University of Alberta

Persistent and Contradictory Comparative Claims of Boys' and Girls'

Reading Achievement: A Historical Interpretive Approach

by

Karen Loerke

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Elementary Education

©Karen Loerke

Fall 2012
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

ABSTRACT

Persistent and contradictory claims that boys are not doing as well as girls in reading achievement have been made since the time of compulsory education in North America (1890). Since approximately half of the school population is comprised of boys, it was critical to understand the extent of, and possible causes for such a gender gap. A plethora of inconsistent research findings across a range of methodologies and perspectives over such an extensive time period (1890 to the present) made it essential to implement a systematic evidence-based historical interpretive (SEHI) methodology to investigate the relative claim of a boy-girl gap in reading achievement over 12 decades. Database searches generated over 3,000 hits and yielded 78 trustworthy studies from four time-periods (1890–1920; 1921–1945; 1946–1980; 1981–2011). A comprehensive examination and interpretation of the evidence-based studies revealed that boys' and girls' reading achievement differ by approximately only 1%. Evidence has indicated confusion around the reporting of statistical information and the use of new scaling systems that have inflated differences in the scores between girls and boys. Findings reveal the critical need to study primary sources when citing prior results. The only select groups of boys underachieving in reading comprehension borne out by the research evidence are those from low SES backgrounds and boys of colour. Reasons that boys underachieve in reading and possible solutions are multi-dimensional that go beyond proposed simplistic solutions such as buy more boy-friendly books, hire more male teachers, or provide boys' only classes. To focus

on boys as a group presupposes that all boys are underachieving in reading and that all girls are doing well, thereby overstating the problem for boys while ignoring girls. Evidence from low-gender gap schools reveals that when teachers have high expectations for their students with additional support for struggling learners, all students achieve. Publicity claims around boys' poor performance in reading comprehension are not only false but have serious implications for both boys and girls. It is time to debunk the myth: boys' are not underachieving in reading and have not done so over the past century.

Table of Contents

LIST OF TABLES

Chapter One:

Introduction

Claims that boys are not doing as well as girls in reading are widespread. Claims of boys' underachievement in reading can be heard on talk shows, reported in the news, written about in the professional literature, commented on in newspapers, discussed in parent magazines, researched in the academic literature, and presented on at conferences. Claims about boys not doing as well as girls in reading are also persistent and contradictory. Going as far back as the early 19th century (Ayres, 1909; Commins, 1928; Lincoln, 1927; Stroud & Lindquist, 1942; Thorndike, 1903), reports have similarly arrived at divergent conclusions. Some studies claim that girls outperform boys in reading achievment (Canadian Council on Learning, 2007), while others challenge the claim, and conclude that the evidence has been greatly overstated (White, 2007). The pupose of this study was to analyse methodically all the evidence-based research studies that have been reported on sex differences in reading achievement in order to bring clarity to the peristent, contradictory, comparative claims regarding boys' and girls' differential reading achievement.

A sample of the claims of a boy-girl gap in reading achievement for over a century are presented next to provide a sense of the controversy in, and discrepancy within the research literature. For example, the Canadian Council on Learning (CCL) (2007) published a report on lifelong learning in Canada, and reported a "gender gap" in reading in favour of girls based on the results of the 2003 Programme for International Student Assessment (PISA) (OECD, 2004b).

The authors assessed the reading skills of about 22,000 15-year-olds from 1,000 schools across Canada and concluded that "Canadian girls outperformed boys in reading, . . . The difference in the reading levels was more significant, with girls scoring an average of 32 points higher than the boys" (Canadian Council on Learning, 2007, p. 28). Please note, the CCL report did not indicate in the previous sentence how the results were "more significant". Morever, in the same year the CCL report was released, analysis of the 2002 Ontario Secondary School Literacy Test (OSSLT) with 113,050 secondary students, showed that gender accounted for less than 1% of the variance in reading achievement (White, 2007). Historically, international studies have reported similar and consistent findings to those reported by White and claimed minimal differences between the reading achievement of school-aged boys and girls (Elley, 1992; Lincoln, 1927; Stroud & Linquist, 1942; Thorndike, 1973).

Most studies on sex differences in achievement prior to the 1980s focused on mathematical and mental abilities (Hogrebe, Nist, & Newman, 1985), and literacy sub-tests around word activities (Thorndike, 1903). Initially, sex differences in school achievement were determined from teacher reports of the numbers of promotions, retentions, and school dropout rates. For example, Ayres' study (1909) was based on records of thousands of students from fifteen cities in the United States. Analysis of the records indicated that 13% more boys than girls repeated grades due to poor achievement levels, and 17% more girls than boys completed high school, and more boys dropped out after the compulsory age of 14. The early comparative studies on gender differences and

achievement results indicated that girls were superior in spelling and handwriting, but that boys and girls were equal in arithmetic (Thorndike, 1903).

Some of the early studies in reading comprehension found no significant sex differences in reading at the elementary and high school levels (Commins, 1928; Jordan, 1937; Traxler, 1935). Contrary results from data collected between 1932 and1940 noted that generally boys underachieved in elementary school but surpassed girls in high school. Specifically, girls were reported to be superior to boys in all subjects in the elementary grades, including reading comprehension. However, during the same period, the reverse occurred in high school where results indicated boys were generally superior to girls in achievement, except for algebra and reading comprehension where evidence of differences was small and not significant (Stroud & Lindquist, 1942). Examination of some of the research on boys' and girls' reading achievement showed two categories (Lietz, 2006a). Some studies claim girls outperform boys on reading measures (Gambell & Hunter, 1999; Gates, 1961; OECD 2001, 2004b, 2007, 2010). Other studies report no differences between boys' and girls' reading achievement (Hogrebe et al., 1985; Thorndike, 1973; White, 2007).

As you see, there was tremendous uncertainty around the reported findings published over the past century on boys' and girls' reading achievement. Like Newton's Third Law of Motion, "for every action, there is an equal and opposite reaction," I sensed that for every study that supported a difference in achievement in favour of boys' reading achievement, there is an equal and opposite study that claimed girls are better. To date, the nature and scope of evidence-based research

findings on the boy-girl reading achievement gap that extends over the time period from 1890 to the present had never been analyzed, synthesized and interpreted.

Motivation

In my attempt to make meaning out of the conflicting findings and to reconcile differences, I was motivated to unravel some accountings for the persistent, widespread and inconsistent findings over the past 12 decades of research. Contradictory claims of boys' and girls' differential reading performance leave classroom teachers, parents, and others with conflicting messages on how best to deal with the claimed underachievement of boys. It is important that we understand the specific nature of the varied claims because generally speaking, classrooms have equal numbers of boys and girls. If half the population of our schools is experiencing difficulty in reading, then it is necessary to have some coherent and comprehensive understanding of the underlying causes. If prodigious differences exist, researchers and classroom teachers will need to explore strategies for teaching reading that specifically target boys' underachievement. Explanations and strategies for classroom practice should be determined based on evidenced-based findings and not on advocacy papers or media hype. However, if significant differences do not exist, then other important questions must be considered, not the least of which is the ethics of appealing to the research evidence prior to making unsubstantiated, overstated, or misinformed claims about boys and their achievement in reading.

Research Question and Goals for the Study

My doctoral research was designed to bring clarity to the research findings over the extensive period from 1890 to the present regarding the comparative claims of boys' and girls' reading achievement. Using an historical interpretive approach, the main question I planned to answer was whether the evidence-based research findings from 1890 to the present, support a difference in boys' and girls' reading achievement that was large and warranted the establishment of such a claim? My specific goals were:

1. To establish the phenomenon of persistent and contradictory comparative claims regarding boys' and girls' reading achievement and determine the degree of nuance.
2. To analyze, synthesize and interpret findings to determine explanation(s) of the difference.
3. To develop a comprehensive interpretation of boys' and girls' reading achievement that provides recommendations for classroom practice and considerations for further research.

Significance of the Study

My study of the persistent and contradictory comparative claims of boys' and girls' reading achievement was important for the following reasons:

1. To date, a historical synthesis on boys' and girls' reading achievement had never been completed. Although there have been research reviews (Hogrebe et al., 1985) and a meta-analysis of large-scale studies from 1970 to 2002 (Lietz, 2006a; 2006b), my study is the first known

interpretation of all the available evidence-based findings on the relative claims of boys' and girls' reading achievement.

2. An examination of the evidence on sex differences in reading achievement from a historical perspective provides a coherent framework for establishing or refuting the claim that boys are underachieving in reading.
3. Public debate surrounding the claims of underachievement of boys in reading has led to policy implications and changes in classroom practices (The Education Alliance, 2007). However, before such policies and changes in classroom practices are implemented, a consolidation of the evidence-based research findings was needed in order to resolve whether the claims are substantiated. If boys are indeed underachieving in reading, it is critical that we examine how best to address the reading needs of boys. Possible explanations that account for sex differences in reading may yield further insight for classroom practice as well as provide a beginning point for further research. If the claim is refuted, then teachers can put the issue to rest and focus on other reading issues and concerns for both boys and girls. Currently, incredible energy is devoted to new books, different approaches, and even segregated collections for boys in libraries despite the fact that evidence is contradictory. Since literacy concerns are a responsibility of all educators, parents, policymakers, and

researchers it is important to have a consistent understanding of sex differences in reading achievement based on evidence-based research.

Chapter Two:

Review of the Literature

Theoretical Framework

As outlined in the introduction, there are persistent and contradictory claims of boys' and girls' reading achievement. What then are some of the reasons offered for the widespread and inconsistent findings and results over the duration of a century or more? Using an interdisciplinary theoretical approach, I considered three proposed explanations: a) test design and interpretations; b) inherent biological factors; and c) sociological and cognitive factors. I left open the possibility that additional explanations, connections, and interconnections may emerge as findings were procured, examined, and interpreted throughout the duration of this study. Although described separately, the understanding is that these components are not mutually exclusive but rather interconnect and influence one another.

Test Design, Overgeneralizations, and Biases

Using results from large international achievement tests, such as the PISA to track student achievement over time is common in many countries today. These large-scale testing programs are frequently challenged. Some researchers have noted discrepancies in test design, data interpretation and testing assessment procedures in international achievement tests. For example, conclusions from a meta-analysis including 139 studies on reading achievement and sex differences indicated that in over half of the studies, differences were the result of test design

(Lietz, 2006a). By way of example, Lietz (2006a) noted that when comparing test results they

> differed markedly not only in design, sample size, scope and the scale of the reading score but also in the reporting of results. Thus, results were frequently not reported in terms of standardized effect sizes but in terms of correlation coefficients, regression coefficients from single-level and multi-level analyses, sums of squares, percentage differences or mean differences. (p. 329)

Test design.

To make fair comparisons, Lietz (2006a) used a meta-analysis approach, a systematic way to synthesize findings by converting results to a common unit of measurement. This method corrects for different sources of error, particularly sampling errors, and thus points to other study characteristics to explain variations. Although test design is one consideration for differences, Lietz (2006a) concluded that other factors such as item selection procedures, scaling of reading scores, and changes in reading in school and society account for additional gender differences.

Mead (2006) noted differences in data interpretation in her comparision of the 2003 and 2005 National Assesment of Education Progress (NAEP) in the United States. Mead, a senior policy analyst at Education Sector in the United States, an independent education think tank, made the claim, "hysteria about boys is partly a matter of perspective" (p. 3). Mead submitted that boys' reading achievement had actually increased, but not at the same rate as the girls (2006, p.

4). She further concluded that groups of boys, particularly Hispanics and Blacks from low-income homes, appeared to be falling behind. Similar results of low performing groups have been documented in Australia, where "the gap between boys and girls has widened over the past 20 years. Especially at risk are Aboriginal boys, boys from working class areas, and boys from homes where the first language is not English" (West, 1999, p. 41). The studies from Mead and West raise concerns about influences on the boy-girl gap that are related to the impact of relationships of power beyond gender and socio-cultural factors, such as race and poverty rather than sex differences.

Overgeneralizations.

According to White (2007) concern for the way the media reports statistical information has resulted in "overgeneralizations" of conclusions. An example of overstated conclusions are the findings from the Canadian Council on Learning (2007). Reading peformance on the 2003 PISA (OECD, 2004b) reading test was compared to reading performance on the PISA 2000 (OECD, 2001) results. The comparison claimed that girls surpassed boys by 32 points. However, no total possible score was reported. What does a difference of 32 points signify? Was the 32-point difference out of a total possible of 50, 100, 1,000? The answer to my question makes a significant difference to the interpretation. When incomplete findings are presented and out of context, without clear and precise language, and without clarification of the data, then opportunities for distortion are heightened. The intrinsic value of the reported research is questionable and the potential for negative fallout bewildering.

Test biases.

Other concerns have been raised over the inherent biases and power of testing procedures in international studies such as how they are administered and scored. For example, Purves (1992) noted in the International Association's Evaluation of Educational Achiemement (IEA) study of written composition that the raters in all the participating countries were female and therefore, results were biased. Yet, Gambell and Hunter (1999) suggested that the scoring bias is overstated and that "sex differences are omnipresent" (p. 11) therefore, differences cannot be explained in terms of anomalies in assessment, sampling, or scoring. The claim of sex differences in reading achievement in the plethora of inconsistent research findings promotes confusion, conflict and debate. Although discrepancies in test design, data interpretation and testing assessment procedures have been documented in the preceding examples, other factors have been noted as possible explanations for sex differences in reading achievement. Biological, cognitive, and sociological factors are some of these and are discussed next.

Inherent Biological Factors

In the early 1900s there was an emerging focus on individual differences coupled with a quest to understand the nature of human characteristics. As a result, scores of quantitative studies investigated individual differences, including the nature and degree of diversity (Anastasi, 1958). A host of studies in psychology were conducted that compared sex differences. The studies of sex differences were varied and included: growth and proportions of the body, senses, muscular strength, intellectual impulses, metabolism (blood, respiration,

excretion, hair, pigmentation), menstruation, dreams, insanity, criminality, emotions, artistic nature, birth and mortality rates, size of brain, rate of hand tapping, handwriting, preference for colour, attitudes, memory, choice of activities, and sexual impulses (Ellis, 1904; Woolley, 1910). Educational studies adopted the concept of measuring individual differences and began to compare differences in boys' and girls' achievement in school (Thorndike, 1903). Sex differences in student achievement were proposed in terms of biological (Pressey, 1918) and sociological factors (Ayres, 1909). These factors are examined next, starting with biological.

Historical Evidence

Havelock Ellis (1904), a medical doctor, spent his lifetime studying human sexuality and the differences between men and women. Based on anatomical data collected over 12 years, Ellis proposed the variability theory, submitting that men more than women diverged from the average on tests of intelligence. Expanding on Ellis' variability theory, Thorndike in 1914 included men's variability of intelligence in his research. He determined that men were about one-twentieth more variable with a greater range than women in mental abilities. His conclusions influenced Pressey and Hollingworth to study further the proposed theory of variability.

Pressey (1918) used a group scale of intelligence with 2,544 students between the ages of 8 to 16 years to examine the variability theory with school-aged children. Results of Pressey's study affirmed Thorndike's conclusion that

the distribution of total scores showed a remarkably greater variability in intelligence among boys than among girls.

Although many psychologists in the early 1900s agreed that men and women did not differ in average intelligence (Pressey, 1918; Thorndike, 1914), the variability theory of intelligence continued to be challenged, especially among female psychologists (Hollingworth, 1914; Thompson, 1903). Almost half a century later Anastasi (1958) reported

> It may be added that the doctrine of greater male variability was regarded by its proponents as a fundamental biological law and was believed to hold for all traits, both physical and psychological. This doctrine enjoyed wide popularity and was adopted by a number of psychologists during the first quarter of the present century. Nor is it completely absent from contemporary writings, especially popular discussions of sex differences. (p. 456)

It is interesting to note the shift in Anastasi's language describing greater male variability, with supporters asserting that the theory was a fundamental biological law. The use of the word, "doctrine" to describe the theory implies that although widespread, researchers agreed to the theory without additional evidence thereby promulgating what grew to be a false dogma. Sex differences in intelligence continued to be debated with the construction of new IQ tests. Items included in revised intelligence tests fuelled debates over how some items favoured either boys or girls (Anastasi, 1958). However, discussion of these debates was beyond

the scope of my study. The purpose of my study was to establish the reported sex differences in reading achievement.

Sex Differences in Achievement

Studies focusing on differences in boys' and girls' school achievement evolved from the comparative studies of anatomical and mental traits. Conclusions from early achievement studies, from the period 1900 to 1930s, indicated that girls outperformed boys in spelling and handwriting, and that boys performed better in mathematics, logical reasoning and gross motor skills (Lincoln, 1927; Pressey, 1918; Thorndike, 1903). Consequently, the perception that girls were superior in literacy skills was spawned. The assumption was that boys underachieved in English and that this condition was innate. Evidence of such thinking is recognized in the quote made by Sir Henry Newbolt on May 2, 1919, in the Report of the Departmental Committee to inquire into the position of English in the Educational System of England. He wrote,

> The English composition of many boys is clumsy and painful to the verge of illiteracy and they are unable to grasp a line of argument or assimilate or criticize the contents of a book. … To be bad at English seems to be regarded by both masters and boys as a natural. (Great Britain. Committee on English in the Educational System of England, 1921, p. 88)

The implication based on the committee report is that girls could "grasp a line of argument," and a view that girls surpass boys in language and reading is prevalent today. However, it is puzzling that boys are reported to be superior to girls in

mathematics and science (OECD 2001)—both subjects that require the ability to read.

Current Research in Brain Functioning

A growing number of popular psychology books make the claim that girls have better developed verbal skills as a result of different brain functioning (James, 2007). Many of these claims are not substantiated; they are based on correlational and not causational data. James, for example, makes the claim that while "cellular differences between male and female brains does not yet explain gender differences in brain function, it is interesting that there are differences even though the studies do not concur" (2007, p. 15). She further proposes that it is difficult to determine if boys and girls differ in the structure or function of the brain, and that the reason "boys and girls think differently is because society says they should think a certain way" (James, 2007, p. 22). Such conclusions imply a socio-cultural explanation for differences in boys and girls, and do not confirm biological brain differences initially proposed by James in the same work. However, researchers who study sex differences in brain functions are encouraged that medical advancements will substantiate their conclusions that boys and girls have different brain functioning. Maccoby (1998) claims that "Techniques for noninvasive study of brain function are being rapidly improved, and we can hope for more detailed information concerning the development of the language functions of the brain in the two sexes" (p. 107). In the meantime, unfounded claims, such as those proposed by James, are not be endorsed in this study. To

avoid personal biases and media hype, only evidence-based qualitative and quantitative studies were analyzed, synthesized and interpreted.

Claims that different brain functioning and delayed maturation affect boys' underachievement in reading are inconclusive. Although it has been confirmed that generally girls mature earlier than boys, it has not been determined that maturation rates affect reading, "the rate of intellectual growth is unrelated to the rate of physiological growth" (Maccoby, 1966, p. 38). Furthermore, Dwyer (1973) claimed that differences in school performance based on maturation usually could be explained by social and cultural factors. Maccoby (1998) continued to study sex differences and reconsidered her initial conclusions (1966), and stated that new evidence pointed to an advantage for girls in early language abilities. She concludes, however, that

> Naturalistic studies have been plagued by small sample sizes, and in laboratory work, boys have sometimes been more resistant to the testing procedures, so that their level of language competence has been more difficult to assess and may have been underestimated. (Maccoby, 1998, p. 106)

Researchers, who support the claim that boys' brains are biologically different from girls', look to new and improved medical advancements in brain research for support for their position.

Although biological explanations for differences in boys' and girls' achievement continue to be reported, since the early 20th century there appears to be a growing body of evidence that supports sociological factors (Ayres, 1909;

Dwyer, 1973; Gambell & Hunter, 1999; Preston, 1962). Boys and girls are socialized differently and how gender stereotypes develop has been proposed as possible explanations for differences in reading achievement. Sociological and cognitive factors are examined next.

Sociological and cognitive factors.

Over the years, major theories proposing biological, sociological and cognitive factors have attempted to explain sex differences and gender development. This section briefly examines some of the sociological factors that historically have been used to account for intellectual differences in men and women. Cognitive and social cognitive theories, and gender development were reviewed as well as some of the current research on group interactions and how sex sub-cultures develop. Lastly, how cognitive and social cognitive theories, and gender development relate to reading achievement is summarized.

Historical Studies

The variability theory developed by Ellis (1904) and extended by Thorndike (1914) proposed that men's intellectual abilities had a greater range than women, thereby providing a rationale for why more men ranged from low to high intellectual functioning. According to Thorndike (1914), "Two times as many men as women will be found in asylums for idiots and imbeciles; and one and a third times as many will be found by a census including those cases (commonly somewhat less stupid) cared for at home" (p. 189). Female psychologists, Woolley (1910) and Hollingworth (1914), shifted from an emphasis on a "biological to a sociological interpretation of the mental

characteristics of sex" (Woolley, 1910, p. 342). According to Woolley, women were prevented from developing their intellect because society, at the time, held a belief that if women pursued higher levels of training, women would marry later, and ultimately have fewer children. "The conclusion seems to be that it is the highest duty of women to refrain voluntarily from developing her own intellectual capacities for fear of injuring society" (Woolley, 1910, p. 342). She hypothesized that the fear of women pursuing intellectual training was that women would lose important reproductive years that would further affect society's ability to perpetuate itself. It seems to have been a time of cultural conditioning of women to stay disengaged from higher education, even though mothers' work is highly intellectual work.

The major studies that had been completed on the variability theory at the time were examined by Hollingworth (1914). She submitted that there was a lack of conclusive empirical evidence to support the claim that inherent biological differences accounted for variance between men's and women's intellectual achievement. She concluded the variability theory precluded women from achieving intellectual equivalence. As a feminist, Hollingworth strongly opposed Thorndike's variability theory and how it restricted women in society. Hollingworth and Woolley are two examples of female psychologists from the early 1900s who proposed that sociological factors accounted for sex differences.

Sex Differences vs. Gender Development

Differences in males' and females' aptitude, achievement, and personality continue to be studied in psychological research under the specific field of

"Individual Difference" (Maccoby, 1998). Additionally, some researchers began to examine gender socialization and how social interactions affect and shape what it means to be a man or a woman (Millard, 1997). The perspective that sex is biological, involving anatomy, hormones, and physiology and that gender is constructed through psychological, cultural and social interactions was taught in university classes in the 1960s by sociologists (e.g., West & Zimmerman, 1987). At the individual level, cognitive and social cognitive theories contributed to explanations of gender development. Some of the prevalent theories on how children learn gender identity are examined next, followed by how gender development is related to reading and achievement.

Gender Theories

Kohlberg's (1966) theory of children's sex-role concepts and attitudes is based not on biological or cultural factors but rather on children's active meaning-making of their world. He drew on Piaget's four stages of cognitive development characterized by changes in cognitive thinking that developed as children matured. Similarly, Kohlberg theorized that children's sex-role attitudes change in relation to children's cognitive development. He proposed that children's understanding of gender develops in three stages, identity, stability, and constancy (Kohlberg, 1966). According to Kohlberg, children understand at about 2 years of age that they are either male or female. From ages 2 to 7, children crystallize their gender identity. A basic sex-role stereotype develops from children's understanding of biological differences and from observations of the social rules and roles of males and females in their lives. Around ages 4 to 8 years, children

develop masculine-feminine values and tend to identify with and imitate like-sex models. Kohlberg's (1966) theory of sex-role identity follows a regular course of development that is largely dependent upon cognitive maturity. Since the introduction of Kohlberg's theory, alternate gender-schema and social cognitive approaches to gender development have emerged.

In a nutshell, gender-schema theory provided an account of sex typing. It proposed that children learn what it means to be a boy or girl from the culture in which they live, and they adjust to the norms and expectations that are embedded in the practices of a culture (Bem, 1981). Gender-schema theory is based on how children formulate a theory of acceptable gender behaviour for themselves and others. Children evaluate information in terms of whether the new information fits with their evolving gender-schema. Sex typing occurs when children are motivated to learn which attributes match their own sex. The development of their self-concept is strengthened when children choose from among the many possible activities that fit their sex typing. Gender-schemata became a guide that regulates behaviour where children learn to conform to a culture's definitions of maleness and femaleness (Bem, 1981). Boys and girls may differ from each other in what they consider to be a male or female activity; the difference, however, depends on the individual child's gender-schema (Katz & Sokal, 2003).

