Student Assessment. Department of Teacher Education and School Development. University of Oslo, Norway.
- Luke, A., Freebody, P., & Land, R. (2000). *Literate futures: Review of literacy education*. Brisbane, QLD: Education Queensland.

- Maccoby, E. (1990). Gender and relationships: A developmental account. *American Psychologist*, April, 513-20.
- Maccoby, E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.
- Martino, W., & Berrill, D. (2003). Boys, schooling & masculinities: Interrogating the 'right' way to educate boys. *Educational Review*, 55(2), 99-117.
- Mead, S. (2006). *The truth about boys and girls: The evidence suggests otherwise*. Retrieved February 21, 2007, from http://www.educationsector.org/analysis/analysis_show.htm?doc_id=378705
- Millard, E. (1997). *Differently literate: Boys, girls and the schooling of literacy*. London. UK: Falmer.
- Miller, L. (2003, December). Ensuring literacy in an age of scrutiny. *Professionally Speaking*. Retrieved July 29, 2007 from http://www.oct.ca/publications/professionally_speaking/december_2003/literacy_at_risk.asp
- Mills, M. (2003). Shaping the boys' agenda: The backlash blockbusters. *International Journal of Inclusive Education*, 7(1), 57-73.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., and Kennedy, A. M. (eds), *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary School in Thirty-five Countries*, Boston MA: International Study Center, Boston College.
- National Assessment of Educational Progress [NAEP]. (2005). *NAEP 2004 Trends in Academic Progress: Three decades of student performance in reading and mathematics*. NCES National Center for Education Statistics. U.S. Department of Education. Authors: Marianne Perle, Rebecca Morgan, Anthony Lutkus.
- Noble, C., & Bradford, W. (2000). *Getting it right for boys and girls*. London: Routledge.
- Organisation for Economic Co-Operation and Development [OECD]. (1999). *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: OECD Publications.
- Organisation for Economic Co-Operation and Development [OECD] PISA (2001). *Knowledge and skills. First results from PISA*, Paris: OECD Publications.

- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CN: Yale University Press.
- Ontario Ministry of Education. (2005). *Me read? No, way! A practical guide to improving boys' literacy skills*. Retrieved December 1, 2006, from <http://www.edu.gov.on.ca/eng/document/brochure/merread/index.html>
- Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, 64(6), 916-924.
- Roe, A., & Taube, K. (2003). Reading achievement and gender differences. In S. Lie, P. Linnakyla, & A., Roe (Eds.), *Northern lights on Pisa: Unity and diversity in the Nordic countries in PISA 2000* (pp. 21-36). University of Oslo, Norway.
- Rosen, M. (2001). Gender differences in reading performance on documents across countries. *Reading and Writing: An Interdisciplinary Journal*, 14(1), 1-38
- Rowe, K. J. (2000, August 21-22). Problems in the education of boys and exploring real effects from evidence-based research: Useful findings in teaching and learning for boys and girls. Background paper of keynote address presented at the Teaching Boys Developing Fine Men Conference, Brisbane. ERIC document.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Contable, R. T., Skudlarski, P., Fulbright, R. K., et al. (1995). Sex differences in the functional organization of the brain for language. *Nature*, 373, 607-9.
- Scheuneman, J., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), 109-131.
- Smith, M. W., & Wilhem, J. D. (2002). *"Reading don't fix no Chevy": Literacy in the lives of young men*. Portsmouth, NH: Heineman.
- Sommers, C. H. (2000). *The war against boys: How misguided feminism is harming our young men*. New York: Simon & Schuster.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.) New York: Harper Collins.
- Teese, R, Davies, M., Charlton, M., & Polesel, J. (1995). *Who wins at school? Boys and girls in Australian secondary education*. Sydney: Macmillan

- Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries: International studies in education III*. New York: Halsted Press.
- Topping, K., Valtin, R., Roller, C., Brozo, W., & Dionisio, J. L. (2003). *Policy and practice implications of the program for international student assessment (PISA) 2000*. Report of the International Reading Association PISA Task Force. International Reading Association, Newark, DE.
- Wagemaker, H. (Ed.). (1996). *Are girls better readers?* Amsterdam: International Association for the Evaluation of Educational Achievement.

Bozena White is a Ph.D. candidate at the Faculty of Education, Queen's University, Kingston, Ontario (whiteb@educ.queensu.ca). Her research interests include the component skills and processes underlying effective reading and reading comprehension, and their relation to different reading comprehension assessment measures.