Human development theories incorporate social cognitive perspectives proposed by Kohlberg (1966) and gender-schema theorists (Bussey & Bandura, 1999). The social cognitive theory of gender development and differentiation incorporates the social aspects that affect people's concept of gender development

as well as the cognitive aspects from Kohlberg's theory of sex-role identity and from gender-schema theory. The interdependence of human development, societal influences, and personal choices affect gender differentiation. Gender development is explained in a "triadic reciprocal causation" (Bussey & Bandura, 1999). The triadic includes personal cognitive influences, biological, and environmental events. The interactions between and among the three are not equal and may fluctuate (Bussey & Bandura, 1999). They claim people are not strictly influenced by social forces that act on them but rather people personally contribute to their self-development as they interact within their social environment (Bussey & Bandura, 1999).

The major themes of cognitive theories and gender-schema theory are inherent in the work of Martin and Ruble (2009). They proposed another perspective that is characterized by three features: a) the emergence of gender identity and its consequences, b) an active, self-initiated view of gender development, and c) development patterns. They suggest that children actively construct what it is to be male or female and develop personal standards for appropriate gender behaviour. Martin and Ruble summarized their central tenets; "Children are gender detectives who search for cues about gender - who should or should not engage in a particular activity, who can play with whom, and why girls and boys are different" (2009, p. 165). When children understand that they are either a boy or a girl, they are motivated to actively construct and learn what it means to be that sex in the social world. Children's understanding of gender changes evolves as their cognitive abilities develop (Martin & Ruble, 2009).

Children's gender-related beliefs and behaviours become more flexible as they mature.

Sex differences and gender development have been studied, for the most part, at the individual level. But, some researchers have studied gender development from the perspective of group dynamics. Maccoby's (1998) latest book, *The Two Sexes: Growing up Apart, Coming Together*, diverges substantially from her previous work (1966). Maccoby made a shift in interest from individual differences to the ways people interact with others in group settings and at different times in their growth and development. Maccoby's (2002) work focuses on children's tendency to segregate themselves into same-sex groups and how group interactions affect their gender development. From ages 4 to 12 years, children for the most part spend their time with their own sex in large groups. Researchers are now interested in how sex distinctive subcultures are formed and how these groups change with age. Maccoby (2002) proposed that understanding individual differences as well as how individuals integrate within group processes are "promising avenues for the next phases of research on gender development" (p. 57). Recently, it has been noted that multiple perspectives are needed in order to include children and adults who have a mismatch between their biological sex and gender (Martin & Ruble, 2009).

Gender Theory and Reading Achievement

It has been reported that a gender gap in reading achievement is the result of boys' and girls' construction of different perceptions and attitudes toward school and reading (Dwyer, 1973; Millard, 1997). Millard (1997) concluded that

parents act as models of literacies for their children. Data from questionnaires and interviews where children were asked who provides support for learning and who reads the most in families show that mothers (> 25%) more than dads (< 15%) were key influences on learning and reading in the home. In addition, mothers were reported to be more likely to share or recommend new books to their children than fathers (Millard, 1997, pp. 82–87). It is important to note that Millard did not report whether the percentages reported were statistically significant. Gambell and Hunter (1999) proposed that reading is viewed as a female activity, a view that is engendered before children go to school. Interestingly Millard (1997) calculated the difference in the amount of time dads and moms read at home and reported a 10% difference, which when considered in realistic terms is not large. By way of an example, the difference is a mere three minutes if a mother was to read for 30 minutes a day, the father would read for 27 minutes.

Comparisons between pastime leisure activities are reported as evidence of differences in boys' and girls' attitudes toward reading. Gambell and Hunter (1999) noted in their study in Saskatchewan that boys spent more time playing sports ($z = 10.19$ higher), watching more television ($z = 7.06$ higher), watching sports' programs ($z = 19.53$ higher) and reading less ($z = 11.04$ lower) than their female counterparts (p. 5). The differences reported were statistically significant at $p = < .001$. Dwyer (1973) proposed that boys' perception of school and reading is in conflict with the male sex role: "In the early grades, boys are subject to pressures to develop the male sex role to participate in active, non-academic

activities, and to express autonomy from mothers and, by extension, from female teachers" (p. 462). Studies dating as far back as 1909 (Ayres) expressed concern about the "feminization of our schools" claiming that schools are "better fitted to the needs of the girls than they are to those of the boys" (p. 7), and thus the reason for the number of boys failing and dropping out. Comparisons of reading achievement in cultures where there are more male teachers, Preston (1962) reported a reversal of sex differences in reading achievement. The reading achievement of 1,453 boys and girls from Wiesbaden, Germany were compared to 1,338 boys and girls in Philadelphia, United States. The boys outperformed the girls in reading comprehension in Wiesbaden where there were more male teachers. In Philadelphia, where there were more female teachers, the girls scored better than boys on reading achievement. However, with the exception of the Grade 6 boys in Germany, overall the children (boys and girls) in Philadelphia, where there were more female teachers, outperformed the children in Wiesbaden. It appears that having a female teacher did not negatively affect the reading performance of the boys.

Some 40 years later, Sokal, Katz, Adkins, Grills, et al. (2005) contested the notion that boys' achievement in reading would improve with male teachers. They studied 18 inner-city second grade boys in a pilot study of a 22-week reading intervention program with either a male or female research assistant. They found that the children made gains in their reading performance, and that the sex of the teacher had no effect on the progress made by the boys. Similar results were found in another study conducted over a 10-week period with 175 third and

fourth grade boys identified as struggling readers (Sokal et al., 2007). Other explanations for the differences between boys' and girls' reading achievement include: a) the lack of choice in reading materials; b) that boys and girls prefer different genres; c) that boys and girls have different learning styles; and d) boys more than girls prefer more activity-based learning and technology (Gambell & Hunter, 1999; Millard, 1997).

Millard (1997), Smith and Wilhelm (2002) proposed changes in schools that were presumed to support boys' underachievement in literacy. Millard argued that boys need to establish a masculine identity toward literacy at home and in school; otherwise they are being prevented from equal access to literacy. Smith and Wilhelm reported,

> It can't be emphasized enough… gender is a socially constructed concept and that the systems of belief and gender roles are susceptible to transformation. If boys are in trouble in the area of literacy because of gender, then our systems of belief both about literacy and about gender can be changed in ways to help them. (2002, p. 18)

Since such assertions are based on emotional and personal perspectives rather than on evidence-based research, then before any action for change is taken, it is important to establish first whether the phenomenon of boys' underachieving in reading exists.

Documented and established was an indication of the persistent and contradictory claim that boys and girls differ in reading achievement. Three explanations were offered as possible reasons for these claims. The historical

notion that sex differences were the result of inherent genetic factors was challenged in favour of sociological explanations (Hollingworth, 1914; Woolley, 1910). However, biological differences continue to be proposed by some researchers as a reason for differences between boys' and girls' achievement and are linked to claims of differences in brain functioning (James, 2007; Maccoby, 1998). Other researchers claim that the gender gap in reading is the result of different perceptions and attitudes toward reading, wherein reading is considered a feminine activity (Dwyer, 1973; Gambell & Hunter, 1999; Millard, 1997). Yet another explanation proposed for gender differences has been the way results are reported and the design of reading achievement tests. A recent meta-analysis by Lietz (2006a, 2006b) indicated over half of the difference in boys' and girls' reading achievement could be accounted for by test design. Variations between and among research findings and conclusions have resulted in conflicting claims about a boy-girl gap in reading achievement. The confounding evidence from these studies is confusing to say the least.

In order to make sense of the contradictory evidence from the literature review, I first established if there is a girl-boy gap in reading achievement and then tracked when this phenomenon first started. When a gender gap was established, I examined the explanations that accounted for the gap. The methodology used to conduct this study is discussed next.

Chapter Three:

Methodology

The purpose of this study was to analyze, synthesize, and interpret the evidence-based research on the comparative claims of boys' and girls' reading achievement from 1890, the approximate time of compulsory education in North America (Ayers, 1909), to the present. Scholarly research over the past 12 decades examined directly and indirectly sex differences in reading achievement from a variety of perspectives, methodologies, and sample sizes. A historical analysis for a period exceeding 100 years encompasses a large time span for educational research. A concise and comprehensive historical synthesis was important for a number of reasons and some of these are explored next.

Systematic Evidence-based Historical Interpretive Approach

A research synthesis and interpretation of the existing evidence over a substantial period would provide a comprehensive understanding of the issue of sex differences in reading achievement (Petticrew & Roberts, 2006). An important quality of historical research is that it goes beyond the narrative story of documenting events. Historians combine "strong storytelling with an organized rationale" (Krathwohl, 1993, p. 504). Historians adduce a rationale by assembling, organizing and interpreting events that correspond to a proposed explanation (Krathwohl, 1993). Making sense of a large body of historical information is like putting together a jigsaw puzzle that requires assembling hundreds of pieces, assessing and searching for key pieces that fit together in order to get a clear picture. A historical perspective on the issue of sex

differences in reading achievement was fundamental in order to clarify some important questions regarding the boy-girl reading gap. For instance, when were differences between boys' and girls' reading achievement first studied? How was reading perceived and how was it measured over the past 100 years or more? What were the prevailing views of learning and how did these influence the way reading was taught? Hence, a systematic, evidenced-based, historical, and interpretative (SEHI) method was used in my study. Unlike a literature review that may be "a partial review of a convenient sample of the author's favorite studies" (Petticrew & Roberts, 2006, p. 6), my study was an exhaustive search for, and examination of the evidence on the topic of sex differences in reading achievement. The SEHI method included a systematic, comprehensive search for evidence-based primary studies that provided a critical appraisal and interpretation of key findings (Petticrew & Roberts, 2006). Six phases characterized the SEHI methodology employed in my study.

Stage 1: Identification of search engines and use of snowballing.

Using Boolean logic, the major and relevant search engines were methodically investigated. Logical relationships were constructed between three operations (*or, and, not*), in order to expand or restrict the studies retrieved from the major databases. The search engines examined included: CBCA Ed, ERIC, SAGE, Education Abstracts ProQuest, Psych Info, Academic Search Premier, Google Scholar, and Web of Science.

Using search engines and databases to secure current studies was fairly straightforward. However, it was necessary to hand search references for

retrospective inquiries as keywords for historical studies were not rigorously entered into the electronic databases. Creative searching, such as "snowballing," was necessary; a method that uses the references from retrieved studies to further locate findings on the same related topic (Petticrew & Roberts, 2006, p. 121). It was critical to keep track of the number of hits, and the types of studies found. Accurate and complete records were kept to prevent duplication and to allow for repeated checks on previous searches or review initial hits as needed.

Stage 2: Identification of criteria for inclusion of research studies.

Hundreds of hits related to my research question were found. Through the use of predetermined criteria, screening out dozens of studies, limiting and sorting through the multitude of findings helped to focus the collection phase, and narrow the number of cited studies into a relevant and manageable number. The criteria used for the inclusion of studies were adopted from the work of Creswell (2008), and Lincoln and Guba (1985).

Lincoln and Guba (1985) proposed a strategy to determine the trustworthiness and authenticity of findings from qualitative research that is based on well-established criteria for judging quantitative research that includes: internal and external validity, reliability, and objectivity. To establish the accuracy and validity of findings, Lincoln and Guba (1985) described four criteria: credibility, transferability, dependability, and confirmability. In a credible study (internal validity) the researcher is persistent in the data collection phase, seeking out possible sources of distortion, and either crosschecking or triangulating different sources, using different collection modes, and sometimes

involving multiple investigators (Denzin & Lincoln, 2000; Lincoln & Guba, 1985). Transferability (external validity) is evaluated within the context of the study by determining whether the descriptions provided are sufficiently clear and detailed so that other researchers are able to use the methods described for application elsewhere (Lincoln & Guba, 1985). Dependability and conformability (reliability and objectivity) involves the evaluation of the described accounts of the results and conclusions so that others reading the findings and results would arrive at similar conclusions. Although Lincoln and Guba (1985) have acknowledged that using the four criteria to assess the trustworthiness and authenticity of studies has advanced rigor in research, after continued reflection around the debate on quality standards, they have developed new criteria for the evaluation of authenticity and fairness. However, new criteria have yet to be detailed and the "Strategies or techniques for meeting and ensuring them (the five criteria) largely remain to be devised" (Schwandt, Lincoln, & Guba, 2007, p. 24). Since Schwandt, Lincoln, and Guba recognize that they have not been able to generate distinct techniques to test the proposed criteria, they submit that the developed standards should be regarded as speculative (2007). Hence, I decided to rely on Lincoln and Guba's conventional four criteria that describe rigor in naturalistic inquiries (1985) and the well-established criteria for quantitative studies.

I also drew on the work of Creswell (2008) who focuses primarily on three processes to determine accuracy of findings, triangulation, member checking, and external audit. Triangulation refers to corroborating evidence that supports the

accuracy of a study. Using multiple sources ensures that the study is both accurate and credible. Member checking refers to whether the researcher(s) checks the accuracy of the findings with the participants in the study (not possible in my study). An external audit requires an outsider to review and check different aspects of the study to determine whether the findings are grounded in the data presented and decisions are justified. Since my study involved many different types and kinds of studies across an expansive period of history, criteria setting needed to be a flexible process. For example, it was difficult to find studies on boys and reading achievement from the turn of the 20th century, therefore I needed to be more yielding with the criteria when retrieving historical investigations. In contrast, in the last decades when thousands of studies were conducted on gender and reading achievement, I found it necessary to be more rigid and adhere to the predetermined criteria. Collected studies were evaluated for trustworthiness. These are some of the criteria that affect credibility that formed the basis for my initial searches:

1. Findings needed to be evidence-based; judicious care was implemented to make decisions about the practices used in the study. Evidence-based does not refer only to research findings and results from randomized or quasi-experimental research (Slavin, 2002). However, evidence-based studies need to have credible findings and results where the researcher engaged in the research for a prolonged period, crosschecked results, or used multiple sources to draw conclusions.

2. Methodologies needed to be clearly articulated. Although it would be impossible in a classroom to duplicate an exact method and expect to get the same results, the methodology had to be described so that other researchers can understand how the study was conducted and how the data were collected.
3. The results and conclusions required support by the data, such that other researchers could corroborate the results and conclusions. The data may have been audited by an outsider or—sufficient evidence provided such that outsiders reading the data would draw similar conclusions.
4. Primary studies only were included. Therefore, advocacy and second-source papers were eliminated from the final comparative analysis, synthesis, and interpretation.
5. Robust studies that met all four criteria (1 to 4) were included in the research.

Stage 3: Specification of database search terms.

Key search terms (e.g., boys and reading) were entered into search engines. Studies from the databases were scanned and recorded using the pre-determined criteria. Studies that fit the criteria for each approximate time period were procured, read, with accurate bibliographic information recorded, and findings summarized for further examination.

Stage 4: Determining time periods.

Each time period is approximately 30 years. However, time periods were not randomly chosen but were determined from the evidence-based studies that indicated a significant shift in reading theories and/or practice. For example, the researchers in the post WWI years (Chapter Five) were more skilled at measuring reading comprehension, as they relied more on standardized tests that were developed during this time. Measurements of student progress in reading moved from testing sub-skills of word lists, spelling and handwriting noted in Chapter Four to standardized reading comprehension passages observed in Chapter Five. In Chapter Six, the growth of reading as a separate field from psychology came about from the advancement in reading research between the war years. Research findings in Chapter Seven focused on the extensive analysis of statistical information and conflicting evidence from national and international investigations.

Studies collected over the 120 years covered all the general types of research including: descriptive, explanatory and experimental. Descriptive studies involve methods that collect information in great detail from stories or observations of individuals. Thick descriptions are used to discuss the interpretations and allow a reader to draw meaning from the experiences described. Case study is an example of a descriptive method (Creswell, 2008). Explanatory methods explain the relationships between and among variables. Statistics or descriptions are used to show the tendency or pattern of how variables are related. Examples of explanatory methods are survey or

questionnaire designs (Creswell, 2008). Experimental methods use random control tests (RCT) or quasi-experimental designs. Quasi-experimental studies are arranged into clusters that may be randomly assigned, but include specific groups (e.g., classes or schools). In both experimental methods, RCTs and quasi-experimental, the researcher establishes possible cause and effect relationships between and among independent and dependent variables. Every attempt is made to control extraneous factors and conclusions are made regarding the independent variable(s) and probable effect(s) on the dependent variable (Creswell, 2008). The major findings from the narrowed set of selected studies for each time period were charted and further analyzed and synthesized. Careful, thorough, and complete examination of the final selection of studies for an approximate time period was critical in order to determine themes.

Stage 5: Organization of analyses and syntheses.

Analyses and syntheses began with a very careful reading and re-reading of the studies that met the criteria through a systematic and comparative examination of the connections and interconnections between and among the selected studies from which relevant themes emerged. My supervisor and supervisory committee validated and crosschecked the findings recorded in each time period. I adopted the three-step process detailed by Petticrew and Roberts (2006) during the analyses and syntheses phases. These steps include: (a) organizing studies into logical themes, (b) analyzing the findings within each of the themes, and (c) synthesizing the findings across all the studies for each time period.

Organizing the studies.

Meaningful themes were created that pertained to the proposed research question. Unifying ideas were grouped as studies were located and assessed.

Within theme analysis.

Careful reading of the studies within each theme was completed. For example, all studies that supported no sex difference in reading achievement were identified, classified and grouped. Similarly, all studies that claimed girls were superior in reading achievement were sorted and assembled. A review of the studies in each theme allowed me to become familiar with the details of the studies and to determine whether the investigations fit within the theme established for each time period.

Cross-theme synthesis.

Reading across the research motifs and writing an overall summary of the variation of themes within each time period provided a cross-sectional view. A cross-sectional view offered an overall summary of all of the evidence-based studies that met the criteria for a given time period. The differences between and among the research findings in a particular time period provided clues either to explain or account for variations in claims of differences in boys' and girls' reading achievement. Evidence was assessed in order to make sense of the complexity of the issue. The accuracy of the themes and findings were confirmed and validated by my supervisor and supervisory committee. It was important to connect the themes into logical recurring patterns that provided insights for the final interpretation (Creswell, 2008).

Phase 6: Interpretation.

The purpose of the interpretation section was to make sense of the data and reveal what had been learned from the search and categorization of the data. Making sense of the data and findings involved stepping back and looking both locally and globally at the larger interpretations of the conclusions drawn from the findings and results of the question, what the evidence said about girls' and boys' achievement in reading? An overview of the interpretation of the data as well as an interpretative reflection on the literature and the connections between and among findings and results were reported. In addition, weaknesses and limitations of the study were noted related to the data collection process, the methodology used, and questions that remain unresolved. Implications included considerations for important decision-making for teachers, parents, administrators, and policy makers. Additional questions for further research were outlined.

The systematic study of the comparative claims of differences in boys' and girls' reading achievement required an indefatigable attention to detail and thus a systematic, evidenced-based, historical, and interpretative (SEHI) method was my constant touchstone throughout this challenging, exciting and groundbreaking study.

Chapter Four:

The Emergence of Compulsory Education

and the Roots of Reading Comprehension: 1890–1920

Introduction

My main goal in this chapter is to situate within the context of the time period 1890–1920, the major research findings that compared boys' and girls' reading achievement. To accomplish this purpose, I first address some key conditions that influenced the adoption of compulsory education. These include: growth of industrialization and immigration, a changing image of women and subsequent need for better childcare facilities, the creation of new child labour laws, a desire to improve opportunities for education and working conditions, and continued industrialization that required increased skills such as reading and problem solving. Next, I explore how the adoption of formal education, and a growing interest in scientific methods affected theory and pedagogy of reading. With increased demand in the workforce for literate skilled laborers, there was a growing interest in reading acquisition. Thus the roots of reading comprehension emerged. Lastly, I examine how assessment of children's academic progress in school gave rise to an interest in sex differences in reading. Compulsory education required that both boys and girls attend school wherein assessment of academic performance was initiated. Although there is evidence to indicate that boys underachieved in spelling and word activities, such as dissected sentences, opposites and word completion (Earle & Thorndike, 1903; Pressey, 1918), there is

no proof that boys were underperforming in reading comprehension. This chapter will be organized in accord with the purposes specified.

Socio-Economic Conditions That Influenced Compulsory Education

The end of the 19th century witnessed the growth of industrialization and mass migration of people from rural areas to cities and from Europe to North America. Although the proliferation in the growth of machines and factories provided jobs and a hope for a brighter future, in reality, it was the citizens with affluence and power who enjoyed the "good times." For the poor, life was filled with long days of strenuous work for meagre wages, often in monotonous and unskilled jobs on assembly lines dominated by machines.

Manual labour was gradually replaced by machine-based manufacturing that increased production capacity. The mechanization of industry contributed to changes in the way people lived their lives.

> Within a span of years . . . a nation in which most men had little or no connection with the land, were largely without property, and worked for other men. The pattern of their lives was now determined, not by the rising and setting of the sun and the slow, steady cycle of the seasons, but by the sharp blast of the factory whistle and the relentless tempo of the machine. (Degler, 1967, p. 2)

For the poor, unemployed, and visible minorities who endured tremendous hardship on a daily basis, 1890–1920 was a period of poverty, overcrowded cities and poor living conditions. The early 1900s was also a time of racism, segregation, and voting restrictions (Degler, 1967; Lears, 2009).

World War I (1914–1918) influenced worldwide socioeconomic conditions and changes in attitude toward women. Continued technological advancements and industrial production geared toward supporting the war effort, furthered the increase in the production of goods. Employment opportunities in cities continued to stoke the growth of urbanization and to signal a major economic boom. As many able-bodied men were away fighting in the war, women flocked to jobs in factories. Since women were proving their strength and independence by working in factory jobs that had been traditionally done by men, attitudes toward women working outside the home started to change. Newspapers publicized notices: "Women Wanted!" and for the first time ever openly requested "coloured" women (a term used at this time period for African American women) (Greenwald, 1980). For the most part, Black women were given jobs only after White women rejected the work: most often the least enviable positions, the hardest and often the dirtiest. Women were demanding equal rights, organizing and participating in strikes, demanding better wages and working conditions (Greenwald, 1980).

When WWI ended and the soldiers returned home, there was a shortage of jobs for men and rising unemployment.

> While the war emergency enhanced labor's bargaining power and secured women new opportunities, the war's end revealed the entrenched strength of management and the relative weakness of women workers against managerial rationalization strategies, on the one hand, and the hostility of male workers, and job security, on the other. (Greenwald, 1980, p. 243)

A decade later, a severe worldwide economic depression occurred that lasted until World War II.

Urbanization and industrialization in the latter part of the 19th century did more than bring about changes in the way work was being done in Canada and the United States. Also changing was a reconceptualized image of "childhood" in society. One of the new constructs introduced was the idea that children were powerless and passive. This new image of the child proposed that children were innocent and that gentle and loving caregivers could mould their "plastic natures." Adults were viewed as protectors who needed to safeguard children's health and welfare. The romantic notion held by social activists was that childhood was separate from adulthood, and should be a time of innocent and imaginative play (Cunningham, 1995).

The romantic view of childhood held by social activists at this time was influenced by European philosophers and educators including Rousseau (1712–1778), Pestalozzi (1746–1827), Froebel (1782–1852), and Montessori (1870–1952). J. J. Rousseau was born in Switzerland and moved to France in his early 20s. There, he was instrumental in the early reform movement in education (Boyd & King, 1965; Monroe, 1911/1968). The early reform movement was a time when new ideas on the worth of the individual gathered momentum, the privileged and governmental abuses were challenged, and a widespread psychological change was noted, all of which influenced education. One of the most significant results was the gradual acceptance of the view that education ought to be the responsibility of the state. Rousseau, who had a passion for

botany, cultivated the metaphor that children were like plants, needing suitable conditions for their physical, spiritual and educational growth.

> Plants are fashioned by cultivation, man by education. . . . We are born weak, we need strength; helpless, we need aid; foolish, we need reason. All that we lack at birth, all that we need when we come to man's estate is the gift of education. (Rousseau, 1762/1974, p. 6)

Rousseau stressed the importance of helping children to grow and develop naturally. He started the change in the dogmatic role of the teacher toward children and what and how they were to learn as well as the view that education was not just for the privileged. The thinking was that children, in order to grow strong and wise, needed a proper education (Janovicek & Parr, 2003).

Among those influenced by the work of Rousseau were Johann Pestalozzi and Friedrich Froebel. The theories of the Swiss reformer, Pestalozzi, are evident in primary education today. He advocated for the development and integration of the "organic", the intellectual, moral, and physical and that education should be a development of the individual's self-power in the context of experience. Froebel visited Pestalozzi's boarding schools for poor children and orphans of the Revolutionary War in Switzerland (1798) and continued the educational reform started by Rousseau and strengthened by Pestalozzi's pioneering work on education for poor children.

Froebel was the founder of the kindergarten movement and promoted an emphasis on constructive play and self-activity in early childhood. The fact that Froebel was devoutly religious is evident in his pedagogical writings. In essence,

it seems that school was a place for both "dictating and giving way." In other words, school was to be a place that did not focus solely on the transmission of knowledge but rather on the development of character and the motivation to learn. He suggested that the true nature of the child was full of goodness, and therefore advocated for new methods for teaching preschool-aged children, methods that included singing, games, fun, and imaginative play. Within 25 years after Froebel's death in 1852, kindergarten classes were established in major cities in the western world. Incidentally, Toronto was reportedly the second city in the world to make kindergarten a part of the regular school system in 1878 (Phillips, 1957, p. 422).

Maria Montessori (1870–1952), the first woman in Italy to become a medical doctor, founded the educational system that bears her name. She also supported the increasing focus on the child, which began in the decade before the 20th century. Montessori further proposed that the key to progress, and a new civilization would be found through the study of the child. "This is the divine choice that no one can still, and which calls men with a loud voice, calling them together to gather round the Child" (Montessori, 1936, p. 236). Montessori (1936) claimed that adults would be better-adjusted citizens if they were exposed to an environment that fostered social and academic skills during childhood. To release the human potential, children needed to be exposed to quality childcare. Clearly the motivation for excellent childcare and improved academic skills indicated a cry for compulsory education that followed in the years to come.

However, there was a contrast between the childcare provided for the working poor in Canada and the United States, and the progressive preschool movements that developed in Europe during the late 1800s (Polakow, 2007). Families who worked long hours in factories could not afford to send their children to expensive facilities, but instead often needed to resort to placing their children in "orphanage asylums", operated by charities (Polakow, 2007). As soon as children were old enough they were sent out to work to help support the family. Not surprising then, the type of childcare program offered to children depended on the amount of money a family could pay. Differences existed between the middle class who could afford nursery schools that focused on child pedagogy and a program of more custodial care for working poor families. For poor and orphaned children, life was not filled with play such as games and songs enjoyed by children in progressive early childhood schools, but one filled with long days of hard work. In fact, throughout parts of England and in some large cities in the United States, many orphaned children and children of poor families were sent to work on farms in Canada and the United States. For example, during the early 1900s, reports indicate that 60,000 children were sent from New York City to Mid-Western farms (Cunningham, 1995 p. 150). The British sent another 80,000 children to Canada because there was a demand for child labour on farms and for domestic work on Canada's western frontier (Cunningham, 1995 p. 150). Relocating children was justified through the claim that removing children from dirty cities and resettling them in a "back to nature" environment of clean air and open spaces would improve the children's health and welfare (Cunningham,

1995). During WWI, nurseries operating within industries allowed women to work and have their children looked after. Unfortunately, when the war ended, there was a campaign to remove women and children from factories, regressing to the view prior to the war that a woman's place was in the home. This perspective further stigmatized poor working mothers. It wasn't until the Great Depression (1929–1940s) that childcare became a national concern in the United States (Polakow, 2007).

Promoters of the "save the children" movement created tension between the capitalists who owned large factories and an emerging social consciousness that proposed that children be spared from the fate of cruel, unscrupulous taskmasters. But, industrialists viewed child labour as necessary in order to sustain a competitive edge. Since children were paid less than adult workers, hiring children helped to maintain a low price for goods (Cunningham, 1995). The lobbyists' policy of targeting the withdrawal of children from the workforce contributed to a widespread concern for the increasing number of idle and delinquent children in large urban cities (Cunningham, 1995). A growth in population due to industrial expansion transformed the social landscape of large cities like New York. Increased crime, housing shortages and disparity between the poor and rich placed new strains on the social fabric of life in cities (Katz, 1976). In an effort to impose a solution to the hectic urban conditions, there was widespread advocacy for a publicly supported school system (the term used in the United States was common schools). Consequently, the common school movement became a moral crusade for some school reformers such as Horace

Mann (1796–1859). Mann, sometimes referred to as "The Father of American Education," was the first Secretary of Education. He proposed that public education would be the "great equalizer" (Cremin, 1965). The position held by the campaigners of public education was that

> free publicly supported common schools would unite Christian morality with democratic patriotism; the common school would stamp out the evils of ignorance, crime, vice, and aristocratic privilege; and finally, the common school would not only assimilate the immigrants but also transform them into virtuous, productive American citizens. (Katz, 1976, p. 15)

Sutherland's (1976) landmark text, *Children in English-Canadian Society* laid the groundwork for a "new social history" one that combined the history of education and the history of the family (Comacchio, 2000). His systematic study of childhood in Canada examined the changing concept of the child in English-Canadian society at the turn of the 19th century. Sutherland (1976) postulated that the changing concept of the child was focused around the view that a prerequisite to the betterment of society depended on proper childrearing, including an opportunity to go to school. These themes were more prevalent in English-Canadian society than in the history of childhood in Quebec (Janovicek & Parr, 2003). In Quebec, children were considered part of the extended family and were integrated early into the labour force. Compulsory school attendance was delayed compared to other parts of Canada, due to the need for children to work and maintain a traditional family farm (Janovicek & Parr, 2003). Although

national policies and practices did not exist in Canada during the 1890–1920 period, the child-saving movement was the beginning of a general reform that began with charity actions and then slowly fell under the control of the government in later years (Janovicek & Parr, 2003).

> Industrialization, in Britain first and then in other countries, led governments into closer control of child labour than previously. A mixture of motives, including child unemployment, led them to impose compulsory schooling, which vastly increased the range and scope of the state's activities. (Cunningham, 1995, p. 159)

The development of public schools and policies regarding mandatory attendance in Canada and the United States differed between and across regions, populations, and rural and urban centres. For example, slaves in the Southern United States received little education prior to the Civil War (1861–1865) and were forbidden by law to learn to read and write (Venezky, 1996).

Beginning in the Eastern United States, the first public school systems became operational. Specifically, Massachusetts began its first public-school system in 1837 and a similar system began in Connecticut in 1843 (Dewey, 1903). Compulsory attendance in the United States was inconsistent, especially in rural areas, and few children went beyond Grade 8.

> In 1890, for example, fewer than 10% of students were enrolled in secondary schools, and an even smaller percentage of females was enrolled. However, there was a rapid acceleration in high school

> attendance, and by 1920 universal high school education was required in most states. (Allington & McGill-Franzen, 2004, p. 6)

Accelerated growth and improvement of schools in Canada in provinces such as Ontario, Nova Scotia, Manitoba, and British Columbia occurred at the same time as the revival of the publicly sponsored schools (common schools) in the United States. The provincial appointment of chief education officers signalled a new energy and a desire for direction in education. Appointments were made in Nova Scotia and Ontario during the time period 1840–1870; Manitoba and British Columbia followed in 1870. Some officers followed the zeal and commitment of American school reformers such as Horace Mann (Phillips, 1957). Similar to the United States, but at a later time, since Canada was more sparsely populated and particularly so in the prairie provinces, there were differences as to when formal education was institutionalized and when compulsory attendance was adopted. However,

> by 1934, attendance was compulsory up to age 15 in Saskatchewan, Alberta, and British Columbia; to age 16 in Ontario and Manitoba; to 13 in Prince Edward Island; and to 14 in rural Nova Scotia and to 16 in urban areas. (Baldwin, 2008, p. 384)

Newfoundland joined Canada in 1949 and though compulsory schooling was established in 1942, it was not enforced until the mid-1950s (Phillips & Norris, 2002). The general consensus for advocating for the removal of children from the workforce and instead to place them in school was the position that an education would allow children the possibility to improve their lives and provide a brighter

future and ultimately a better society (Cunningham, 1995; Montessori, 1936; Sutherland, 1976).

In addition, due to increased demands caused by industrialization and global competition, more skilled labourers were needed. Schools were expected to provide workers with increased abilities (Allington & McGill-Franzen, 2004). Publicly supported schooling and compulsory attendance were considered the chief educational influences upon the young and the best organization through which to promote democracy and the development of a better society (Cremin, 1965). The tumultuous political, social, and economic environment in Canada and the United States led to changes in attitudes toward the status of children and the role of education in society. From the emerging concept of childhood and a focus on education for the poor, arose compulsory education in which a whole new set of developments took place such as educational standards and the assessment and measurable achievements of individual academic skills, including reading.

Education and the Shift Toward the Individual

Venezky, in "The Development of Literacy in the Industrialized Nations of the West" (1996), outlined literacy trends from a historical and cross-cultural perspective. Prior to the 19th century, literacy was primarily accessible to individuals from religious groups and to the upper-class elite who had the time and money to pursue such endeavours. In the early stages of industrialization, few technical skills were needed, and therefore, there was little pressure for workers to improve their literacy skills. Furthermore, employment of children in

factories allowed poor families an opportunity for increased household income. As industry increased in complexity of production, there was an increased demand for literate workers. Reliance on oral communication became secondary to the skills of reading and writing, as new jobs required new skills. With technological improvements in the printing press, books were made affordable and accessible to the masses. Education was seen as an important contributor to increasing worker production and the building of the nation's place in the world marketplace and school was seen as a way to create future success (Venezky, 1996). With the promotion of formal education and compulsory attendance, children were increasingly encouraged to attend school. Interestingly, the level of a person's education became a convenient and universal criterion for determining social stratification in America in the mid-19th century. Primary education belonged to the lower classes, the middle class could afford secondary education, and higher education belonged only to the upper middle class and the elite (Hobsbawn, 1992).

Increased standardization of factories paralleled the shifting changes in schools and teaching methods (Heath, 1996). From 1850 to the early 1900s in the United States, the organized machinery of educational institutions gradually developed, including teachers' associations, teachers' journals, superintendents, and state universities. Also, there were efforts to make improvements in school facilities, school playgrounds, textbooks, and equipment; for example, maps, globes and scientific apparatus (Dewey, 1903). Teacher-based assessment was replaced with standardized tests, established grades supplanted non-graded

schools, and the focus on what and how students learned shifted attention from teaching students "to think" to teaching specific measurable skills (Heath, 1996).

In the early 20th century some American philosophers, in particular John Dewey (1859–1952) a leading exponent of pragmatism, presented an alternative position on the purpose of schools. He founded and directed the Laboratory Schools at the University of Chicago wherein he opposed the rigid school subject lines. Rather than adhere to prescribed curricula, Dewey promoted an activity-based program that allowed children opportunities to develop their potential to the fullest extent as well as their individual worth. He urged schools to promote independent logical thinkers. He viewed education as a moral and philosophical endeavour and proclaimed that schools should be leaders in democracy, furthering "the emancipation of [the] mind as an individual organ to do its own work" (1903, p. 193). He was opposed to the transferring of school-based decision-making to superintendents and board authorities on issues of curricula and the conditions of schools, declaring that schools should exemplify the democratic process by solving these issues themselves. One of the most prevalent views of Dewey is education is life and not merely a preparation for life. The purpose of an education was not to learn a set of pre-determined skills but to realize one's potential and to use these skills for the greater good. In this way education and school are instrumental in creating social change, or what Dewey termed " social reconstruction." This progressive thinking during the early decades of the 20th century influenced public school systems to varying degrees in many places around the world.

However, an influx of immigrants and southern Blacks in schools between 1910 and 1913, one of the greatest immigration periods in North American history, brought about a desire for standards in schools. Concern over a great influx of students from non-English cultures speaking a variety languages, school boards made the decision that students needed to adhere to a rigid prescription of subjects, to follow acceptable behaviours, to learn proper English, and to become good citizens (Heath, 1996). In the United States, although the major cultural events of immigrants were "tolerated," the educational system was used to promote political socialization. This socialization included learning English, applying for citizenship and pledging allegiance to the American flag (Hobsbawn, 1992). The concern for schools was that they should be helping children from non-English homes to learn the necessary literacy skills to be successful in the competitive market economy. "Key terms of both manufacturing and schooling came to be 'standards,' 'management,' 'control,' 'rational,' and 'predictable'" (Heath, 1996, p. 9).

Research in educational psychology brought about the thinking that education outcomes could be measured by assessing the individual learner. Teachers were required to impart specific knowledge within approved standards and academic performance (Heath, 1996). At the turn of the 20th century, theories on reading were formulated and explorations into how reading should be taught were examined and studies were conducted on sex differences and literacy activities.

Classroom Practice and Reading Research: Divergent Perspectives

The scientific movement in Education began to flourish in the decade from 1910 to 1920. "Almost four times as many studies were published in this period as had been since 1881" (Stauffer, 1971, p. 443). Emphasis on measuring and providing evidence led researchers to assess individual skills through the use of instruments such as word-speed tests and spelling drills. What was measured was what was observable and testable (Bohannon & Bonvillian, 2009). In addition, standardized tests were being produced and theories of how students learn to read were proposed which led to debates on the best methods to teach reading. However, even though researchers such as Huey (1898) and Thorndike (1917) proposed reading theories and laid the foundations of reading as we know it today, "there was little concerted effort to marry research knowledge and instructional practice until much later in the 20th century" (Alexander & Fox, 2008).

The standard practice in schools for teaching reading involved a systematic and orderly learning of sub skills with a heavy reliance on oral recitation (Allington & McGill-Franzen, 2004). Children were regarded as having a passive role. Language as a skill was regarded to be like other skills, one that could be controlled and shaped by the environment. Development of language then was largely dependent on training rather than maturation. The teacher was the source of knowledge and s/he imparted this knowledge to his or her students (Binkley, Phillips, & Norris, 1995).

However, examination of the work of two reading researchers of the time, Huey and Thorndike, suggests that reading theories were proposing a different perspective—one of problem solving, with a heavy reliance on comprehension and silent reading. Edmund Burke Huey, one of the most creative pioneers of reading research, attempted to apply psychological methods to resolve conflicting perspectives regarding reading instruction. Huey (1898) initially researched eye movements and experimented with the speed at which children read words horizontally and vertically. He wanted to measure the efficiency rates of eye tracking during reading. Did children find it easier and faster to read left to right or by going down the page? Results showed that children were able to read faster horizontally if the words were between five to eight letters in length. Since most of the words in the English language are within such a letter range, Huey concluded, there is "no grounds for changing the present order of printing, as the words of the English language average only five letters each" (1898, p. 675).

Huey's seminal work, *The Psychology and Pedagogy of Reading* (1908/1968), is an intriguing mixture of science and practice (Buchner, 1909). As an information-processing activity, Huey examined the reading process from a physical and cognitive basis. In addition, he also explored and proposed methods for teaching reading. A summary of his pedagogical conclusions reveals insightful recommendations, some of which address issues discussed today, over 100 years later. In particular, Huey promoted the use of good literature in the classroom, with methods that emphasized teaching meaning through context and

that included more practice in silent reading than in read-aloud strategies that focused on practice and drills (Carroll, 1968).

Edward L. Thorndike, another prominent early 20th century American psychologist and researcher published a study in 1917 on reading, disputing the commonly held view that reading was a simple compounding of habits, a calling up of each word using sound and meaning in series to produce meaning and thought. Focusing on reading comprehension, Thorndike argued that reading was a complex process similar to that of problem solving in mathematics. From his analysis of responses collected from reading comprehension questions, he coined the notion of *powerful words*. Children's understanding of groups of words, and sometimes even single words, influenced their interpretation of an entire sentence or paragraph. Thorndike theorized that students interacted with texts; that there was interplay between the words in relation to one another. Reading went beyond the meaning of single words strung together.

> In educational theory, then, we should not consider the reading of a textbook or reference as a mechanical, passive, indiscriminating task, on a totally different level from the task of evaluating or using what is read. While the work of judging and applying doubtless demands a more elaborate and inventive organization and control of mental connections, the demands of mere reading are also for the active selections which is typical of thought. (Thorndike, 1917, p. 332)

Thorndike also promoted the practice of silent reading. He observed that students during choral reading and oral round-robin reading, common classroom

practices in the early 1900s, were not engaged in actual thinking or attending to meaning. He maintained that reading was more complex than learning a hierarchy of individual skills or the practice of rote memorization.

The theories proposed by Huey and Thorndike were in contrast to the classroom practices of the day. Since theory usually precedes practice, this difference is neither alarming nor unexpected. Rehearsed and oral recitations were common classroom reading practices in the early 20th century. For the most part, reading was assessed in terms of what was observable and measurable. In addition, from 1890–1920, with the inception of the scientific movement in educational research, studies were conducted that compared boys' and girls' literacy skills. Discussions of the major findings follow.

Early Influences and Research Findings Prompting Comparisons of Boys' and Girls' Abilities in School

An exhaustive investigation for and examination of the evidence on the topic of sex differences in reading achievement was completed using a systematic, comprehensive search for evidence-based primary studies for the time period 1890–1920. Collected findings were evaluated for trustworthiness using the criteria described in Chapter 3 resulting in a careful, thorough, and complete examination of three studies. A synthesis of the pertinent conclusions that relate to boys' and girls' reading achievement is described next.

The influence of Thorndike's studies on measuring individual abilities.

Unlike John Dewey who argued for the study of children's interests and capacities for learning, Edward L. Thorndike (1874–1949) began his study of the

laws of learning with animals and from there migrated into education wherein he argued that psychology and physiology were instrumental in the formation of intellect, character and behaviour. Building on his numerous studies of animal intelligence and transfer of learning, Thorndike's work evolved to individual differences, learning, and mental measurement. His handbook, *An Introduction to the Theory of Mental and Social Measurements* (1904), stimulated a series of studies on the application of psychology to the teaching and learning of arithmetic, algebra, reading, handwriting, language, and other areas.

It may be said that specialization in and isolation of perspectives marked the early years of research in education that overlapped in part with the period 1890–1920. In particular, it was a time when education became a subject of university research influenced by psychological testing and school surveys. Thorndike was known as the father of the measurement movement. He co-authored and edited a collection of studies with his Masters students from the Department of Educational Psychology at Teachers' College, Columbia University in 1903. Six of the studies examined differences between abilities in different subjects. Two studies specifically measured sex differences in arithmetic and spelling ability.

Arithmetic abilities and sex differences.

"The Relationship Between the Different Abilities Involved in the Study of Arithmetic" (Fox & Thorndike, 1903) included 77 students in high school, 28 boys and 49 girls (the location of the high school is not mentioned in the paper). The students ranged in age from 14 to 20 years. Students were given tests on

mental operations of addition with five sums, multiplication of four digits, fractions, rational computations, and two sets of four-word problems. The same individual administered all subtests, in the same manner, and within the same amount of time, using a fixed scoring criterion. The Pearson coefficient of correlation was used to determine the relationships between and among the different subtests (addition, multiplication, fractions, rational computation, and word problems). Coefficients between sexes were calculated separately. Fox and Thorndike (1903) concluded that the small differences between the relationships on each arithmetic subtest indicated that the abilities required in arithmetic are independent of one another with "separate abilities that together make up the composite 'arithmetical ability'" (p. 37). For example, students who did well on fractions did not necessarily have the same success in addition or multiplication. Girls performed about 5% better than boys. However, Thorndike and Fox (1903) noted that reliable conclusions could not be made, considering data were collected from a small sample and from a limited selection of schools. Thus the results were inconclusive.

Spelling abilities and sex differences.

Although claims had been made that girls were superior to boys in spelling, Earle and Thorndike (1903) in their study, "A Sex Difference in Spelling Ability," set out to measure the extent of these claims. The spelling ability of 1,600 boys and girls from three different school systems (locations were not indicated) was measured using four sets of words. The median scores for the boys and girls on each test of spelling were calculated. Earle and Thorndike

(1903) determined that assessing the median score was a better comparison than using a straight average, because comparisons could be made between all four spelling tests. Overall, results indicated that 33% of the boys reached or exceeded the spelling ability reached or exceeded by 50% of the girls who had the same amount of training. Furthermore, Earle and Thorndike (1903) concluded that the sex difference in spelling

> cannot be explained by the earlier maturity of girls, for in a majority of mental and motor tests, boys of the same age as girls surpass them. It does seem to be one feature of a general superiority of girls in tests involving the perceptive power (success in spelling is largely due to the ability to perceive and attend to the formation of words). Such are tests in marking out words containing certain letters or tests of memory-span, and in these also [we] find girls to be superior. (p. 49)

Interestingly, no references were found to support the claim that boys surpass girls of the same age on mental and motor tests.

In general, Thorndike (1903) and his students, after examining the transferability of performance rates between drawing, math, and spelling, found no significant correlations between subjects or between subtests in arithmetic. Performing well in spelling did not mean a student would automatically be competent in math or drawing. Strong abilities in addition did not mean a student would be proficient in fractions or math problem solving. Boys and girls did not differ on tests of arithmetic. However, girls outperformed the boys in spelling (33% of the boys reached or exceeded the spelling ability reached or exceeded by

50% of the girls). This was one of the first studies that measured differences in achievement levels between boys and girls in specific subject areas in school.

"Laggards" in the school system: Identification and interpretation.

With an increase in the perceived value of education, tremendous concern arose for the number of children who did not do well in school. Success at school was taken to be of great importance to the welfare of the country, because education was believed to provide children with a successful future. Ayres (1909) claimed that extensive studies (no references are made to specific studies) in Britain and the United States were done on the causes of children "lagging behind" (not progressing and therefore being held back) in school (p. xiii). It was noted that the number of "defective children" (this term is defined by Ayres, as children who can never be educated or become independent members of the community) did not account for the number of children who were not at grade level.

Using statistics from the Commissioner of Education Report of 1907, Ayers systematically studied what he considered to be the causes of "retardation" (retention). The results are outlined in *Laggards in our Schools: A Study of Retardation and Elimination in City School Systems* (Ayers, 1909). Ayres proposed that children in New York were not successful in school because of: 1) irregular attendance, 2) an influx of non-English-speaking children, 3) enrolment of children at an advanced age in Grade 1, 4) physical defects, 5) inefficient teaching, 6) frequent moves by the families, and 6) unsuitable courses of study (pp. 1–7). Included in the concerns raised was the number of boys who were

failing and dropping out of school. After examination of the records of 7,624 students in the four grades in high school (presumably Grades 9–12), Ayres noted that 43% of the population was boys. For every 100 girls who entered high school there were only 79 boys. He declared that these statistics were quite alarming, noting that the United States "is the only nation having more girls than boys in her secondary schools" (1909, p. 150). The common explanation was that boys had better opportunities to get jobs, and therefore they dropped out of school at an earlier age (Ayres, 1909).

Statistical information comparing boys' and girls' achievement in elementary school had been quite rare prior to the 1907 Commissioner of Education Report, and hence the hypothesis that boys were being retained had not been tested. Using statistics from 752 towns and cities across the United States, Ayres' results indicated that although boys slightly outnumbered girls in the first grade, by Grade 8 there were more girls than boys. Examining the number of children who were retained indicated that boys exceeded girls by 13%, revealing that more boys than girls repeated grades in elementary school. Ayres concluded:

> We have indisputable evidence that there is more retardation among our boys than among our girls in the elementary schools. As this condition exists before the close of the compulsory attendance period it can have no relation to the alleged greater desire for seeking employment on the part of the boys which has often been put forward as an explanation of the more rapid falling out of school of the boys. There are more repeaters among the boys than among the girls and the boys leave school earlier and in

> greater numbers. This latter condition arises in the elementary schools and continues through the high schools. The percentage of promotions is less among boys than among girls. (1909, p. 157)

Ayres (1909) concluded that schools were "better fitted to the needs of the girls than they are to those of the boys" (p. 7). However, he cautioned that it would be false to attribute the "retardation and elimination" of boys to the number of female teachers in the school system. Making a comparison was not possible, since there were no schools where only males taught. Interestingly, Ayres did not indicate why children were retained in school or what measures were used to assess them.

Ayres (1909) proposed a series of remedies. He argued that the government needed to enforce attendance laws and demand better school census taking. Since school was optional from age 14, many children, and in particular boys, were leaving for the workforce. Also, children were starting school at different ages, some entering Grade 1 as late as ages 7 or 8 years. Ayres advocated for compulsory education beyond the age of 14 and for a more systematic way of enforcing attendance. Ayres was also very concerned about the physical health of children, calling for medical inspections on a regular basis. He was critical of the "lock-step" approach that advocated children move through the curriculum at the same sequence and pace (Manzo & Manzo, 1990, p. 25) and recommended flexible grading and differentiation to fit the abilities of the students. Ayres argued for continuous progress and programming that matched students' levels. Recommendations were made for varying the amount of time

given to students to complete their work, thereby providing slower students more time to bolster their skills.

Gulick and Ayres (1913) followed up on Ayres's original recommendation of medical inspections in schools. A number of cities in the United States were starting medical inspections and introducing the concept of a school nurse. The idea was that if children could be safeguarded from diseases, they would be happier and better able to take advantage of the free education being offered them. With urbanization and compulsory education, children were in close quarters with one another for longer and schools were becoming centres of infection. However, medical inspections were not solely intended to detect diseases but were also used for preventative purposes. Gulick and Ayres claimed that there was a direct link between children lagging behind in school and their health. Medical inspections in schools for defective eyesight, hearing loss, enlarged tonsils, swollen glands, and poor oral health were initiated. Gulick and Ayres proposed that schools, as places of public trust, needed better child protection measures to ensure the best learning environment and ultimately a better society for the future.

1918–Expanding on research methods: Individual mental measures.

Luella Winifred Pressey (1918), a psychologist, conducted a study on sex differences in the general intelligence and special abilities of school-aged children. Concerned by the lack of adequate studies and the limited sample sizes of those that existed, Pressey used large, random, equal samplings from both sexes. Using a group scale (a mental survey developed at Indiana University), Pressey assessed 2,544 children: 1,342 girls and 1,202 boys, ranging from ages 8

to 16 years. The group scale consisted of 10 tests, each with 20 items including: 1) rote memory for words, 2) logical selections between words, 3) practical arithmetic, 4) opposites, 5) logical memory, 6) word completion, 7) moral classification, 8) dissected sentences, 9) practical information, and 10) analogies. Pressey's sample was drawn from three small cities in Indiana. Results from the younger ages (8 and 9) were excluded from the final analysis because the samples sizes were not equal. A greater number of boys in the early years were held back and so by Grade 3 there were more girls than boys. Pressey also excluded results from the 16-year-old group since more boys dropped out to go to work, the numbers at the upper ages were also unequal. For the remaining groups, the number of correct items on the 10 tests was then added. Pressey calculated the averages and the medians. To test the theory of variability, she compared the percentage of boys above the girls' 90th percentile and below the girls' 10th percentile. The boys averaged 10.4% above the girls' 90th percentile and 20.6% below the girls' 10th percentile. In addition, using only the data from ages 10 to 14 years, Pressey drew four main conclusions:

1. Averaged general abilities were higher for girls than boys. The boys' test scores were about 6 months below that of the girls. An explanation provided was that girls matured before boys.
2. On the 10 subtests, the girls excelled at rote memory, opposites, word completion, dissected sentences, moral classification, and analogies. The girls were slightly ahead on tests of judgment, and on logical

memory. Boys were superior on arithmetic tests and practical information.

3. The distribution of the scores indicated a remarkably greater variability among boys than girls. Pressey concluded that the distribution differences could be due to a greater variability in the rate of mental development or general intelligence between boys than between girls.
4. The degree of the nature of the variability depended on the subtests, some showing greater variability than others. Pressey was sceptical of the concept of a "general" ability, and concluded that the differences between boys and girls depended on the skills being tested.

Pressey speculated that the measurements used in her study could have been easier for girls and that another scale using arithmetic, practical information, special relations, and puzzle tests would result in boys excelling. Pressey's findings supported other work at the time that proposed that men were more variable than women, suggesting that men more than women had a greater range in intellectual functioning (Ellis, 1904; Thorndike, 1914). Pressey noted, however, that it was conceivable that the comparative variability between the boys and the girls might be due to the nature of the tests used. Although differences between boys and girls are noted in Pressey's findings, with the girls performing better on word completion, opposites, and sentence completion, and boys on arithmetic, it is important to note that reading comprehension was not measured. Pressey's findings, therefore, were inconclusive with respect to determining whether boys and girls differed on reading achievement.

Synthesis of Findings

The period between 1890 and 1920 was a time of continued urbanization, industrialization, and immigration. By the early 1900s in Canada and the United States most children from ages 6 to 14 years attended school. With an increased emphasis on the individual and on improving standards, productivity began to be measured both in the workforce and in schools. In schools, students were tested using standardized testing measurements and new test designs were being developed and implemented with a focus on empirical evidence. Educational testing in schools was viewed as the panacea to solve all school issues.

Three robust studies from the early 1900s on sex difference in school achievement were analysed and summarized. Although Pressey (1918) concluded that generally boys surpassed girls on tests of arithmetic, Fox and Thorndike (1903) established that there was little difference in performance on arithmetic tests. Fox and Thorndike, however, noted that the sample size and school selections used were limited and inconclusive. Earle and Thorndike (1903) indicated that girls were superior in spelling, and Pressey's (1918) findings denoted higher achievement in rote memory, opposites, word completion, dissected sentences, moral classification, and analogies; evidence that supports the conclusion that girls surpassed boys in spelling and word activities. The findings of Ayres (1909) on how well boys were doing in school revealed disturbing data. Generally, in public schools more boys were being retained (13%) and more boys were not completing high school (17%). However, Ayres did not provide details as to why boys were leaving school early or what they left

school to do. Although there were suggestions as to why boys were not achieving in school, explanations and details were not provided. After studying the conditions in schools, a number of suggestions and remedies were made, with solutions ranging from increased medical inspections to changes in teacher practice (Gulick & Ayres, 1913).

Implications for Boys and Reading Comprehension (1890–1920)

Withdrawing children from the workplace and changing child labour laws led governments in Canada and the United States to impose compulsory education (Cunningham, 1995). Concern for how well students were performing in school, in particular boys, invited research on sex differences on attendance, retention, and achievement in specific subjects. The period from 1890 to 1920 was a time when educational researchers adopted scientific research methods to seek "truths" based upon empirical evidence, an optimistic way of solving social problems. Although reading comprehension was theorized and there is evidence of disputes over pedagogy of practice, reading comprehension measures were not yet designed. There was no conclusive evidence to suggest that boys were underachieving in reading comprehension. What was conclusive was that educators generated more questions than answers in dealing with ability and sex. Nevertheless, based on spelling assessments and word activities, the perception was spawned that girls were superior in language arts while the boys underachieve.

Although there was a plethora of studies investigating a variety of issues related to sex differences during the late 1800s and early 1900s, few comparative

studies between boys and girls in school were completed. Thorndike, a prominent and influential researcher, concluded that there was little evidence to support sex differences; therefore, few investigations comparing boys and girls were subsequently conducted, and as a result, few standardized tests with relative norms were developed (Lincoln, 1927). In the chapter that follows, you will note that Lincoln once again raises the topic of sex difference in school achievement in his study on *Sex differences in the growth of American school children*. The findings from Lincoln, as well as other evidence-based studies that investigated gender and reading will be examined, juxtaposed within the socio-cultural conditions that influenced educational research during the time period 1921–1950.

Chapter Five:

The Growing Status of the Field of Reading (1921–1945):

Laying New Ground

Introduction

The period between the end of World War I (WWI) and the end of World War II (WWII) was marked by turbulence. These three decades began with one of the greatest social, cultural, and economic periods of growth in modern history; continued into the most widespread economic depression ever recorded; and ended in major global bloodshed. "The decades from the outbreak of the First World War to the aftermath of the Second, was an Age of Catastrophe . . . For forty years it stumbled from one calamity to another" (Hobsbawn, 1994, pp. 6–7). However, in the midst of this turmoil, progress prevailed in the emerging field of reading pedagogy. Furthermore, significant advancements in reading research took place within an atmosphere of professional unanimity and with a "level of agreement by its leaders on key educational points" (Monaghan, 2007, p. 20). The aim of this chapter is to provide a perspective on research in reading education that took place from 1921–1950, particularly the major research findings on sex differences in reading achievement.

To illustrate the nature of the research that was being done, I have organized major studies from this 30-year period into five unifying ideas or themed topics that recurred throughout the multiple sources that were read.

1. First, I highlight the influence of early reading pioneers, and examine how their research reshaped the conception of reading as instruction

moved away from oral reading and memorization toward an integrated approach.

2. Next, I explore the enthusiasm that emerged for engaging, age-appropriate reading materials, including the use of basal readers and other quality children's books. Although reading pedagogy and the increasing availability of good literature attracted much attention, researchers were also concerned about the number of children who failed to learn to read.
3. A third focus I examine is the reading research that centred on the development of a medical framework for identifying and treating reading difficulties. When studying how children approached reading, investigations were conducted on how to successfully organize and implement reading programs to maximize student progress.
4. Subsequently, in the fourth section, I survey the studies that attempted to merge research and classroom practice for the purposes of improving instruction and student achievement. While focusing on the reasons for the number of students experiencing difficulty in reading acquisition, factors beyond the school realm were considered suggesting among other things that not all children were ready to learn to read at the same age.
5. The 1930s and 1940s was a time when scores of studies focused on reading readiness and the use of child-centred approaches designed to

match children's cognitive, physical, and emotional growth. These topics are explored in the fifth section.

Insights gleaned from these themes provide a context for the examination of studies on sex differences and reading comprehension. During the time period 1920–1950, eight evidence-based studies that met the criteria outlined in chapter three regarding sex differences in reading are described in detail and then analyzed.

Theme I: The Influence of Early Reading Pioneers and the Changing Conception of Reading

Although many researchers contributed to the growing field of reading education between 1920–1950, "[t]wo men in particular, both trained as educational psychologists, established themselves as important figures—Gates and Gray" (Monaghan, 2007, p. 17). William Gray (1885–1960) began his career as a teacher and school administrator. He was introduced to the scientific approach to education through his work at Columbia University under Edward Thorndike. Gray's interest in assessment began with his master's thesis in 1915, in which he developed an analytical scale, Oral Reading Paragraphs, which has undergone only minor changes over the last 50 years. Gray (1956) completed a study for the United Nations Educational, Scientific, and Cultural Organization (UNESCO). He was also one of the first researchers to study literacy internationally, and his recommendations are still taken under consideration today. Gray's main focus in reading research was how readers derive meaning. Stressing the importance of reading to the mastery of all subjects, he promoted the

viewpoint that reading should be taught in all grades. Gray advanced the idea that improvements in reading comprehension are critical for higher mental processes of all kinds and are central to learning and improving one's quality of life (Lauritzen, 2007; Luke, 1987).

> Regarded as the dominant figure in reading education, Gray can be credited as a founder of the field itself. He actively participated in and published major works about every aspect of reading instruction: testing, diagnoses, developmental programs K-college, remediation, reading and writing, adult literacy, instructional materials, teacher training, evaluation, theory, and research. (Gilstad, 1985, p. 510)

In all, Gray published over 500 titles on the topic of reading and was the founder and first president of the International Reading Association (Gilstad, 1985).

Another leader in reading research between 1920–1950 was Arthur Gates (1890–1972). While enrolled at Columbia University as a student, he was invited by Thorndike to join the teaching team at Columbia Teachers College. Gates worked with leading minds in education, including John Dewey, Robert Woodworth, James Cattell, and Charles Judd. Gates' research in reading led him to the realization that students differ in their interests and abilities, a key concept in the reading readiness movement. Understanding that children come to school with a variety of backgrounds, reading interests, and strengths, Gates proposed that classroom teachers need to use an assortment of methods to teach reading effectively (1928). He also made contributions in the area of remedial reading instruction. He developed one of the first individualized prescriptions for children

having difficulty with reading. Over a period of two decades of research, Gates outlined a system for the identification and remediation of reading difficulties. His research is highlighted in *Theme III: Diagnostic and Remedial Assessments to Support Children Who Find Reading Difficult*. "In many ways, Gates was ahead of his time because many of the topics he wrote about speak to reading educators today" (Sailors, 2007, p. 343).

Influenced by the research of reading investigators such as Gray and Gates, a shift in the concept of reading emerged during the time period 1920–1950. Previously, oral reading and memorization were the predominant focus of reading instruction, with a special emphasis on elocution and correct pronunciation. However, with the influence of scientific methods in educational psychology came a heightened interest in reading research, which in turn promoted an integrated approach for the teaching of reading. Thus, it is key to look at the influence of scientific developments on reading research.

An important example of the impact of scientific experimentation on reading is the systematic work on eye movements conducted by Huey (1908/1968) and the plethora of investigations that extended his research. Huey's landmark experiments led to the realization that the eyes do not continuously move across the page when reading, as previously thought, but rather dart quickly and then pause. Huey hypothesized that reading involves a complex interaction between the eyes and the brain, a somewhat controversial theory; in fact, the relationship between what the eyes capture and how the brain processes the information collected continues to be debated (Reed & Meyer, 2007). Charles

Judd (1873–1946), then director of the School of Education at the University of Chicago, extended Huey's research and initiated a change in emphasis from oral reading to silent reading in classrooms. By photographing readers' eyes during silent and oral reading and then measuring the number of pauses, Judd (1918a) discovered noticeable differences. In oral reading, the eyes make more pauses, resulting in slower and more laborious reading. In silent reading, there are fewer and shorter pauses, allowing a reader to read faster. Consequently, Judd initially claimed that silent reading should replace oral reading in classrooms. Compared to previous decades, there was a dramatic increase in the amount of material published, and consequently students needed to read more, read more rapidly, and read a greater range of topics (Judd, 1918a; Stone, 1922; Wheat, 1923).

Based on subsequent research into eye movements and reading, Judd (1918a) reconsidered his insistence on the exclusive use of silent reading in classrooms. Since the oral language skills of primary-aged children are more developed than their reading skills, Judd considered it natural to allow beginning readers to read aloud. He called for reforms in classroom reading instruction, balancing both oral and silent reading according to children's developmental needs. When the mechanics of reading were mastered (around Grade 4), students would benefit from reading a wide variety of materials with an emphasis on silent reading (Judd, 1922; Judd & Buswell, 1922).

Initially, during the early 1920s, there were conflicting viewpoints as to when formal phonics should be taught and whether or not they should be taught systematically. "Conflicting opinions regarding the place of phonetic training

have prevailed for many years . . . . [However, t]here is abundant evidence that phonetic training has some value" (Jenkins, as cited in Whipple, 1925, pp. 85–86). In an attempt to settle the phonetic controversy, a thorough report describing the most effective methods of instruction was published in the 24th yearbook of the National Society for the Study of Education (NSSE). "The investigation consists of an analysis of eighteen well-known widely-used reading manuals published since 1900" (Jenkins, as cited in Whipple, 1925, p. 88). The general conclusions of when and how phonics should be introduced were then outlined. For example, recommendations were made to start phonics instruction with the analysis of familiar words taken from sight vocabularies, and to introduce single consonants and long and short vowel sounds in the primer work with compound consonants and vowel combinations appearing in the first-reader work. It was cautioned, "Phonetics badly taught may do more harm than good" (Jenkins, as cited in Whipple, 1925, p. 89). It is interesting to note that the 1925 edition of the NSSE Yearbook assigned its entire volume to reading instruction,

> for reading is so obviously the key subject of the elementary grades and has, as this *Yearbook* rightly stresses, so many ramifications and contacts with the other school subjects and with important attitudes and habits of mental work that we cannot have too much light thrown upon its aims and methods. (Whipple, 1925, Editor's Preface)

Gates (1927a) examined the results of a study on phonetic and non-phonetic methods, and concluded that rather than accepting or rejecting the whole system of teaching phonics, a more intelligent solution would be to determine

which phonetic concepts were of value and then ascertain how and when to teach the skills. He suggested that there were advantages to using the intrinsic approach to reading, a method of integrating contextual clues with word analysis, even though "[n]o other single teaching device had been so widely adopted by American teachers as the phonetic training of beginning readers" (Gates, 1927a, p. 217). He argued that English phonetic translations were too varied and complex to be readily mastered by young children and that experimental techniques in phonetic methods were few and inconclusive. Gates (1927a) further claimed that phonetic training yielded slow and mechanical reading with an insufficient focus on thought. Considering the complexity of the reading process, Dolch (1931) hypothesized that the best pedagogical approach would be a balance of different methods, because all methods have advantages and disadvantages, and children do not all learn in the same way: "No one method alone is self-sufficient, so the problem is to make the most efficient use of them all" (p. 101).

Gray (1937) extended the concept of balance to include all aspects of reading. In his chapter "A Decade of Progress" (1925–1935), Gray cautioned educators against taking extreme positions in reading and proposed instead an integrated approach employing oral and silent reading, recreational and remedial activities, and basic processes. "One of the urgent needs today is a well-balanced reading program that gives adequate emphasis to the various types of reading [in] which children should engage" (p. 10). In order to integrate a balance of approaches in the classroom, Gray (1937) emphasized that teachers would need to be cognizant of the strengths of each method as well as when and how to use

specific strategies with specific children. A successful reading program would depend on having knowledgeable teachers who were able to apply a variety of teaching methods appropriate to the needs of the children in their classrooms (Cole, 1938; Gray, 1937). Experiments were conducted to integrate scientific studies with classroom instruction. These are described in more detail in *Theme IV: Teacher Knowledge and Supporting Supervision.*

The dawn of reading as a separate field in education was fostered by the work of many researchers, including William Gates and Arthur Gray. The method of uniting research techniques and psychological testing within an educational focus on improving reading acquisition contributed to a host of new topics and investigations. A shift in the conception of reading as oral performance to that of a complex interaction of processes fostered a strategy that emphasized a balance of instructional approaches. One topic that dominated reading research during the time period 1920–1950 was the use of engaging, age-appropriate reading materials geared toward the different interests of children.

Theme II: Improvement in Reading Materials and Investigating Reading Interests

The period 1920–1950 saw a substantial increase in emphasis on selecting age-appropriate reading materials. Judd (1922) made several comparisons about the need for school-based reading programs to keep up with changes in society. In 1840, for example, he commented that the average child went to school on an irregular basis and for a total of about 450 days; just enough time to learn how to read a few passages. By 1922, children attended school more regularly, for longer

periods, and spent more time reading. Consequently, to keep up with changes in attendance, schools needed to purchase a greater variety of quality books appropriate for children.

The emphasis on providing quality children's books was also apparent in Gray's *Summary of Reading Investigations* (1929). He recommended that books in school be interesting, be carefully matched to grade levels, and have educational value. Gray co-authored (with May Arbuthnot) the *Elson Basic Readers* series (1930s) (the popular Dick and Jane readers), which were based on an underlying philosophy of intrinsic phonics and offered stories that would interest children. The early basal readers replaced rote phonics drills and were designed using a system of controlled vocabulary, in which familiar words were repeated frequently and new words were introduced gradually. The text included many rhymes as well as repeated words and phrases, with each successive story gradually increasing in difficulty. Narratives involved everyday adventures familiar to most children of white-collar families. They were situated predominately in suburbia, with colourful pictures that depicted traditional family values based around a stay-at-home mother, and a father who went to work in a business suit (Luke, 1987). The reading series, at the time, focused on a narrow White middle-class view of life. Revisions to the *Elson Basic Readers* continued up until the 1950s and 1960s, with the eventual introduction of minority families and inner-city settings (Luke, 1987). It is estimated that 85% of schools in the United States used the Elson and Gray readers in their classrooms (Lauritzen,

2007, p. 313). Although the basal readers were criticized as "blatantly classist, sexist, and racist" (Luke, 1987, p. 92)

> … [i]t would be erroneous, however, to assign some sort of conspiratorial intent to the makers of these texts.—they were riding a wave of post-war optimism based on the belief in "social progress" through scientific and industrial expansion. Gray and Arbuthnot—in their commitment to make reading failure virtually impossible—were unabashedly idealistic in their intents. (Luke, 1987, p. 110)

Highroads to Reading, a series similar to the *Elson Basic Readers*, but a Canadian version, was authorized for use in public schools in Western Canada (Roy, Roy, Sheffield, & Bollert, 1932). Similar to Dick and Jane, the *Highroads to Reading* series depicted two school-aged children, Jerry and Jane, with family values similar to those in the American series. However, the major difference was the stories displayed a northern setting where Jerry and Jane are seen tobogganing and playing in the snow. What influence did the basal readers have on young Canadian readers?

> From 1920 on, anthologists continued to help redesign the received shape of the national culture. . . . During the middle decades of the 20th century, several anthologies came to be influential textbooks. . . . such as the *Highroads to Reading Series*. (New, 2002, p. 33)

These anthologies appeared to be powerful in portraying and promoting a narrow image of Canadian society—urban, White, and middle class, and not reflective of the growing majority of Canadians.

David Russell, a Canadian reading researcher, was a senior author on the *Ginn Basic Readers*, written between 1948–1960. Russell worked as a research assistant for Gates after completing his PhD in Educational Psychology in 1937 at Columbia University. Influenced by Gates, Russell developed what were termed reading readiness books. Reading readiness described the developmental stage at which a child is ready to read, which varies from child to child. Reading readiness has been replaced with the more appropriate term, emergent reading. Early experiences in pre-reading activities are critical to success when learning to read. Using common children's experiences, Russell developed quality, realistic stories (Barry, 2007). Most reading experts in the 1940s promoted the use of the basal reader: "Reading professionals were united in designating good basal reading series as [a] key to orderly, cumulative reading instruction" (Monaghan, 2007, p. 23). Basal reading series allowed teachers to read with small groups at different levels, which was critical since research on individual differences revealed that a range of reading levels existed within classrooms (Monaghan, 2007).

In addition to the use of basal readers, there was a growing interest in school libraries and the publishing of quality children's books. May Arbuthnot, Gray's co-author on the *Basic Readers* series, was well known for her reviews of children's literature. "*Children and Books* was a ground-breaking historical survey of children's literature that over the next five decades became the standard college level text in schools of education" (Monaghan, Israel, & Dahl, 2007, p. 211). Arbuthnot attempted to promote the love of literature and reading among

children and called upon parents, teachers, and librarians to work collaboratively toward this goal.

During this period (1921–1950), the reading preferences of children became topics of study for librarians and reading researchers. After surveying students, Danylu Belser (1926) and David Russell (1941) independently came to the conclusion that sex differences influence children's preferred choice of reading materials by adolescence. Belser visited six libraries in Montgomery, Alabama, and surveyed the reading interests of 112 boys and 125 girls enrolled in Grades 4 and 5. Results indicated that up to ages 8 and 9, boys and girls did not differ in their preferences of reading material. The greatest difference in reading interests occurred between the ages of 10 and 13, with boys favouring books about war, scouting, sports, strenuous adventure, and physical bravery. Girls preferred love stories and tended to read adult fiction earlier than boys. Boys were more independent in their choices of books and gave more reasons for choosing books. Girls, on the other hand, were more open to reading a variety of genres. Belser (1926) stressed the value of illustrations in "awakening and developing a love for reading" among both boys and girls (p. 296).

In 1941, David Russell surveyed 300 boys and girls in Grades 9 and 10, enrolled at two high schools in a Western Canadian city. He compared his results to eight similar studies completed in the United States to create a cross-national sample. Russell noted marked differences in the reading interests of boys and girls. Younger children would read anything that was available, whereas early adolescents developed specific preferences in reading materials. Adolescent boys

preferred science, sports, and news events and adolescent girls favoured love and romance, fiction, and women's magazines. Both boys and girls read a variety of magazines: 87% indicated that they regularly read one or more magazines as well as the newspaper. When Russell compared the findings from his survey to American studies on a similar topic and with a similar age group, he noted parallel results.

How better to instil a love of reading than to target the interests of students of all ages? In Patterson's 1930 summary of the general aims of the reading teacher, the first goal was to provide vicarious experiences that were rich and varied: "The selection of material consonant with this aim should be a prime pursuit of classroom effort in all grades from the beginning through the high school. None other is more enduringly important" (p. 15). In educators' quests to make reading interesting and available to all ages, quality books addressing children's reading interests became the major focus of conversation and research. However, the adoption of engaging reading material in classrooms did not ensure success in acquiring skills in reading for some children. Alarming were the large numbers of children in every grade who struggled to learn to read. Using a medical framework, researchers began to develop assessments that identified signs and symptoms of reading difficulties. They also experimented with systematic methods for the treatment of reading deficiencies.

Theme III: Designing Diagnostic and Remedial Assessments to Support Children Who Find Reading Difficult

Lack of progress in reading was a major cause for retaining students in school, particularly in the primary grades. Walter Percival (1926) conducted a study on the causes and subjects of failure in schools from Grades 1 to 8, sampling schools of varying sizes from random cities and counties across California. The results revealed that the ratio of students held back due to deficiencies in reading was 99% in Grade 1, 90% in Grade 2, and approximately 70% in Grade 3, with a continual drop to about 25% by Grade 8 (pp. 20–21). Data were collected from 9,217 responses: 7,476 from cities and 1,741 from counties (Percival, 1926, p. 2). Despite advances in teaching methods, the improvement in reading materials, and the increased time spent on reading instruction in class, "pupils of good, or even of superior intelligence, as well as the duller ones [a term used at the time to describe children with cognitive delays], do not always attain satisfactory skill despite every advantage and incentive" (Gates, 1927b, p. 5).

After many years of research, Arthur Gates (1927b, 1935, 1947) outlined a systematic technique of measuring reading achievement and a plan to treat reading difficulties. He updated his diagnostic methods through 3 editions and 20 years of continued research. Using data from over 13,000 pupils and 46 self-designed subtests, Gates devised a series of group tests that teachers could use to "measure and to diagnose ability at the same time" (1927b, p. 9). The tests were used to assess the reading abilities of students, both those who were struggling

and those who read well. Using the analogy of a malfunctioning automobile, Gates hypothesized that if the teacher could break down the components of reading and identify those causing the problem, a targeted, effective remediation program addressing the reading deficit could be implemented. Adopting the concept of average as a norm representing "mediocre" ability, Gates devised tables for each grade and component of reading. These components included word recognition, phrase and sentence reading, and paragraph comprehension. In conjunction with each other, the three tests were designed to diagnose reading difficulties. Examining the discrepancies in the results, Gates (1927b) described four types of reading profiles and detailed a specific course of remediation for each type (pp. 54–66):

1. A low score in paragraph comprehension with an average or higher score in phrase and sentence reading and word recognition indicated a deficiency in precise comprehension of thought. Children with this profile needed practice in reading for meaning with a gradual increase in the size and complexity of paragraphs.
2. A low score in paragraph comprehension and phrase reading but an average or high score in word reading indicated that the pupil had a fair knowledge of single words but was not able to combine the words into larger thought units.
3. A deficiency in all three tests indicated that the student needed instruction in all components of reading.

4. The last profile identified was a relatively low score in word reading but a higher score in phrase reading and paragraph comprehension. This profile was not frequently found because most children cannot read phrases and sentences if they cannot recognize individual words. However, Gates noted that this combination sometimes appeared, and that students with such a profile were able to draw meaning by relying on a few words in a sentence. They often omitted, transposed, or mispronounced words when reading orally but used contextual clues to derive meaning. Gates stated that these students required more focused instruction on literal reading with caution against becoming too "word conscious." (p. 65)

Gates proposed that children's reading scores on average should match their intelligence scores:

> Although we do not know exactly how much higher or lower the reading score should be, the best approximation is that the reading score should be about as much above or below the average of the pupil's intelligence is above or below the average for his grade or age. It is expected, in other words, that the "reading grade" should about equal the "mental grade" and that the "reading age" should about equal the "mental age." (1927b, p. 57)

Embedded in Gates' research was the assumption that the IQ tests used were reliable and valid measures of student abilities. The concept of using a reading–IQ comparison was revolutionary, as was the theory that a discrepancy between the scores indicated a learning difficulty. Although controversial today as

researchers have since demonstrated that many poor readers do not exhibit a reading–IQ discrepancy, this approach was widely accepted at the time when diagnosing learning disabilities (Siegel, 2003) and it remains in use.

Combining the increasing use of intelligence and achievement testing with the study of the relationship between capacity and intelligence, there was a growing interest in children who, although they had average intelligence, failed to learn to read. "Until recent years teachers have assumed that any child who attended school regularly could learn to read, and that if he did not he must be either lazy or stupid" (Monroe, 1932, p. 1). Marion Monroe completed an in-depth study of reading disabilities in an effort to diagnose causative factors related to reading difficulties. Using diagnostic tests to inform instruction, Monroe completed a follow-up study assessing the effectiveness of the selection of remedial methods used to overcome reading difficulties. She worked in both regular public school classrooms and at the laboratory of the Institute for Juvenile Research in Chicago for over an 8-month period, and instructed teachers on the use of remedial techniques for students with reading disabilities. The results were compared to classrooms with students in public school, presumably in Chicago, who had reading disabilities and whose teachers had not received instruction in remedial methods. To determine the extent of each child's reading disability, a battery of tests was administered, including achievement in reading as well as social, physical, and psychological tests. The study used three groups: 1) regular public classrooms with teacher support in remedial methods, 2) students at the Institute for Juvenile Research with teacher support in remedial methods, and 3)

regular public classrooms with no teacher support in remedial methods. All groups had very similar initial reading achievements. Students ranged in age from 6 to 15 years. Students who received remedial instruction from trained teachers in both clinical and school settings improved their reading scores on average by 0.8 to 1.5 grade levels while students in classrooms with no remedial instruction made a gain of 0.14 grade levels on average (Monroe, 1932, p. 146). Monroe concluded "that children who have difficulty in learning to read do not usually overcome the difficulty under ordinary school instruction but are able to make normal and accelerated progress under special methods adapted to their difficulties" (p. 143). In addition, Monroe observed that remedial instruction in reading not only improved scholastic success but also led to what she saw as improvements in the students' personality and behaviour (p. 147). Connecting poor performance in school, and in particular in reading, to behavioural and emotional problems was not common in the reading literature of the time. Recognizing the uniqueness of Monroe's research, Ernest W. Burgess, Director of the Behavior Research Fund, noted in the foreword of *Children Who Cannot Read: The Analysis of Reading Disabilities and the Use of Diagnostic Tests in the Instruction of Retarded Readers*,

> This study [referring to Monroe, 1932], therefore, embodies in unusual degree the combination of behavior research of high scientific standards with successful therapy. . . . It is not often that one may present a product of scientific research, which is also of immediate help in removing a

handicap common to thousands of school children. (as cited in Monroe, 1932, Editor's Preface)

Marion Monroe furthered her research as chief psychologist at the Child Guidance Center in Pittsburgh (1932–1936), where she developed the Monroe Reading Aptitude Tests for the Houghton Mifflin Company in 1935. She was a member of the team of writers, along with William Gray, who wrote the *New Basic Reading Program* (Dick and Jane series) in 1937.

A change in school promotion policies, as the result of studies that indicated children develop best when they are with children of the same age, size, and stage of social and emotional development, discouraged the retention of children. As a result, a new issue emerged: how to cope with 30 or more students in a classroom with a vast range of reading levels. Prior to the policy change, students who could not read were retained until they were successful or dropped out of school (Cole, 1938). As Cole aptly stated, with "this modern emphasis upon the child as an individual rather than upon the curriculum" (p. 20), teachers in regular classrooms can "expect to find a range of ability equal to at least three grades, and probably more" (p. 26). As an aside, it is interesting to note the influence of John Dewey's progressive philosophy in this statement made by Cole. To cope with such a range of reading levels, remedial reading classes became common in many school districts, particularly in junior and senior high schools (Cole, 1938). Luella Cole (1938), researched methods for teaching remedial reading, and emphasized that since reading is a complex process, it requires a knowledgeable classroom teacher who understands the nature of

reading and is able to administer a repertoire of developmentally appropriate techniques. Successful programs included reading material that matched the varied reading levels of students. Thus, although remedial teachers were already versed in specialized techniques for reading difficulties, the need for regular classroom teachers to have such training began to be explored (Cole, 1938).

Although Luella Cole continued to research on her own, between the 1920s and 1930s she worked with her husband, Sidney Pressey at Ohio State University (OSU). Their research was an original blend of Freud's psychoanalysis and Thorndike's behaviorism (Petrina, 2004). Summarizing the uniqueness of their approach, Petrina noted in his abstract how these two psychologists were able to merge education and psychology

> to individualize students within the confines of mass education . . . Exemplifying the hybrid nature of educational psychology, these two psychologists' practices at OSU (Ohio State University) were nonetheless common and indicative of relations between education and psychology at the time. If psychoanalysis was primarily oriented toward personality and counseling, and behaviorism toward performance and testing, Cole and Pressey thoroughly mixed the two in practice. They recognized that personality was inseparable from mundane educational practices—personality was normalized in the minutiae of everyday schooling. (2004, p. 525)

Gates (1947) concluded that in order to diagnose and remediate reading difficulties, classroom teachers needed to be able to determine the strengths and

limitations of students, understand the reading process, and become well versed in methods of instruction. Subsequently, attempts at marrying research and classroom practices in order to foster better reading instruction appeared.

Theme IV: Bridging Research With Classroom Practice

Work in the area of integrating scientific methods in reading research with classroom instruction was noted as early as 1918 when Charles Judd made the following statement: "A scientific study of reading should point out the way in which the experiences of the school and the investigation of the educational laboratory may be combined to supply certain principles of procedure which will surely improve instruction" (1918b, p. 5).

An attempt to integrate scientific research with teaching in classrooms was undertaken in a study completed by Gray between 1925 and 1930. Five elementary schools participated. They differed widely in size and location, and were selected from rural areas and small cities in Northern Illinois. The study was divided into three parts:

1. Initially, a survey was used to determine the kinds of reading activities employed in the schools, the variety of reading resources available, and the levels of reading achievement among the students. The data provided a starting point from which to evaluate the reading practices of each school in relation to then-current reading research.
2. In the second part, classroom teachers were taught new reading methods, given demonstrations, and invited to conferences with supervisors.

3. The third part examined whether an increase in student reading achievement and a change in teacher practices resulted.

Due to changes in school administration and staffing, data were collected from only three of the five schools. Of those, two of the three schools made distinct gains while the third did not. According to Gray (1933), the failure of the third school was due to ineffective supervision: "The findings show clearly the value of well-planned supervisory activities directed by the supervisory officer of each school. Increasingly, principals must provide vigorous leadership in reorganizing and improving reading" (p. 207). He outlined the need for common goals, a cooperative spirit, effort, and desire to learn on the part of the teacher; a strong leadership that believes in the need to change; ongoing professional development; continuous review of data; support of teachers experiencing difficulties with students who are not progressing; provision of quality reading materials; and cooperation with outside agencies such as libraries. "Experience teaches that numerous difficulties are encountered in any campaign to improve teaching. These vary in type from inadequate facilities to inertia among staff members or open hostility to educational reforms by members of the community" (Gray, 1933, p. 213). He noted that there had been 28 studies between 1916 to 1932, with samples drawn from city, rural, and district schools, that focused on improving classroom teaching of reading (Gray, 1933, p. 9). The results of these studies indicated that there was convincing evidence that improvement in classroom reading occurs when:

> …stimulating leadership is provided. They also show clearly the value of a carefully planned program, including such steps as defining specifically the problems to be studied; and providing adequate help in such matters as measuring progress, diagnosing difficulties and needs, and providing remedial training. (Gray, 1933, p. 26)

A summary of 83 scientific studies conducted between 1931–1934 is provided in Gray's *Review of Educational Research* (1935). Methods to improve reading instruction in the classroom were again addressed. Suggestions included the need for a designated reading coach to help adapt instruction and provide support and guidance to students, the reading of current bulletins and research, and adequate supervision. The mounting evidence suggested that knowledgeable teachers and strong leadership were essential in meeting the diverse reading needs of the classrooms. "Gray, like many Progressive-era educators, would have teachers be reading 'experts,' knowledgeable in the ways of instructional psychology" (Luke, 1987, p. 109). The need for continued professional learning to improve reading instruction was apparent because educational policies were moving toward meeting the needs of individual differences within the classroom. The number of students failing due to reading deficiencies led researchers to a better understanding of the variance within same-aged children in physical, emotional, social, and cognitive development. Sex differences were examined, and of specific interest are the studies related to gender and reading discussed in the next section.

Theme V: Individual Differences in Reading Readiness and Sex Differences

The concept of reading readiness found its historical roots in the works of Rousseau, Pestalozzi, Froebel, and Dewey. As discussed in Chapter Four, these philosophers and educators were instrumental in the early reform movement in education and promoted the philosophical stance that education should be harmonious with the process of growth and development through life (Smith, 1950). Given the appalling number of children failing the first grade, investigations were conducted on the stages of early development and identified the need for a differentiated curriculum that matched children's cognitive, physical, and emotional growth (Durrell, 1940; Smith, 1950; Traxler & Townsend, 1946). The number of articles published on the topic of reading readiness was highest between 1937 and 1940. A decline occurred during WWII (1939–1945) and a gradual increase appeared after the war years (Smith, 1950). Previous studies claimed that on average a child learned to read at age 6 years, 6 months (Gray, 1935, p. 55). However, continued research revealed that the age at which a child is ready to read varies greatly. In addition, studies on reading readiness and achievement indicated that there were sex differences.

> Differences in physical, intellectual and emotional maturation is most frequently given as the reason why, on the whole, girls seem to be ready for successive levels of development in reading sooner than boys. Other reasons given are that the interest and disposition of girls make them more favourably inclined to learning, that pupils in the elementary schools have women teachers almost exclusively, that girls are promoted on lower

> standards than boys, and that there is a need for more reading material which appeals especially to boys. (Smith, 1950, p. 10)

Although it was stressed that teachers could not generalize because of the overlap of abilities between boys and girls, they could expect that boys would need more attention in early reading as they generally develop later than girls (Smith, 1950). Analyses of reading readiness studies as well as other investigations on sex differences in reading achievement conducted during the time period 1921-1945 are detailed next beginning with Lincoln's extensive examination of school-aged children.

First comprehensive study on sex differences in schools.

As part of his doctoral dissertation in educational psychology at Harvard University, Edward Lincoln (1927) wrote the book, *Sex Differences in the Growth of American School Children*. It was one of the first comprehensive studies on sex differences and school performance. Previous comparative investigations either dealt with only one or two traits, or studied sex differences in conjunction with other issues. Lincoln was in favour of equality of education and a promoter of coeducation (Blakemore, Berenbaum, & Liben, 2009). Although rigorous, "Lincoln's work on sex differences does not appear to have made much impact on the field, although virtually all of his conclusions would still be considered reasonable in light of the data that has been collected in the 75 years since the book was written. Perhaps the time just was not right" (Blakemore et al., 2009, p. 24).

Lincoln (1927) examined both the physical and mental growth and development of students. In terms of physical growth, he determined that boys overall weighed more and were taller than girls, except during puberty. On average, girls matured earlier than boys by 12 to 18 months (p. 27). Using the *Stanford-Binet Intelligence Test*, (reference not included) Lincoln (1927) deduced that there were no differences in cognitive abilities between boys and girls. The difference in general intelligence between boys and girls was less than the range of differences within each sex (p. 164). Batteries of achievement tests were administered, including arithmetic (computation and reasoning), reading (oral, silent, and speed tests), handwriting, spelling, history, language tests (understanding sentences), and composition. The overall conclusions drawn were that girls were generally superior in school achievement with the exception of math reasoning and history (Lincoln, 1927, p. 104). Results from the reading comprehension tests were inconclusive (Lincoln, 1927, p. 72). The following is a précis of Lincoln's analyses of the four studies that specifically pertain to sex differences in reading.

St. Louis survey.

A survey of the St. Louis Public Schools completed in 1917, using the *Gray Oral Reading Test*, (reference not included) compared the reading achievement of 5,118 students in oral reading (approximately 300 boys and 300 girls from each grade from Grades 1 to Grade 8). Girls were found to be one-fourth to one-half of a grade ahead of the boys in their ability to read orally. Unfortunately, it was not clear if the superiority in oral reading translated into an

increased comprehension and reading achievement. The results on the Silent Reading Test with 4,463 students in Grades 2 to 8 (1,804 boys and 2,659 girls) in rate and quality of silent reading revealed that girls did better on the speed test but that boys were "superior" in the quality of responses (Lincoln, 1927). Students were assessed by grade (rather than by age). "In both oral and silent reading the differences are small, and it is hard to estimate their significance, as no reliability figures are shown" (Lincoln, 1927, p. 68).

O'Brien experiment.

In 1921, a second study was conducted on the average reading rates of boys and girls, and was known as the "O'Brien Experiment." A small number of students (166 boys and 201 girls) from Grades 3 to 8 were assessed. Unfortunately neither the reading test nor the location of the schools was identified. Although boys showed gains in reading rate from the beginning to the end of the study, the girls' rate of reading was higher in every grade except Grade 5 and 8.

Black Hawk, Iowa study.

Using the *Monroe Silent Reading Test* (reference not included) with students from 20 rural schools in Iowa from Grades 3 to 8, girls fared better on both rate and comprehension than boys. Although the differences were reported as extremely small, and the author did not report the number of students tested.

Missouri study.

One of the first studies to use age scores in reading took place in Missouri. The *Holmes Reading Test* (reference not included) was administered to

approximately 270 boys and 350 girls ages 7 to 15 enrolled in Missouri training schools (incidentally the term training schools is not described). Results from the quality scores (inferring that this means reading comprehension) showed that girls were superior to boys at every age on the percentage of mean, ranging from 58.7% at age 8 to a 1.4% difference at age 12. In speed of reading, girls also outperformed boys on percent of mean scores, with an average difference ranging from 5.4% at age 11 to 35.4% at age 15. There is no indication if the results are significant. Results from the Missouri study contradict the St. Louis Survey. Since each investigation used different reading tests, contrasting methods of reporting—grade versus age scores, and varying numbers of participants, making a comparison and drawing any definitive conclusion is impossible. Therefore, Lincoln determined that the evidence from all studies combined indicated that girls read better orally and at a faster rate; however, due to conflicting evidence, a clear case could not be made as to which sex was superior to the other in reading comprehension (1927, p. 72). For more reliable results, since the age distribution at every grade varied greatly due to retention and late entry to school, Lincoln emphasized the need for researchers to use age rather than grade comparisons for reading comprehension tests. As researchers designed intelligence and achievement tests, questions were raised about the validity and reliability of the tests. The issue of test design was noted as a possible explanation for sex differences in the studies conducted by Commins (1928) described next.

Examining the design of tests: Intelligence and achievement.

Commins (1928) completed two separate studies on sex differences, one on intelligence and another on achievement. The studies took place "in the schools of a small city with a total school enrolment of about three thousand" pupils (p. 599). No locations are provided; presumably the studies took place in the United States.

Comparisons were made between two intelligence tests, the *National Intelligence Test* (NIT) and the *McCall Multi-Mental Scale*. Detailed descriptions of these intelligence tests were not provided. Since the focus of this investigation was sex differences in reading ability, results of the intellectual comparisons will be briefly summarized. The NIT was given to 200 Grade 5 students. Results showed that the median scores between boys and girls were virtually identical. *The McCall Multi-Mental Scale* was administered to 1,130 students between 9 and 14 years of age in the same school system, with approximately equal numbers of boys (560) and girls (570). On *The McCall Multi-Mental Scale*, the girls' median score was higher than the boys' score by 4.7. However, it was not reported whether the difference was statistically significant. Two reasons offered at the time for the difference in results were test-design bias or a difference in maturation rates. Citing the research from Lincoln (1927) as support for maturation differences, Commins (1928) deduced that girls were brighter than boys "at least up to fourteen or fifteen, because of the advanced maturity of the girls as indicated by most anatomical and physiological tests" (p. 599). Though, after close examination of Lincoln's (1927) study as discussed previously,

Lincoln concluded that there were no differences in cognitive abilities between boys and girls (p. 164). Commins' interpretations of Lincoln's results are different from those I analysed from the original source (Lincoln, 1927). In addition, it is difficult to draw comparisons between the *National Intelligence Test* (NIT) and the *McCall Multi-Mental Scale*. Two different intelligence tests (NIT and *McCall Multi-Mental*) were administered to two different groups (NIT tested only students in Grade 5, while the *McCall Multi-Mental* tested students between the ages of 9 and 14). Both studies did not use the same age group; therefore this is not a valid comparison. This leads me to presume that the design of the test was the contributing factor in the differences noted between the two IQ tests.

The second study on school achievement was reported using the *Stanford Achievement Test* (SAT) results from 175 Grade 5 students: 90 girls and 85 boys. Results showed sex differences between the subtests. Boys scored higher than girls on arithmetic (6 points), nature study (5.7 points), and history and literature (3.7 points). Girls surpassed boys on reading (5.2 points), language (4.2 points), and dictation (10.9 points). Despite the fact that girls surpassed boys in reading and language, boys achieved higher scores in nature study and history and literature, tests that also involve reading comprehension. It is thus inappropriate to conclude that these results are indicative of boys' underachievement in reading comprehension. Rather they are attributable to the design of the reading tests. If boys can achieve in history and literature tests, which involve reading, why did boys underachieve in reading and language tests? A description of the SAT used was not provided. Considering the date of the study (1928), presumably the first

published version (1923) was used. A review of the reliability and validity of the first edition of the SAT indicates that although the psychometric quality was relatively high, the test relied to a great extent on norm-referenced interpretations. There was criticism that the test produced invalid results due to the gap between its norms and what was being taught at school (Haertel & Calfee, 1983, p. 120). This criticism leads me to challenge the trustworthiness of the results of the SAT, as we are uncertain what the test was measuring.

Rate of reading examined.

The purpose of Arthur Traxler's study (1935) was to determine if separate norms for boys and girls were needed when measuring rates of reading. The *Iowa Silent Reading Test* Form A (Traxler, 1935, p. 351) was administered to seniors at the University of Chicago High School in October 1933. Form B of the same test was administered to classes from sub-freshmen to senior levels in January 1935 (references were not included). A total of 6 classes of high school–aged students, 256 boys and 283 girls, were assessed. Mean scores favoured boys in several areas: sub-freshmen (difference of 1.9), freshmen (0.6), seniors from 1933 (2.1), and seniors from 1935 (0.3). No difference was noted in the junior year, and girls' mean scores exceeded boys' in the sophomore year by 2.1. Overall, Traxler concluded that boys and girls as separate groups at the high school level did not differ in terms of reading rate.

A previous study at the college level conducted by Berman and Bird (1933) with 790 sophomore psychology students from the University of Minnesota, 463 women and 327 men, indicated that on average women read

approximately 20 words per minute faster than men. They noted, "Women are significantly superior to men in speed of reading but not in college ability scores" (p. 225). It is fair to conclude that speed of reading does not necessarily imply improved achievement. Even today, discussions of fluency and its effects on comprehension are debated because improvement in rate of reading does not necessarily imply an improvement in reading comprehension. Although many studies on young children conducted during the time period 1920–1950 focused on the underachievement of boys, particularly in language and reading, studies of the high school grades concentrated on whether or not girls were as capable as boys.

Sex differences and high school examinations.

In the winter of 1935, more than 19,000 senior high school students (approximately 8,000 boys and 11,000 girls) were assessed using the *North Carolina High School Senior Examination*, a test constructed and administered by a committee representing the North Carolina College Conference (reference was not included). Students were from mostly English-speaking homes and small rural towns and cities (50,000 or less) in North Carolina, United States. The test included 25 multiple-answer-type questions for each of 7 subtests, including: 1) literature, 2) reading I, 3) reading II, 4) English usage, 5) mathematics, 6) general science, and 7) American history. "The committee attempted to select those items which represented typical experiences of the high school students" (Jordan, 1937, p. 255). The literature subtest consisted of questions where students had to identify authors, characters, books, and quotations. The two reading subtests

were based on questions from selected paragraphs. The English usage subtest consisted of sentences that needed to be analyzed for correctness and syntax. Mathematics questions were selected from algebra and geometry concepts. Questions in science and American history were related to what was taught. Students were given an hour and a half to complete the 175 multiple-choice questions (25 questions in each of the 7 subtests). Results indicated that on average, there were no differences in achievement between boys and girls on the literature and reading subtests. Although girls surpassed boys in English usage, boys surpassed girls on the tests of mathematics, general science, and history. Jordan (1937) concluded that although boys' scores were slightly ahead as a whole (average for boys was 82.9 and average for girls was 80.1), "girls were as able to perform the intellectual problems incident to high school work" (p. 261). Although the differences between boys and girls were negligible, it is interesting that Jordan (1937) hypothesized that the differences between boys and girls were not simply biological features, but proposed that girls did not perform as well as boys as a result of environmental factors that could be resolved by "adjusting teaching procedures to individual needs" (p. 261). This hypothesis suggests that the culture of the region of North Carolina weighed against girls and that further studies would needed to verify whether girls were indeed as capable as boys and could reach similar achievement at the high school level.

School achievement differences between elementary and high school students.

Investigations thus far have examined sex differences in either an elementary or a high school setting. The following study explored achievement differences between elementary and high school students. Stroud and Lindquist (1942) examined two separate studies on sex differences in school achievement using data from, respectively, the Iowa Every-Pupil Testing Program for high school and the Iowa Every-Pupil Basic Skills Testing Program for Grades 3 to 8. Results from both studies were reported and comparisons were made.

The College of Education and the Extension Division of the State University of Iowa designed the Iowa Every-Pupil Testing Program. Each year, on average about 300 high schools and approximately 50,000 students participated in this state-wide program. The tests were administered in all schools in the third week of May. Stroud and Lindquist's (1942) study used the assessments from 26 of these schools over the period from 1932–1939 and gathered data from 12 different subjects, including algebra, geometry, general science, biology, physics, world history, American government, American history, contemporary affairs, economics, reading comprehension, and Latin. Random samplings of students were selected from each subject. Information was provided on the number of boys and girls for each test, the mean scores for the two sexes, and the differences between the mean. Boys outperformed girls on all subtests except for reading comprehension and algebra. On the reading comprehension tests, the difference in the mean scores between boys and girls

was very small. On Part I of the reading comprehension test, the mean score for boys was 79.87 and for girls it was 79.71. On Part II of the reading comprehension test, the average score was 22.96 for boys and 23.27 for girls.

To compare results from high school with those from elementary school, data were collected from the Iowa Every-Pupil Basic Skills Testing Program from Grades 3 to 8 and then analyzed. Two batteries were administered: an advanced battery for Grades 6 to 8, and an elementary battery for Grades 3 to 5. Each battery consisted of four subtests: 1) Silent reading comprehension in two parts: reading comprehension and vocabulary; 2) work-study skills; 3) basic language skills; and 4) basic arithmetic skills. Students were randomly chosen (every tenth pupil from alphabetical lists). Differences in the mean scores of each test were then ranked. In all subjects, with the exception of arithmetic, girls outperformed boys. Small differences were noted in all the subtests with the exception of basic language skills where the girls showed superiority over the boys however, no significant differences were reported.

Stroud and Lindquist (1942) concluded that although the results indicated that girls generally did better in elementary grades, boys caught up and surpassed them in high school. They hypothesized that the reason boys underachieved in the elementary grades and then outperformed girls in high school was due to the change in subject matter with an increase in history, geography and science from elementary school to high school. Boys outperformed girls in high school because they were more interested in the subjects offered in high school, such as science and social studies, although Stroud and Lindquist offer no evidence to

support these conclusions. That boys appeared to underachieve in the elementary grades and that this was due to differences in maturation rates was a topic mentioned. The next section examines the relationship between the nature and character of preadolescent growth in reading achievement.

Sex differences related to maturation and achievement.

The 1930s brought about increasing emphasis by educators on the whole child. Reading, for example, was viewed as a way to enhance children's overall development rather than as an activity on which to assess their achievement for the purpose of passing or failing them. Cecil Millard at Michigan State College noted this shift in the pedagogy of teaching methods and seized the opportunity to study its underlying principles in *The Nature and Character of Pre-Adolescent Growth in Reading Achievement* (1940). Millard was critical of group tests. Specifically, he was critical of the use of reading norms and standards to compare similarly aged children, because these systems failed to take into account individual differences in growth and reading. Millard (1940) emphasized the need for longitudinal studies that successively examined children's reading growth (p. 72).

Millard (1940) devised a mathematical equation for determining reading growth by comparing time and maturity. For example, it takes a cognitively delayed child a longer time than an average child to reach a mental development of 10 years (1940, p. 74). Growth equations for reading achievement were based on the time it took to reach a given growth level (grade level of achievement). An equation for the growth of each individual was derived from the actual

measurements taken from the *Stanford Reading Achievement Test*. Millard does not indicate which version of the *Stanford Reading Achievement Test* was used. Predicted reading scores were compared to actual Stanford Reading Achievement measures assessed approximately every 6 months. This work was completed over a 6-year period; using only data collected from students Millard observed, for more than 3 consecutive years preceding adolescence (ages 7.6 to 11). Using the *Stanford Reading Achievement Test* in combination with "growth equations," Millard collected 576 reading scores from 55 boys and 62 girls attending Henry School in Dearborn, Michigan. These students were predominately from middle-class homes in an industrial community.

The purpose of Millard's (1940) study was threefold: 1) to examine if there were individual differences in the development of reading achievement; 2) to plot the overall pattern of preadolescent reading growth; and 3) to measure the effect of sex and IQ on reading development (p. 72). He discovered that it was possible to describe a child's growth in reading in mathematical terms. Reading achievement did not develop in an erratic or variable way as he anticipated but rather individual performance followed a regular pattern of development. Since not all children develop at the same rate, Millard determined that individual rates of learning were constant. By calculating developmental rates, he was able to estimate expected growth in reading. The distribution of deviations of predicted reading scores from the actual Stanford Reading Achievement measures were plotted, and the average deviation for the whole group was found to be less than ±

2.9 (p. 80). He concluded, "It seems remarkable that individual performances follow such a regular pattern of development" (p. 79).

Millard (1940) then plotted (using "Curves of Constants"), and compared, the reading performances of boys and girls. Two major findings were noted. First, the growth in reading had a curvilinear characteristic that was not comparable to the straight-line norms presented in the Stanford reading achievement tests. Based on these findings, Millard concluded that the norms for the SAT test required revision. Secondly, girls had different growth patterns than boys, with girls initially showing greater rates of growth. Millard indicated that although many studies had investigated reading performance, few had explored sex differences as a significant factor. This shift led him to further investigate sex differences in reading performance.

Millard's (1940) main objective in this last section of the study was to assess the differences between the reading scores of boys and girls at the preadolescent level and to examine the relationship between IQ and reading achievement. Comparisons were made between two groups: 1) the first group consisted of unpaired students, including 55 boys and 63 girls; 2) the second group of children was paired (43 pairs) according to their IQ, which ranged from approximately 85 to 125, with a mean of 106.1. (Although Millard does not specify which IQ test he used, he does indicate using Stanford norms. Considering the time this study was completed, at the time the Stanford IQ Test was commonly used, and since Millard used the *Stanford Reading Achievement Test*, it is likely he also used the Stanford-Binet IQ test). Using the *Stanford*

Reading Achievement Test scores collected every 6 months, comparisons were made in terms of the actual scores and their relation to the predicted growth patterns of the students.

Results of the non-grouped students showed that the reading performance of girls was superior at each age level, with the greatest differences at 8 years 6 months. As well, on average, girls started to learn to read at a slightly younger age of 72 months, whereas boys started at 73.2 months (Millard, 1940). However, when participants were matched with similar IQ scores, no significant gender differences were found in the age that reading started, nor were there significant differences in their growth rates or their levels of performance.

Millard (1940) concluded:

1. Although there were individual differences in growth, these followed a predictable pattern of development.
2. There were sex differences in reading growth that favoured preadolescent girls from the time the girls began to learn to read. However, the groupings of boys and girls were not equal in IQ.
3. "When allowances were made for differences in intelligence, no significant differences were found to exist between boys' and girls' scores" (Millard, 1940, p. 104). Students with higher IQ scores, both boys and girls, began reading earlier and continued to grow to a higher level than students with lower IQ scores. Students with higher IQs had markedly superior reading achievements than the groups with lower IQs. (p. 104)

Millard (1940) concluded that although children differ in their rates of reading growth, as individuals they continue to grow at an amazingly regular rate. Growth patterns of the individual are quite different from the straight-line pattern of growth represented by the Stanford Reading Achievement reading norms, which did not appear to accurately reflect a child's actual growth in reading, according to Millard. When differences in intelligence were taken into account, sex differences in terms of reading growth were negligible. However, many educators of the day observed that girls tended to make better progress in elementary school, particularly in the early years, which led to an increased interest in and focus on sex differences in reading readiness.

Sex differences in reading achievement and reading readiness.

The goals of Fra Samuels's (1943) research were fourfold. 1) to examine if there were sex differences in readiness and reading achievement; 2) to calculate the correlation between various measures of readiness; 3) to analyze if there was a correlation between measures of readiness for the use of predicting reading achievement as measured by a standardized reading test; and 4) to determine sex differences in readiness and achievement when boys and girls are paired by mental age (p. 594). During the first 2 weeks of school, the *Kuhlmann-Anderson Intelligence Test*, the *Monroe Aptitude Test* (citations not provided), a drawing test (specific test not specified), and teacher-based evaluations were conducted on 216 boys and 237 girls in Grade 1 in the Phoenix Public School District. The *Kuhlmann-Anderson (KA) Intelligence Test* measures individuals' academic potential. Unlike other IQ tests, which measure innate skills, the KA Intelligence

Test examines what the child had already learned in order to predict subsequent performance. Released first in 1927, the KA underwent eight editions and remained in publication for 55 years (Hiscox & Rodger, n.d.). *The Monroe Aptitude Test* was intended for students entering Grade 1 and could be administered to individuals or small groups. The test included subtests in visual, motor, auditory, and language skills (Standish, 1960). At the end of Samuels's 2-week period, children were grouped according to the results of all pre-tests combined, and the lowest students were identified and grouped for the purposes of offering additional instruction and support. Before the end of the school year, all students' reading achievement was assessed using the *Gates Primary Reading Tests Form 1* (no citation provided). It was difficult to show precisely how much progress boys and girls made in the readiness groupings, as these units were fluid and students moved out once their teachers saw improvement. However, Samuels noted that of the 49 boys who scored poorly on the *Monroe Aptitude Test* and who were placed in reading readiness groups for all or part of the year, 20 of them achieved a grade level of 1.95 on the Gates Reading Tests. Samuels concluded, "It seems reasonable to assume that their presence in the readiness group was a determining factor in their progress" (p. 599). Unfortunately, Samuels does not identify how many girls were in the reading readiness groups nor what their progress was. Results from the study indicated low coefficients between the various measures of reading readiness, suggesting that no one measure was sufficient to determine the accuracy of a child's eventual reading achievement. A high score on one measure did not correspond to a high score on another. The

best predictive measure for both boys and girls was the intelligence quotient (.49 for boys and .51 for girls). Teachers were more successful at predicting success in the reading achievement of boys (.37) than of girls (.35). The test with the lowest predictive rate was the *Monroe Aptitude Test* (.36 for boys and .43 for girls), a test specifically designed to measure reading readiness. Overall, girls scored higher on the reading achievement tests: scores ranged between 1.25–3.44 for girls and between 1.25–3.08 for boys, with an average score of 2.08 for boys and 2.51 for girls. Interestingly, reading achievement scores were recorded as a composite score, with no breakdown of the individual components (e.g., word reading, vocabulary, comprehension). Thus, it is impossible to examine which reading component(s) boys scored lower than girls. To test the conjecture that the difference between the sexes was not just a matter of IQ, since the girls' mean IQ score was higher than that of the boys (difference of 5.07), a comparison was made by pairing 100 boys and girls on the basis of mental ages. Although Samuels indicated that students were paired with students of similar chronological age (within 3 months), he did not indicate why he used mental age over IQ for comparison. Also interesting is the explanation Samuels provided for not including the method of the comparison: "The length of this article does not permit a full description for the arbitrary method used, but the conclusions drawn were that the girls were superior in more pairs than were the boys" (p. 601). Also, the specific scores of the paired results were not included. Samuels provided the method he used for calculations, which was the formula outlined by Lindquist in *Statistical Analysis in Educational Research* (1940), and he states, "The mean

differences in favor of the girls in the two measures of the achievement are statistically significant" (p. 601). Details of the measures of achievement were not provided. Samuels concluded:

1. To optimize individual children's capacity, school administrators need ways to determine differences in reading readiness and to address these differences.
2. Teachers need to provide instructional strategies to cope with individual differences that exist in their classrooms.
3. No one measure can be used to evaluate reading readiness, and groupings should be flexible and fluid with teachers consistently assessing and working with individual students.
4. Further research was needed to determine the causes of sex differences. (p. 603)

Although the first three conclusions are justified by the results of Samuels's study, it is questionable to conclude that sex differences exist in the population studied. In order to verify and confirm such conclusions, an explanation of the method used in the paired comparison would be needed as well as the inclusion of the data and not just the formulae used.

A paper presented by Frank Pauley in 1950 in Atlantic City at the meeting of the American Educational Research Association and later published in the *Journal of Educational Research* (1951) indicates that boys develop at a slower rate than girls. Therefore, recommendations were made to admit girls into school earlier than boys. However, after I reviewed this study in detail, I found that it

did not meet the criteria of a rigorous study. The paper did not include data or describe a methodology; in actual fact it was a position paper; a growing phenomenon in research journals post-WWII. Interest in sex differences in early literacy development was a topic of interest before and after the Second World War. Additional studies conducted on reading readiness after WWII are included in the following chapter.

Synthesis of Findings

Almost a century ago, Lincoln (1927), one of the first researchers to thoroughly investigate the topic of sex differences, examined both the physical and mental development of boys and girls. In terms of reading achievement, Lincoln's findings were inconclusive due to conflicting evidence. He emphasized the need to report reading achievement by age and not grade since many classrooms at the time had students with a range of ages. The design of reading achievement tests was debated (see Commins, 1928). How is it that boys could be superior to girls in tests of history and literature and then underachieve in reading? A much later analysis of the *Stanford Reading Achievement Test* questioned the validity of the instrument. Findings from Berman and Bird (1933) indicated that although girls could read faster at the college level, boys had a higher achievement level (however, Berman and Bird did not provide data to support these claims). Traxler (1935) determined that boys and girls at the high school level read at equal rates. Regardless, the speed of reading did not translate into increased comprehension. Stroud and Lindquist (1942) found that academically girls generally did better in elementary school and that boys did

better in high school. The magnitude of sex differences in reading comprehension, although greater in elementary than in high school, were small and no significant differences were reported. Jordan (1937) studied the achievement of high school students and concluded that girls were as competent as boys. Millard (1940) was interested in the growth of boys and girls in reading over time. When reading achievement was matched to rates of growth and IQ, no differences were noted between the sexes but the superior maturity of girls was observed. Assuming that physical maturity matched mental development, researchers like Samuels (1943) examined reading readiness to determine whether or not boys indeed needed extra support in the classroom. However, due to insufficient data on paired groupings and an incomplete explanation of methods used to make comparisons, definitive conclusions were not and are not possible. Nevertheless, Samuels noted that primary teachers needed to be aware of individual differences in reading readiness and to adjust their teaching methods to meet the needs of the students in their classrooms.

Implications for Boys and Reading Comprehension (1921–1945)

Between 1921 and 1945, research in education, and in particular reading, flourished. Reading as a field of study emerged separate from psychology. Led by pioneers such as William Gray and Arthur Gates, reading researchers worked together with general agreement on matters of reading theory and pedagogy. The belief was fostered that reading instruction should incorporate many approaches and feature the use of engaging reading materials that matched children's interests. Reading achievement took front stage in elementary schools and was

often the criterion for passing or failing students in school. Researchers advocated for knowledgeable teachers able to provide for the wide range of student abilities and interests in their classrooms.

Studies indicated that by adolescence, boys and girls generally had different reading interests. Researchers focused on individual needs and delved into how to identify and support students who found reading difficult. Although some studies supported the conclusion that girls read faster than boys (Berman & Bird, 1933; Lincoln, 1927), others noted no differences (Traxler, 1935). Sex differences were explored regarding the belief that girls appeared to be more successful in reading in the early grades while boys surpassed girls in high school (Commins, 1928; Jordan, 1937; Lincoln, 1927; Stroud & Lindquist, 1942). Was the superiority of girls in the primary grades related to maturity? If so, should boys start school later? The relationship between physical maturity and mental growth was hypothesized but not confirmed. The correlation between IQ and reading achievement appeared to be strong (Millard, 1940; Samuels, 1943). Throughout the ebb and flow of reading research, sex differences formed a recurring theme, resurfacing in many topics in reading research. Notwithstanding, no conclusive evidence was produced that boys and girls differed in their reading achievement.

Chapter Six: Reading Research and Theoretical Models in Gender Development: 1946–1980

Introduction

There were a number of social-political pressures after WWII that influenced educational reforms. Social-political events ignited demands in the 1950s and 1960s for federally funded programs to improve schools especially for the economically disadvantaged. Record numbers of children entered school and many experienced reading difficulties which prompted calls from parents and teachers to revise classroom-reading practices in order to reduce the gap in achievement (Allington & McGill-Franzen, 2004). Rudolph Flesch (1955), in his bestselling novel, *Why Johnny Can't Read?,* castigated the use of basal readers and pedagogical approaches adopted by most schools at the time. Flesch advocated for the return to a systematic phonics approach—a way in which children were taught to sound out unknown words (Monaghan, 2007). In contrast, within the reading community, based on new theoretical models, alternative advances were proposed. The view that language was a natural process found favour amongst researchers who promoted the whole language movement, and preferred the use of quality literature to instruct reading rather than to teach isolated skills such as phonics and decoding rules (Goodman & Goodman, 1980). The intertwining of more sophisticated theoretical models within the reading field as well as external social-political forces shaped reading research during these 35 post-war years. Separate from research within the reading community, was the scholarly work conducted on gender development.

Initially, developmental psychologists examined sex differences in terms of measurable variables such as physical growth and achievement. In the 1960s and 1970s, influenced by social learning theory and cognitive developmental theory, fundamental changes in gender research took place that remain influential today (Blakemore et al., 2009).

My goal in this chapter is to analyze in detail the evidence-based studies on sex differences in reading during the time period 1946–1980. Through the analyzes of the 35 years of research findings, I examine specifically the interplay between reading and gender research in order to gain insight into how the two fields influenced one another. To retrieve relevant research studies, the keyword terms "reading achievement" and "sex differences" were used to pursue all major and pertinent sources. Prior to the 1980s the search term "sex differences" was applied since it was the term adopted in the literature at the time. The terminology psychologists preferred to use after 1980 was gender differences (Blakemore et al., 2009), which will influence how studies will be searched and retrieved in the next chapter. The database searches provided 471 hits in Google Scholar, 181 in PsycInfo, 31 in ERIC, with fewer numbers in ProQuest, Web of Science, Academic Search Premier, and CBCA Ed. In addition, the search technique of "snowballing" was employed wherein references from retrieved studies were crosschecked and used to locate related and relevant investigations. In all, 26 studies deemed topical were procured, analyzed and synthesized.

To effectively identify the central concepts and make connections between and among the research findings, the 26 studies were organized into four themes

with approximately 5 to 8 investigations in each section. In theme one, I examined sex differences in reading readiness when aspiring readers are beginning the journey of reading. Studies that challenged the view that boys underachieved in the early years led to the exploration that supported no sex differences in reading achievement detailed in theme two. Issues around test design and data interpretation are raised in these studies. Theme three discusses the studies conducted during the post-war decades that proposed female superiority in reading achievement. Additional variables such as race, socioeconomic status, and IQ were considered in these studies. In an attempt to find the cause of sex differences in reading achievement cross-cultural studies were conducted. The purpose of these studies examined in theme four, was to determine if differences between boys' and girls' reading achievement were biological and therefore uniform across nations.

The major findings from the four themes described, provide insights into the approaches and perspectives researchers took while investigating sex differences in reading that followed WWII. Detailed descriptions of the 26 studies are summarized next.

Theme I: Reading Readiness and Sex Differences During the Post War Period 1945–1980

As reported in Chapter Five, a number of children failed their first grade in school due to reading difficulties (Percival, 1926), resulting in an increased interest in the emergent reader or what was known at the time, as reading readiness. The number of articles published on the topic of reading readiness was

highest between 1937 and 1940, declined during WWII, and then gradually increased after the war years (Smith, 1950). Reading readiness studies prior to WWII are included in Chapter Five. The studies completed after the Second World War are examined and explored in more detail in this chapter.

Marjorie Wight Carroll (1948) analysed the statistical data compiled from four pre-reading tests to ascertain if there were differences between boys' and girls' reading at the emergent reading stage. The tests analysed included: 1) an unpublished survey completed by Zeta Brown at Boston University using the *Stone and Grover Classification Test for Beginners in Reading*, and the *Gates Primary Tests* (1-2-3), 2) the *Monroe Reading Aptitude Tests*, 3) the *Gates Jim and Judy* tests, and 4) the *Ready-To-Read Games*, an unpublished test by Walter Dearborn and Marion Cushman (no references for any of the tests). Amassing data from 1,100 children and 38 items (e.g., visual discrimination, knowledge of letters, articulation, oral language) girls performed better on 24 items and the boys on 14. These tests were administered before formal reading instruction and thus were purported to be a measure of children's aptitude for reading and not the result of teaching methods. Is there a relationship between reading readiness and reading achievement? As noted in Chapter 5, Samuels (1943) concluded that reading readiness tests were poor predictors of future reading achievement. Regardless of the validity of reading readiness tests, Carroll's (1948) study did not measure reading comprehension, and therefore does not support the claim that there are sex differences in reading achievement.

Sex differences in reading readiness as well as reading achievement were studied by Virginia Konski (1951) in two cities in Missouri with a group of first grade Caucasian boys and girls. The author did not indicate why only Caucasian children were included in the study. One can presume that the researcher was attempting to create homogenous groupings for the purposes of comparison. Although it was cited that the groups were similar on age and intelligence, the total number of children included in the study was not cited. Students were assessed at the beginning of Grade 1 using a battery of reading readiness tests (names of the tests used were not provided). Then at the end of the school year, after a full year of reading instruction, the students were reassessed on four areas of reading achievement (language, paragraph reading, word pictures, and word recognition). However, the name of the reading comprehension test used to assess the students was not provided. The results of the reading readiness tests administered at the beginning of the school year did not indicate any significant differences between the boys and girls as a group. Konski (1951) concluded, "There is no need to prolong the reading readiness program for boys as a group, or to raise the age of school entrance for boys" (p. 920). Results from the reading achievement tests administered at the end of the school year revealed that there were no significant differences on the word recognition and word picture tests, but that there were significant differences on paragraph comprehension. However, no data were provided to allow for a cross check or to determine the level of significant differences. Since reading readiness was not a factor in reading achievement differences, Konski concluded that the disparity between the

scores was due to environmental conditions such as home factors, reading materials or differences in interest in reading. In conclusion, rather than delaying the entrance age of boys, Konski recommended that educators focus on preventing reading disabilities. Details on how this prevention should be implemented were not mentioned.

Contrary to Konski's (1951) findings were the studies reported by Frank Pauly (1951) from the Tulsa Public Schools in Tulsa, Oklahoma. He presented his analysis, *Sex Differences and Legal School Entrance Age*, at the 1950 Atlantic City meeting of the American Education Research Association. Pauly challenged the legal school entrance age and made recommendations for a flexible, individualized entrance age for boys in kindergarten and first grade. These recommendations were based on data from Pauline Freeny's research, on which Pauly provided statistical assistance. Freeny assessed 1,411 pupils in Grades 6 to 8 using the *Stanford Achievement Test* for paragraph and word meaning. Results indicated that girls scored better than boys (although neither data, nor citation of the test were provided), even though boys on average were older in every grade. Pauly provided data on only age differences between boys and girls and from this data noted that boys were older, thereby implying that more boys were retained. However, it was not stated that boys were retained based on poorer reading achievement. Pauly speculated that if boys enter school at an older age, they will make greater progress in their reading, would not be retained as often, and would have greater success in school. The leap in conclusion from age differences in

Grades 6 to 8 and the delayed admittance of boys to Grade 1 is based on incomplete information and is thus highly questionable.

One year later, Pauly (1952) published another position paper that proposed a later entrance of boys to school and for schools to consider a longer pre-first grade period. Included in the paper were supportive and non-supportive comments from a first-grade teacher, a reading specialist, superintendent, principal, and parent but no evidence. Interestingly, there was a wide range of views on the topic, some who agreed with Pauly, and others who criticized Pauly's interpretation of the data and claimed that it was misleading. Although the discussion paper was fascinating to read, it was merely an unsystematic collection of views on the topic of delaying the entrance age of boys to school.

A third article "Let's Give BOYS a Break!" was published by *Phi Delta Kappan* (1959) on the subject of a differentiated legal school entrance age based on Pauly's research. The author of the article was the publisher (*Phi Delta Kappan*) who synthesized Pauly's research findings. Pauly assessed the achievement of 29,992 students from Grades 2 to 8 from Tulsa in 1956–57. The *Chicago Reading Tests* were administered in Grades 2 and 3, the *Metropolitan Achievement Tests* in Grades 4 and 5, and the *Stanford Achievement Tests* in Grades 6, 7, and 8 (no references for the tests were provided). The results were plotted by grade that compared age and achievement. The girls' scores were used as the norm and the boys were compared on age and test scores. For example, on average in Grade 2, boys were 1 to 2 months older than the girls, but their test scores were 4 months lower than those reported for the girls. This result led Pauly

to conclude that boys should go to school 6 months later than girls. The test scores reported were compiled from the overall reading tests in Grades 2 and 3, and a composite of achievement tests including mathematics and reading in Grades 3 to 8. The specific breakdown of the scores was not provided, only the mean, therefore it was impossible to determine the specific details of the various components of the achievement tests (e.g., word recognition, comprehension, arithmetic facts, problem solving). In addition, the statistical significance of the means at each grade were not reported. The composite test scores revealed higher scores overall for the girls in Grades 2 to 8,

> The total deficit when age and scores are both considered varies from nearly 5 months in the second grade to more than 8 months in the eight grades, but apparently this deficit is overcome by the freshman or sophomore year in college. (*Phi Delta Kappan*, 1959, p. 282).

Incidentally, data from the freshman or sophomore year in college were not provided to substantiate this conclusion. Overall, Pauly's studies on sex differences in reading readiness and reading achievement are inconclusive to support the claims that either boys underachieve in reading comprehension or that there should be a change for boys in the school entrance policy.

Inez King (1955) from Highland View Elementary School in Oak Ridge, Tennessee examined the effect entrance age had on achievement of Grade 1 students. Students in the study were grouped by the age of entrance in Grade 1. One group consisted of 54 children (29 boys and 25 girls) who entered Grade 1 between the ages of 5 years 8 months and 5 years and 11 months. The second

group of 50 children (25 boys and 25 girls) were chronologically between 6 years and 5 months and 6 years and 11 months. All children remained in Oak Ridge school for 6 years. Both groups' IQ scores fell within the average range of 90–110 (mean of 100.08 for the older group and 102.04 for the younger group). The test used to measure the IQ was not specified. Folders on each child were kept and included information on 1) academic or grade standards, 2) average daily attendance, 3) progress through the grades, and 4) social or personal adjustments. The *Stanford Achievement Test* (no reference) was administered at the end of the sixth year of schooling. Results indicated that students who entered school after age six, on average, had a higher reading achievement rate, were retained less often, had better attendance, and fewer personal and social adjustments. Although King does not indicate any sex differences in reading achievement from the *Stanford Achievement Test*, she does note that of the 104 children, 11 students who were retained, 10 (7 boys and 3 girls) were from the younger group. This finding offered modest evidence for late entrance of boys in Grade 1. The one student from the older group was a boy. King did not outline why students were retained, so it is impossible to determine if reading was the factor. She concluded that the results from her study supported the proposal, "Sex Differences and Legal School Entrance Age" presented by Frank Pauley in 1951. However, as previously stated, results from Pauley's study do not justify the conclusion that there should be a different entrance age for boys, nor does King's study support that girls are superior in reading achievement.

In 1955 George Prescott set out to determine if there were sex differences in the performance of first grade boys and girls using the *Metropolitan Readiness Test* (no reference). The test consisted of six subtests: word meaning, sentence meaning, information, visual perception, number knowledge, visual perception, and number control. Data from 7,821 boys and 7,138 girls were collected during the first month of school in 1948 from 56 communities in 26 states. The results showed a lack of consistency in the superiority of girls over boys when chronological age was considered. Prescott concluded that different norms for each sex would be of little value for the *Metropolitan Readiness Test.* And, furthermore, that the results did not support the claim that girls were superior to boys in reading readiness tests or in reading comprehension.

Robert Dykstra and Ronald Tinney (1969) from the University of Minnesota completed a study on sex differences on three measures: 1) reading readiness (initial Grade 1 testing), 2) first-grade achievement (end of Grade 1 year), and 3) second grade achievement. Data were collected from 1,659 boys and 1,624 girls from eight projects involved in a Cooperative Research Program involving school systems in Pennsylvania, Michigan, New York, and New Jersey. A description of the program was not provided, but it was clear that these data represented the combined work of eight research groups from four different states. A battery of seven reading readiness tests was administered (group test of intelligence, auditory discrimination, letter knowledge, learning rate, visual discrimination, oral test of general vocabulary, and ability to follow oral directions). The girls surpassed the boys on six subtests with significant

differences reported at the 0.01 and 0.05 levels. The one readiness measure in which the boys were significantly better (at the 0.01 level) was the oral test of general vocabulary. It is interesting to note how cautious Dykstra and Tinney (1969) were in reporting their conclusions. They noted that although the general pattern of female superiority was evident in reading readiness, they pointed out that the magnitude of the differences varied considerably and were inconsequential in some cases (p. 624).

Sex differences in reading achievement at the end of Grade 1 were measured using the *Stanford Achievement Test, Primary Battery 1* (no reference) which consisted of word reading, paragraph meaning, vocabulary, spelling and word study skills. Mean achievement for the girls was significantly better (at the 0.01 level) on all but the oral vocabulary test. Using the *Stanford Achievement Test, Primary Battery 2* (no reference), a wider variety of measures was used to assess the students at the end of Grade 2 including: word meaning, paragraph meaning, spelling, word study skills, language, science, and social studies concepts independent of reading (presumably this means it was administered orally), and two arithmetic tests (computation and numerical problem solving). Girls scored significantly better (at the 0.01 level) in word reading, paragraph comprehension, spelling, word study skills, language, and arithmetic computation. Boys scored significantly better (at the 0.01 level) in understanding science and social science concepts, and they surpassed the girls (at a non-significant level) in arithmetic concepts for solving numerical problems which involved reading (Dykstra & Tinney, 1969, p. 626). Since sex differences were noted across all

eight projects, conjectures were made that differences between boys and girls were not related to the different approaches used in reading instruction but rather to biological differences. Furthermore, Dykstra and Tinney noted that although boys did not do as well as the girls on paper and pencil tests, they did better than the girls on measures of oral vocabulary tests, a finding that proved consistent between each test. Dykstra and Tinney (1969) cautioned teachers against drawing the conclusion that boys are less able to learn, because in most cases the mean differences on the subtests were not large (p. 628). Since instruction needs to be geared toward individual students, Dykstra and Tinney concluded that teachers should not concern themselves with the sex of a student, but rather the child's individual needs.

Results on reading readiness and achievement studies after WWII to 1980 were inconsistent and contradictory. Most of the studies examined pre-reading skills such as phonological and phonemic awareness, letter and word knowledge, and auditory-visual discrimination, but not reading comprehension (Carroll, 1948; Konski, 1951; Prescott, 1955). Other studies examined the success of students in later grades to test the reliability of reading readiness tests to predict reading success, and to determine the most ideal entrance age of Grade 1 students (Dykstra & Tinney, 1969; King, 1955; Konski, 1951; Pauly, 1951, 1952; *Phi Delta Kappan,* 1959). The popular belief that boys do not achieve as well as girls in reading in the early years was not supported in the research presented. Many studies reported that girls consistently did better than boys in reading subtests, however the differences in the scores were minimal and their importance largely

overstated. Secondly, the evidence presented is not convincing that reading readiness tests are good predictors of reading achievement.

Theme II: No Evidence of Sex Differences in Reading Achievement

Researchers (Anderson, Hughes & Dixon, 1956, 1957; Clark, 1959; Parsley, Powell, O'Connor, & Deutsch, 1963; Sinks & Powell, 1965; Yarborough & Johnson, 1980) during the time period 1945–1980 challenged the claim of sex differences in reading. Others questioned what the discrepancies in reading attainment would mean in terms of educational implications. Consensus among most researchers was the need for educators to focus on individual needs regardless of the gender of the student. As the quantity of evidence on the topic of sex differences in reading increased, it became critical to evaluate and record only the most valid evidenced-based studies.

Anderson, Hughes, and Dixon (1956, 1957) completed two studies with children from the University Elementary School at the University of Michigan. The first study was related to reading readiness, however since the study was extended and students were followed for 6 years, the results from both studies are recorded together in this section. The overall findings from the study conducted by Anderson, Hughes, and Dixon (1956) study were fourfold:

1) The age at which boys and girls learned to read varied greatly. 2) Girls tended to learn to read sooner than boys. 3) The age at which girls and boys learned to read correlated with intelligence (0.57 for girls and 0.54 for the boys). 4) Students' reading achievement in the sixth grade was correlated with the age at which students learned to read (0.67 for girls and 0.65 for boys). In an extended

follow-up study with the same students, Anderson, Hughes, and Dixon (1957) investigated the relationship between sex, age of learning to read, intelligence, and the rate of reading development (p. 481). Longitudinal records of 107 boys and 102 girls were analysed using the *Gates Primary Reading Tests*, the *Stanford Reading Achievement Test*, and the *Stanford-Binet Intelligence Test* (no references) The study revealed five main findings: 1) The age at which children learned to read did not relate to the rate at which children learned to read. Older students learned to read at a faster rate than students who started to read at a younger age. 2) The individual rate at which students learned to read, regardless of their IQ, varied greatly. 3) There was no sex difference in the rate of reading development in general, nor when rate was correlated with when students began to read. 4) Children with high intelligence scores (130 and higher) began to read earlier with more rapid progress, regardless of their sex. Children with lower IQs (scores of 100 and lower) learned to read at a later time, and developed at a slower but variable rate. Boys in the IQ range of 100 or less, although they started to learn to read later, when they did begin the reading process progressed at a faster rate than the girls. Students in the high IQ group (130 and greater) continued to increase their intelligence quotient, while the group with an IQ of 100 or less remained the same. The IQ of the residual group (IQ between 100 and 129) showed modest increase, and the increase was related to the rate of reading development. An observation from the results is the weight the *Stanford-Binet Intelligence Test* (no reference) gives to language acquired through reading rather than through non-reading tasks (Anderson, Hughes & Dixon, 1957, p. 491). In

order to understand the context of this research, it is important to note the demographics of the sample and nature of the study. The mean IQ of the entire group was 120; therefore the pupils in this study were a superior group. As well, reading instruction was not conducted in a formal way commonly observed in regular classroom settings. At the University Elementary School at the University of Michigan, children were introduced to reading according to their readiness. Children selected the books they wanted to read and progressed at their individual rates. Due to the approach taken in teaching reading, the researchers assumed that the progress students made, was not the result of reading interventions or teaching methods but rather based on the child's individual rate of growth (Anderson, Hughes & Dixon, 1957, p. 482). Overall, the conclusions from this study revealed the tremendous variability in the age of learning to read and in the individual rates of reading development. The overall conclusion is that the rate at which a child learns to read is related to intelligence (which is not distinct from environmental influences) and not the sex of the child.

Educators debated educational implications based on sex differences in achievement. Should boys be admitted to school later than girls, or should legislatures lower the legal entrance age for girls? Also contentious was the position that mental age norms needed to be revised, using separate tables for boys and girls. These were some of the questions Willis Clark (1959) attempted to answer in his study. Clark was also sceptical of the design used in previous studies on sex differences. He noted that these studies, usually involved restricted populations—a single school, community, or environmental group. Further, most

of these previous studies have analyzed differences between obtained scores with little or no effort to control such factors as chronological age when analyzing differences in mental ability or mental ability when analyzing for differences in achievement performance.

> The true test for sex differences is to equate or match so that the school grade, age, and mental ability characteristics of a group of boys and a group of girls are the same, and then test to determine achievement test performance. (Clark, 1959, pp. 75–76)

The *California Test of Mental Maturity* (CTMM) and the *California Achievement Test* (CAT) (no references) were used to determine sex differences. Random samples of 75 boys and 75 girls were drawn from Grades 3, 5, and 8, from 69,354 students across 341 systems in 48 states, 18 geographical areas and four community-sized categories. No significant differences were found between the boys and girls in all three grades on reading vocabulary, reading comprehension, and arithmetic reasoning. Girls did better than boys on mechanics of English at Grades 5 (significant at the 0.01 level) and 8 (significant at the 0.001 level), and on the spelling test the girls performed better in all three grades (significant at the 0.05 in Grades 3 and 5, and 0.001 in Grade 6). Clark (1959) concluded, since no differences existed in the basic skills of reading and arithmetic, other educational factors such as interest or instructional material were responsible when differences between boys and girls are found (p. 76). Also, since no differences were noted on the intelligence tests, separate norms were not required. Clark (1959) stressed the variability of the results on both intelligence and achievement

at every grade, which indicated the need for educators to deal with individual differences irrespective of the sex of the student (p. 76).

Parsley, (1963, 1964) and his associates at the University of Alabama conducted two research studies on sex differences and achievement tests. In addition, Powell, a member of the research team from the initial studies, completed a third analysis. Their review of the pertinent literature left the researchers with the supposition that there were differences between boys' and girls' achievement levels in math and reading. The results of their first study "were so definitely contrary to this assumption that it was felt that they should be publicly noted" (Parsley, Powell, O'Connor, & Deutsch, 1963, p. 210). Five scores were obtained from 2,651 boys and 2,369 girls using the *California Reading Achievement Test* (reading vocabulary and reading comprehension), the *California Arithmetic Test* (arithmetic reasoning and arithmetic fundamentals), and the *California Test of Mental Maturity* administered to children in Grades 2 to 8 in an urban-suburban school district in Ohio (no references included for all tests). Results from the total group were subdivided into five IQ groups (75–94, 95–104, 105–114, 115–124, and 125 and up). Analyses of the data indicated that there were no differences between the sexes in any of the grades for any of the achievement tests and that "differences between the sexes fail to approach significance and are, if in fact, very small" (Parsley, Powell, O'Connor, & Deutsch, 1963, p. 212). This preliminary study prompted the researchers to investigate further sex differences in achievement. The results of the second study were different to the findings from the first. The achievement test scores

within each IQ group for Grades 4 to 8 only were used to subdivide the groups into under-, average, and over-achievers. This was done by determining the average at each grade and then determining a plus or minus 0.6 grade placement. The 0.6 score was computed using the mean standard error of achievement for each IQ group. In addition, student *t* test scores were determined as a measure of statistical significance between the differences in mean scores. In reading comprehension, at all IQ levels, the girls excelled (at either the 5% or 1% levels of confidence).

> …female under-achievers do not under-achieve as much as male under-achievers; female average-achievers achieve at a higher level than male average-achievers; and female over-achievers achieve at a higher level than male over-achievers. Even where the differences are not statistically significant, the direction of the difference generally favors females. (Parsley, Powell, & O'Connor, 1964, p. 269)

Although the results of the two studies are contrary to one another, Parsley, Powell, and O'Connor report that even though sex differences in reading were found that "too much emphasis has been placed on sex differences" and that "one must be very cautious in generalizing them (results) into every-day curricular practices" (1964, p. 269).

In addition, Powell, one of the researchers in both studies, together with another researcher (Sinks), re-evaluated the reading data (vocabulary and reading comprehension), and a year later (1965) reached a **different** conclusion. The results from Sinks'and Powell's (1965) analyses indicated that from their results,

"one conclusion may now be apparent: that no generality of relationship as to reading achievement, with respect to reading vocabulary and reading comprehension, may be made on the basis of intelligence and sex for the population of this study" (p. 78). Differences were noted between the two researchers' analyses of the data. Sinks and Powell (1965) focused on only the reading data and they used a different formula to determine differences in the scores (Sinks and Powell used a chi square, while Parsley, Powell, and O'Connor used a *t* score). Because Sinks and Powell used a more advanced statistical analyses and provided a comprehensive set of tables, their study is considered more robust and trustworthy, and therefore their conclusion more reliable and valid. In summary then, it is reasonable to assume that **no** sex differences on reading achievement between Grades 4 to 8 were determined in this Ohio study.

Yarborough and Johnson (1980) reviewed 200 studies related to sex differences and more than 40 related to sex differences in reading and found the data fragmented with little evidence to support the claim (p. 56). They completed a 6-year study of 52 girls and 42 boys matched according to age, socioeconomic status, readiness test scores, and IQ scores. At the end of 6 years at school the students were tested on a battery of tests including three measures of intelligence—IQ, relational thinking, and cognitive abilities, four achievement tests including reading comprehension, vocabulary, language arts, and spelling, and 20 affective variable tests based on self-reliance and social adjustment. Even though the location of the study is not provided, the researchers were from the Departments of Curriculum and Instruction and Educational Foundation at Old

Dominion University in Norfolk, Virginia. The study was presented as a preliminary draft at the annual meeting of the American Educational Research Association in Toronto, Ontario in March 1978. The results disclosed that girls outperformed boys on measures of language arts, spelling, and six affective measures. There were no other significant differences or meaningful consistencies. However, the researchers provided only the data that indicated significant differences and not the entire battery of results therefore, crosschecking all data was not possible. Yarborough and Johnson (1980) concluded, "[T]he superior language arts and spelling achievement of girls, along with their affective advantages, may be mistaken as superior reading skills" (p. 55). This position was in opposition to the popular 1961 study conducted by Gates, who according to Yarborough and Johnson, "has been extremely influential and has, in many respects, colored the views of reading educators since its appearance, i.e., popularized the belief that girls' reading ability generally exceeds that of boys" (p. 55). The Gates' study will be examined in detail in the next section, *Sex differences that indicate that girls are superior in reading*.

No sex differences, or small and insignificant variances in reading achievement were noted in studies completed by Clark (1959), Parsley, Powell, O'Connor, and Deutsch (1963), and Yarborough and Johnson (1980). Sinks' and Powell's (1965) study revealed such a scattered range of results that they could not make any general statements about sex differences and achievement. Studies conducted by Anderson, Hughes, and Dixon (1956, 1957) found more variation within the sexes than between sexes. They also discovered that differences in

reading achievement were correlated with IQ but not sex. Issues of differences in methodologies and how best to analyze data surfaced in the studies completed by Clark (1959), Sinks and Powell (1965), and Yarborough and Johnson (1980). In some cases, differences in methods led to differences in results even for the same researchers.

Theme III: Sex Differences Indicate that Girls are Superior in Reading

A number of literature reviews were completed during the 1960s on the topic of sex differences and reading, an indication of the mounting interest in the topic (Cardon, 1968; Criscuolo, 1968; Flaherty & Anderson, 1966; Maccoby, 1966; Wyatt, 1966). The evidence-based studies cited in these reviews either have been examined in previous chapters or are included in this one. These reviews provided an opportunity to verify the completeness of my search for studies and to further cross-examine my appraisals of the research studies described.

As previously mentioned, Arthur Gates' (1961) study "*Sex Differences in Reading Ability*" was pivotal in swaying the thinking that girls generally surpass boys in reading achievement (Yarborough & Johnson, 1980, p. 55). An example of Gates' influence on further research on the topic is apparent by the number of scholarly articles and publications that cite his research; 114 in Google Scholar dating from the 1960s to present time, and over 57,000 hits recorded in Google. Gates' authority is understandable considering he was a research giant and renowned pioneer in the field of reading. He contributed over 300 books, articles and presentations (Sailors, 2007; Tostberg, 1971) on a variety of reading topics

including: reading readiness, reading methodology, reading materials, and remedial reading instruction (Sailors, 2007). No doubt Gates was a prolific writer and respected scholar, interestingly however, he conducted only two studies on sex differences and only one that pertained to gender and reading variances ("Experiments as the relative efficiency of men and women in memory and reasoning" [1917] and "Sex differences in reading ability" [1961].) It is incredible that a single study could have had such authority on the issue of gender and reading achievement.

Gates' study was conducted in the spring of 1957, a year after he retired from Columbia University. Gates remained affiliated with Columbia as he continued to research and write for many years (Sailors, 2007). Test scores were gathered from three subtests from the *Gates Reading Survey Test* (Speed of Reading, Reading Vocabulary, and Level of Comprehension, but not referenced) from 13,114 students (6,646 boys and 6,468 girls) from Grades 2 to 8 from twelve school systems in ten states (Gates does not identify the school systems or the states). The population "was approximately typical in intelligence or scholastic aptitude, socioeconomic level, and other pertinent respects" (Gates, 1961, p. 431). He did not indicate how he arrived at this information, nor did he use IQ tests to establish that the population was typical. Gates also reported that students were rarely required to repeat grades, which might be why he used grade rather than age scores. Data from the Grade 2 sample were eliminated as no significant sex differences were identified. Also, Gates indicated that there were issues of validity on the tests used at this level and that the test was not a good predictor of

reading achievement. Data were combined for Grades 3 and 4, 5 and 6, and 7 and 8, and raw scores were converted into grade equivalents. The results revealed that the girls were superior to the boys by 0.2, (Grades 3 and 4), 0.3 (Grades 5 and 6), and 0.33 (Grades 7 and 8) grade levels. All results, indicated, ($p = 0.01$ level) that girls were superior in reading comprehension. Gates claimed that no consistent trends were found, and that there was a slight increase in the difference between boys and girls from Grades 3 to 8. Interestingly, after close examination of the data it appears that the boys were closing the gap. A 2-month difference in Grade 3 is roughly 7% (a 2-month difference in 30 months of schooling) while a 3-month difference in Grade 8 is just over 3% (3 month difference in 80 months of schooling), which suggests that the boys were indeed closing the gap.

When Gates (1961) examined the lowest reading comprehension scores, he noted that the boys outnumbered the girls in all the tests by 2 to 1. The boys however, did not demonstrate a consistent superiority within the higher scores. "The greater variability of the boys seems therefore to be due primarily to the fact that a greater proportion of boys got low scores" (Gates, 1961, p. 432). It is interesting that Gates, once a former student of Thorndike, continued to assess and comment on the variability theory that was promoted by psychologists from the early 1900s. As reported by Anastasi (1958), the doctrine of variability was regarded as a fundamental biological law that "enjoyed wide popularity and was adopted by a number of psychologists during the first quarter of the present century. Nor is it completely absent from contemporary writings, especially

popular discussions of sex differences" (p. 456). Gates' comments on boys' reading achievement in his study is proof of Anastasi's summation.

Marian Wozencraft (1963) from State University College in Geneseo, New York conducted an investigation on sex differences comparing a number of abilities using the *Stanford Achievement Test* (no reference). Using a stratified sampling method from 121 public elementary schools in Cleveland Ohio, 564 Grade 3 students out of a total possible 5,708 and 603 Grade 6 students of out of a total 5,059 were assessed. Students' "Probable Learning Rate" (PLR) or intelligence scores were assessed using the *Kuhlman Anderson Test* in Grade 3 and the *Cleveland Classification Test* (no references) in Grade 6. Students were divided into three groups (low, average, high) based on their IQ scores. In Grade 3 no significant differences were noted in paragraph meaning (reading comprehension) in the upper and lower groups. Sex differences favouring girls ($p = 0.01$ level) were detected in the middle group. However, overall the middle group of girls had a marginally higher mean PLR than the middle group of boys (100.1 for the boys and 101.9 for the girls). In the overall group in Grade 6, sex differences were noted ($p = 0.05$ level) however, when the group was subdivided into high, middle and low groups, no significant differences were noted in paragraph meaning. Overall, with the exception of one middle group in Grade 3, no sex differences were noted in the Wozencraft study when students' IQ scores were matched to reading comprehension scores.

Nita Wyatt, after a review of the then current literature, was determined to resolve whether different approaches to reading instruction in Grade 1 would

yield greater reading achievement for boys. Two experimental and one control group for a total of 633 children were organized from three elementary school districts. The specific location of the study was not provided. Students from each group were from similar backgrounds (based on parents' level of education and socio-economic status) and schools, and had teachers with similar training and experience. All 30 of the teachers in the study were women. In experimental group one, the boys and girls were separated during reading instruction. The boys were given boy-friendly books, provided with manipulatives and games to move about during reading group sessions, and given concrete objects, such as pictures when teaching word meanings. The teaching procedures were "designed to help teachers adjust their teaching methods to the behavioral tendencies of boys as well as girls" (Wyatt, 1966, p. 598). In a second experimental group, teachers used a linguistic approach to reading instruction that introduced spelling patterns in context and focused on comprehension and fluency. Since no published materials were available that supported this method, teachers used several different published textbooks and supplemented their teaching with teacher-made materials. The control group was taught reading using basal readers and was grouped on reading ability –a typical approach to reading at the time. After 140 days of reading instruction, students were assessed using the *San Diego Inventory of Reading Attitude* and the *Stanford Achievement Test* (no references). Results from the achievement tests showed significant differences with girls having higher means for paragraph meaning, spelling, and for word study. The level of

significance was not stated. As well, data tables were not included to enable others to confirm the claims made.

Richard Jantz (1974) from the University of Maryland investigated the effects of sex, race, socioeconomic status, and intelligence on reading achievement of 3,188 students who attended Grade 6 between the years 1968–1971 in one urban, mid-western school district. The purpose of the study was to determine students' reading achievement in Grade 5 and then examine the gain in performance by the end of Grade 6. Data were collected from each students' permanent records including: identity, sex, race, IQ scores, parental or guardian occupation, school attended in Grade 6, reading scores in Grades 5 and 6. Although the researcher described in detail how the socioeconomic status of each student was calculated, no information was provided on which IQ or reading assessments were used. Results indicated that on reading performance, girls had a higher mean score than boys, White students performed better than Black students, pupils from higher socioeconomic groups outperformed students from lower groups, and pupils in the higher IQ grouping did better than pupils in either the upper-middle, lower-middle, or the lower groups. The differences were significant at the 0.01 level. The mean differences were greatest between I.Q. (ranging from 4.54 in the low IQ group to 7.88 in the high IQ group), followed by SES (7.73 for the high SES group and 6.05 for the poor SES group), race (6.35 for the White population, and 5.10 for the Black), with sex differences having the lowest mean difference (6.34 for the girls and 6.08 for the boys). However, Jantz found no significant differences related to sex, race, intelligence, or

socioeconomic status in the gains in performance between Grades 5 and 6. He observed a wide range in students' reading scores that were attributed to many factors (sex, race, IQ, SES) and that on average, students did not make a year's growth in reading from Grade 5 to Grade 6. In summation, Jantz (1974) concluded, "If educators are to continue to use standardized tests to assess the performance of pupils, then differences in the levels and gains in reading performance ought to be included in the formulation of educational objectives" (p. 94). Although Jantz makes an excellent claim for an individualized perspective on student achievement and instructional needs, his study reveals that IQ and SES are more influential when comparing reading achievement than are race or sex. Since he did not provide data, other than the average mean scores, or indicate the standardized tests used, it is not possible to confirm, based on his reported work, that sex differences in reading achievement existed as suggested by the researcher.

Edward Dwyer (1980) from East Tennessee State University assessed 157 students in Grade 2 from a large elementary school in rural Georgia on a number of language tests including reading achievement, listening comprehension, sight vocabulary, and paradigmatic language (word association) on the basis of race and sex. About one-third of the students were Black and two-thirds White. "The mean income of families residing in the area serviced by the school was substantially lower than that of the national average according to the Bureau of Commerce", 1974 (Dwyer, 1980, p. 209). A number of assessments were used including, the *Oral P/S Language Inventory* (paradigmatic language) (Dinnan,

1971), the *Metropolitan Achievement Test* (1971) for reading, the *Slosson Oral Reading Test* (Slosson, 1963) to measure sight vocabulary, and the *Diagnostic Reading Scales* (Spache, 1972) to measure listening comprehension. The scores of White girls were higher than those of White boys at the $p = 0.05$ level on reading, sight vocabulary, and listening comprehension. However, minimal differences were reported on the language tests between the Black boys and Black girls. I was left with the question of why there were significant differences between Whites but not among Blacks. This study highlights the strong relationship between socioeconomic differences and reading achievement rather than sex differences.

In all, during the period 1950 to 1980 five studies, 3 experimental and 2 correlational, purported that girls were superior to boys in reading achievement. Five literature reviews were procured, crosschecked and examined (Nita Wyatt's investigation included both a literature review and an experiment). Arthur Gates' (1961) study determined that girls were superior to boys in reading achievement by 2 to 3 months from Grades 3 to 8. The fact that Gates used grade rather than age in his analyses is questionable. There are wide ranges in ages in every grade, which makes such comparisons untenable. As well, Gates concluded that his groups were similar in IQ but did not use any measures to support this conclusion. The fact that Gates' study had such influence on the topic of sex differences in reading is one that needs to be challenged.

Other researchers investigated sex differences in addition to factors such as race (Dwyer, 1980), IQ (Wozencraft, 1963), instructional methods (Wyatt,

1966), or a combination of factors—sex, race, socioeconomic status, and IQ (Jantz, 1974). Jantz (1974) and Wyatt (1966) did not provide any data nor specify the standardized test(s) used to measure growth. Dwyer's (1980) study disclosed a sex difference between White girls and boys but no difference between Black boys and girls. The issues of race and reading achievement and possible socioeconomic factors were unanswered in his study. Wozencraft's study (1963) signified that there were no sex differences in reading when IQ scores were considered. In sum, there are many inconsistencies in the investigations presented, and no clear indication that girls are superior to boys in reading from any of these studies.

Theme IV: Sex Differences in Cross-Cultural Studies of Reading

Debates over sex differences in reading achievement included disputes over the cause for such differences. Many researchers from the 1930s to the 1950s viewed discrepancies between boys and girls based on biological attributes, claiming that girls matured earlier, had greater oral language skills than boys, and therefore had a general superiority over boys in literacy (Lincoln, 1927; Pauly, 1951, 1952). A transition in research took place in the 1960s and 1970s because of the development of more sophisticated theoretical models and a rising feminist movement (Blakemore et al., 2009). Researchers started to focus on environmental factors related to children's social and cognitive development. The appeal for cross-cultural studies was to test the hypothesis of female superiority in literacy. If "boys in countries with different cultures are found to surpass girls in reading, then the theory of biological causation would be effectively demolished

and supplanted by cultural causation" (Klein, 1977). Six studies that examined sex differences in reading across cultures are reviewed next, with reference to literature reviews that were completed during the period 1950 to 1980.

The first international comparison of achievement conducted between 1959–1961, was sponsored by the United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Education. One of the primary goals of the project was to determine the feasibility of undertaking future large-scale international studies (UNESCO Institute for Education, 1962). Data were collected from four achievement tests developed by the international representatives from the participating countries involved in the study (mathematics, reading comprehension, geography, and science) and one non-verbal aptitude test developed by the National Foundation for Educational Research in England and Wales. The study was a pilot project targeting 13-year-olds in 12 countries. Samples ranged from 300 (Switzerland) to 1,732 (Israel) with a total number of 9,918 students assessed in all. The tests were originally prepared in English, French, and German and then translated into eight languages (English, Finnish, French, German, Hebrew, Polish, Serb-Croatian, and Swedish). Although many of the subtests consisted of only 4 to 10 items, Thorndike who presented the data (UNESCO Institute for Education, 1962, p. 21), described the reliability of each test as "fairly satisfactory" with each subtest ranging from: 1) non-verbal (0.89), 2) mathematics (0.81), 3) reading (0.81), 4) geography (0.70), and 5) science (0.62) (p. 21). A general estimate of reliability was obtained using the "Kuder-Richardson Formula Number 20 from the average standard deviation

and the average item difficulty over 11 national groups" (UNESCO Institute for Education, 1962, p. 21). The results from Yugoslavia were not included in the study because the data arrived after the analyses were completed. Since the pilot project had limited resources,

> it was not practical to try to get a truly representative sample of the 13-year-old population in each country. Sampling procedures varied from country to country . . . limited to one or a few communities or regions that were thought to be representative of the country as a whole. (UNESCO Institute for Education, 1962, p. 21)

Raw scores were converted into means for each country. The mean of all 11 countries was calculated to form a grand "total mean." This average was set at O. The mean of each country was then compared around the grand "total mean," with positive scores showing an average above the total mean and a negative score as one lower than the total mean. This index provided a scale to compare the magnitude of the differences between each country. A key finding from the pilot study was a discrepancy in sex differences among the 11 countries. In all countries except for the United States, the boys, on average, did better than the girls. "On average, over all countries and tests, the boys fall about a fifth of a standard deviation above the girls " (UNESCO Institute for Education, 1962, p. 27). More specifically, a difference between the boys and girls was smallest in Sweden and Scotland, and largest in Poland, Germany, and Belgium. On specific tests, the girls outperformed the boys in reading comprehension and boys

achieved better results in science. Since this study was a feasibility pilot, with questionable sampling procedures, the results must be considered with caution.

However, encouraged by the consistency and item statistics from the Reading Comprehension items from the feasibility study, Thorndike in 1973 went on to launch an international study with 15 countries. Within the overall study, reading comprehension was one subtest. Also included were subtests measuring science, literature, civic education, English and French as foreign languages, with 3 subgroups: 10-year-olds, 14-year-olds, and students in the final year of secondary education. Two chapters in the final report describe the instruments developed for the study, the selection of samples to be tested, and how the testing was conducted. The remaining eight chapters of the report discuss the results of the survey. In this international study, sex differences were not detected. "Correlations of Reading Comprehension scores with sex are small and inconsistent from country to country" (Thorndike, 1973, p. 78). The two overall findings from this cross-cultural study revealed: 1) a large difference in the reading levels between developed and developing countries, and 2) within developed countries reading achievement of individual students depended on the economic resources of the home such as education of the parents, home income, and availability of reading materials. The fact that sex differences were not included in the final summary indicates that sex differences were not a significant factor in determining the success of students' reading comprehension. The zeal for international studies continued into the 1990s and are still conducted. These studies will be addressed in the chapter that follows.

A frequently cited cross-cultural study, entitled "Reading Achievement of German and American Children," was completed by Ralph Preston (1962), Professor of Education from the University of Pennsylvania (In Google Scholar, 92 scholarly articles cite Preston's research). Post WWII publicist Rudolf Flesch (1955) in his book *Why Johnny Can't Read*, openly criticized the reading pedagogy adopted by schools in the United States. The purpose of Preston's study was to test the reading approaches used in the United States and compare them to the contrasting approaches adopted in European countries. International comparisons between countries that used different approaches to teaching reading were one way to verify which approach yielded the greatest gains. Also, cross-cultural comparisons were a way to determine if sex differences in reading achievement were biological or culturally determined. If the differences were the result of biological factors, then cross-national studies should indicate similar global trends (e.g., boys uniformly underachieving). If, however, the cause for sex differences in reading achievement were socio-culturally related, sex differences would vary from country to country (Johnson, 1976).

Preston (1962) assessed students in Grades 4 and 6 from Wiesbaden, Germany (1,053), and from public schools in Philadelphia and vicinity (1, 338) (p. 351). The average IQ scores in the American subgroups ranged from 104 to 110. Preston did not indicate where these data were derived. The IQ scores of the German students were determined on the basis of the above-average socio-economic status of their parents and the type of school from which the students were selected. Students tested were from the Mittelschule and Gymnasium

schools which prepared pupils for medium-level business careers. Therefore, it was resolved that the German students had above average IQ scores, within a similar range of the American students (Preston, 1962, p. 351). Two reading tests were administered, the *Frankfurter Test* and the *Gates Reading Survey* (no references). The authors and dates were not provided. Each reading test was translated and made available in both German and English. Preston cited two major findings from this study. Overall, the American students did better on both the *Gates Reading Survey* and the *Frankfurter Test*, evidence that the reading approach in the United States was not inferior as suggested by Flesch. In Grade 4, the American children, both boys and girls, performed better on the reading comprehension tests on both the *Gates Reading Survey* and the *Frankfurter Test* than the students from Germany (differences between the girls' scores were significant at the 0.5 level, but were not significant for the boys). On the Grade 6 tests, the German boys on both reading assessments outperformed the girls in Germany and the boys and girls in the United States (differences between the boys' scores were significant at the 0.5 level, but no significant differences were noted between the girls). Since this result was deemed an anomaly – boys performing better than girls on a reading test, Preston's study became a signpost as proof that reading achievement was related to cultural differences.

Preston (1962) further hypothesized that sex differences in reading were related to the sex of the teacher, as Germany had more male teachers in the elementary grades than did the United States. Preston also made the deduction that reading in Germany was more commonly a male activity than in the United

States. Both statements were speculative and not based on research evidence. Although the average scores for the American students were higher than the students from Germany, analyses of the details indicated greater variability among the boys in the United States and among the girls in Germany, and further "suggested" that sex differences in reading were not biological but rather a cultural phenomena. So alluring were these findings on sex differences that the primary purpose of the study, which revealed differences in early reading methods favouring the American approach, was overshadowed. The results from Preston's study have been overgeneralized. A small sampling of one city (Philadelphia) is not a fair representation of the whole of the U.S.A. In addition, Preston used grade rather than age score comparisons and assumed that both populations had similar IQ scores. Unfortunately, the purpose of the study to compare reading methods was derailed by the identification of sex differences in reading achievement.

Dale Johnson (1973–74) completed a preliminary investigation on sex differences in reading achievement of 1,081 students in Grades 2, 4 and 6 between four English-speaking countries, Canada, England, Nigeria, and the United States. Six tests were administered and they included reading comprehension, vocabulary, structural analysis, and three phonics tests. The communities that represented each country included: St. James-Assiniboia, a school district near Winnipeg; Birmingham in England; Zaria in Nigeria; and Stoughton in Wisconsin, USA. Students were randomly selected from two schools in each of the communities. Selection of the four communities was not

discussed nor reasons provided. Reading comprehension was assessed using the *Gates-MacGinitie Tests* (no reference). In Canada, significant differences were reported at the Grade 2 ($p < .03$) in favour of the girls, but no significant sex differences were noted at Grades 4 and 6. In England and Nigeria, no significant differences existed at any of the grades in reading comprehension. The only test that favoured the girls in the United States was at Grade 4 (significant at the $p < .0002$ level). Even though Johnson concluded that his study supported Preston's study, previously described, Johnson noted the limitations of his study and cautioned readers against overgeneralizing. One small area is not representative of the population of each nation. Furthermore, sample sizes were small with about 50 students from each grade in each country with fewer numbers from Nigeria (8 and 20 students from Grades 4 and 6 respectively). From the 12 comprehension tests (three grades in 4 countries), only two revealed significant differences in reading comprehension that favoured girls. Based on the evidence provided, Johnson's study does not provide conclusive evidence to support sex differences in reading achievement. As indicated by Johnson, this study was an exploratory investigation that supported the need for additional comprehensive cross-cultural studies in reading achievement.

John Downing from the University of Victoria, with researchers from Denmark, England, Finland, Israel, Japan, and the USA, was interested in investigating how reading was perceived in terms of gender identification (Downing, Dwyer, Feitelson, Jansen, Kemppainen, Matihaldi, Thomson, 1979). About 100 students from six subgroups (Grades I, IV, VII, XII, college, and adult

populations) were asked to decide from a series of pictures if the various activities were male or female, including reading books. Some countries were not able to obtain all six-subgroup testings. Furthermore, the age of students tested in Grade 1 varied from country to country because there was a range in the legal age requirements at which students started school. Some limitations were also noted in the research methods such as cultural differences when interpreting the pictures (e.g., the stick person was not neutral but considered to be a boy, since the legs looked like pants). Overall, the results revealed that in Canada, England, and the USA, reading is deemed to be a girl activity; in Finland and Israel, reading is viewed as both a girl and boy activity; and reading is primarily considered a boy activity in Denmark and Japan. The data supports a previous study completed by Downing and Thomson (1977) who showed that reading is stereotyped as a female activity in some countries and not in others. Although these findings reveal the role culture plays in determining gender identity and reading, they do not resolve the question of sex differences in reading achievement.

The last study to be reviewed is one completed by Jeremy Finn (1980) from the University of New York at Buffalo. Finn analyzed the records of 10,294 students from 356 schools from the USA, England and Sweden, using data from the International Association for the Evaluation of Educational Achievement (IEA). The cross-nationally standardized tests were administered in May to students 14 years of age distributed between Grades 8 and 9. Scores were collected from measures of reading comprehension (52 multiple-choice questions), word knowledge (word pairs), reading attitudes (checklist), science

achievement (80 items including Biology, Chemistry, Physics, and Practical), and science attitudes (interests and participation). When analyzed for sex differences and attitudes in science and reading "[M]arked similarities in patterns of sex differences were found: In all three countries, male and female pupils have similar reading comprehension and vocabulary levels" (Finn, 1980, p. 24).

The debate over the cause of sex differences fostered a zeal for cross-national comparisons. If sex differences existed globally, then differences between boys' and girls' reading achievement levels were biological. However, if comparisons yielded contrasting results, then the cause for sex differences were culturally based. The first international comparison, completed by UNESCO Institute for Education (1962), was a pilot study. Although sex differences in reading comprehension were noted, the results were not reliable due to inadequate sampling procedures. No conclusive results were found in another exploratory study completed by Johnson (1973–74). Significant differences that favoured girls were found in only two of the 12 reading comprehension subgroups. A decade later, Thorndike (1973) chaired the second international study comparing the reading comprehension of three age groups from 15 countries. In this more rigorous study, sex differences in reading comprehension were small and inconsistent. The findings indicated that differences in reading comprehension between and among countries were determined by economics and related to resources of individual households. Finn (1980), 7 years later, re-examined the data from Thorndike's study and used the results from three of the 15 countries (USA, England, and Sweden). Finn's results indicated that in all three countries

no sex differences in achievement existed, albeit girls revealed a more positive attitude toward reading, although the results are not statistically significant. Preston's (1962) study was problematic because he assumed, without verifying, that the populations for his study were similar in IQ and were representative of their respective countries. The study completed by Downing et al. (1979), although important in verifying that reading is stereotyped as a female activity in some countries and not in others, it did not resolve the question of sex differences in reading achievement. Overall, no definitive conclusion regarding sex differences in reading could be drawn from the cross-cultural studies in the time period from 1950 to 1980.

Synthesis and Interpretation of Findings

During the three decades following WWII, interest in sex differences continued. Hundreds of related research studies were examined, 26 were identified that met the criteria of evidence-based, primary studies with methodologies clearly articulated, and conclusions supported by data. Even though established criteria was used to select the studies, after careful analyses and syntheses of the 26 investigations, inconsistencies were found in 17 of the studies. Seven investigations did not provide data to corroborate the results and conclusions stated (Jantz, 1974; Konski, 1951; Pauly, 1951, 1952; Phi Delta Kappan, 1959; Wyatt, 1966; Yarborough & Johnson, 1980). An additional nine studies revealed discrepancies in test design, or were deemed to have contrary data interpretations (Carroll, 1948; Dwyer, 1980; Unesco, 1962; Gates, 1961; Johnson, 1973–74; King, 1955; Parsley, Powell, & O'Connor, 1964; Preston,

1962; Wozencraft, 1963). One study assessed cultural differences in gender identification but not sex differences and reading comprehension (Downing et al., 1979). In all, nine studies met the requirements as outlined in the methodology (Anderson, Hughes & Dixon, 1956, 1957; Clark, 1959; Dykstra & Tinney, 1969; Finn, 1980; Parsley, Powell, O'Connor, & Deutsch, 1963; Prescott, 1955; Sinks & Powell, 1965; Thorndike, 1973).

Of the nine investigations deemed trustworthy, two studies within the first theme of reading readiness challenged the assumption that boys underachieve in reading in the primary grades (Dykstra & Tinney, 1969; Prescott, 1955). For many years, the belief prevailed that boys underachieved in early reading achievement because they matured later than girls (Pauley, 1951, 1952; Phi Delta Kappan, 1959). Although girls perform better in many of the pre-reading subtests, the differences were often small or inconsistent (Dykstra & Tinney, 1969; Prescott, 1955) and did not measure reading comprehension. Attempts to isolate and measure the pre-reading skills needed for reading during this time period were not valid predictors of future reading success (Samuels, 1943). This finding indicates the complexity of the reading process and the difficulty in using isolated pre-reading tests to determine future reading competencies.

New theoretical perspectives during the 1950s and 1960s focused on oral language development and its foundational role in learning to read centring on the importance of early literacy activities (Alexander & Fox, 2008). Since children come to school with diverse literacy experiences, the time at which a child is able to read is highly individualistic. How theory influenced practice can only be

speculated, yet evidence from the studies from this time period (1945–1980) reveals that researchers held various positions on the topic of reading readiness. Some researchers examined reading readiness in terms of a chronological age at which a child is ready to read (innate), thus promoting policies regarding school entrance age (Pauley, 1951, 1952; Phi Delta Kappan, 1959) while others advanced the position that early literacy teaching needed to involve developmentally appropriate activities geared toward the needs of individual children regardless of age or sex (Dykstra & Tinney, 1969; Konski, 1951). Theoretical disputes over the nature vs. nurture debate in early literacy development were evident in the studies examined. Some researchers speculated that differences between boys and girls were biological and related to physical maturity (Dykstra & Tinney, 1969; Pauly, 1951, 1952; Phi Delta Kappan, 1959) while others pointed to environmental factors pertaining to home and school conditions (Konski, 1951).

Five studies indicated that when differences in reading achievement were correlated with intelligence and/or chronological age there were no sex differences in reading achievement (Anderson, Hughes & Dixon, 1956, 1957; Clark, 1959; Parsley, Powell, O'Connor, & Deutsch, 1963; Sinks & Powell, 1965). Psychologists from the time of the early 1900s agreed that men and women did not differ in average intelligence (Pressey, 1918; Thorndike, 1914) but differences in reading achievement varied from individual to individual depending on IQ. These studies suggest a biological cause for differences in reading achievement, not related to sex, but rather innate individual differences.

Additionally, in *Theme II: No Evidence of Sex Differences in Reading Achievement,* the issue of test design was raised to explain possible differences in data interpretation. Studies that looked at restrictive populations with few controls rendered conclusions that were not reliable (Clark, 1959). The results from Clark's (1959) study determined that the norms tables used in reading achievement assessments did not need to be adjusted and that separate tables for boys and girls did not need to be created.

Issues over test design were also found in *Theme III: Sex Differences Indicate that Girls are Superior in Reading*. Inconsistencies in data collection and analyses of data were detected in the five studies that purported that girls surpassed boys in reading achievement. It is evident how critical it is to be cognizant of factors that affect validity, reliability and transferability when reading research investigations. In order to attain the most trustworthiness studies, applying evaluative criteria is essential.

In an attempt to settle the nature-nurture debate, cross-cultural studies emerged to test the hypothesis of female superiority in literacy. If sex differences in reading were detected worldwide then the cause for female superiority was biological. However, if sex differences between countries varied then the cause was environmental. Six cross-cultural investigations were procured, and two studies met the final criteria. These two studies found no differences between sexes (Finn, 1980; Thorndike, 1973). Differences in reading achievement were related to socio-economics rather than sex. These findings suggested

environmental differences between individual countries that were not related to innate differences such as sex.

Implications for Boys and Reading Comprehension (1945–1980)

Although not formally organized, prior to the Second World War the leading researchers in the emerging field of reading, agreed on most major points (Monaghan, 2007). However, after the release of the hypercritical bestseller, *Why Johnny Can't Read*, "when the entire field was threatened, as it was by Flesch's attacks, the need for an organization devoted only to the interests of the reading community" (Monaghan, 2007, p. 27), brought about the formation of the International Reading Association. Pressures from post-war political and social influences as well as heightened attention in solving the "problem" of reading acquisition, resulted in an increase in literacy research and the advancement of new theoretical positions (Alexander & Fox, 2004). Reading research was paralled by investigations on gender. However, after examining the research studies during the time period 1945–1980, it is evident that there was little exchange between these two fields of research, demonstrated by few theoretical positions presented in the studies completed on sex differences and reading achievement. Reading researchers suggested only either biological or environmentally causations in their findings. How reading research and gender theory influenced one another is explored more fully in the following chapter.

Results of the nine final studies overwhelmingly reveal a lack of evidence to support female superiority in reading from primary grades to high school. The concept of reading readiness was challenged and proposed instead was the

individual nature of acquiring language and the influences of language experiences and reading. The relationship between IQ scores and reading achievement was again noted, however factors such as socio-economic status and age scores not previously measured were reported in the literature and seemed to be more promising variables to explain differential performance. Discrepancies in study design, procedures, sample sizes, data analyses, data interpretation, and conclusions in studies conducted from the 1945–1980 time period is alarming. With increased interest in the topic of gender differences and reading achievement from 1980 to the present, it will be critical to adhere to the criteria of a trustworthy study outlined in chapter 3.

Chapter Seven:

The Reemergence of the Boys' Reading Achievement Crisis: 1981-2011

Introduction

Several social-political challenges gained ground as the world moved into the last decades of the 20th century. The emergence of scholarship on multicultural education (Gorski, 1999) heightened awareness of social justice issues such as power, privilege and economics. Indeed the 1980s were seen as the education decade, "not a decade in which specific education problems were solved, but a decade in which the problems of education as a whole achieved national significance" (Doyle, 1991, p. 185). It was the decade where public pressure was brought to bear to force educators to change practice and theory and to establish standards of performance. These calls for standards became the impetus for the decade of accountability in the 1990s especially in the assessment practices that evolved in Grades Kindergarten to 12 and which subsequently grew into the high stakes testing programs evident today.

A focus on standardized testing reported gender differences in achievement that promoted an increased interest in research on a variety of topics related to school-aged boys and girls. Understandably, research on boys' achievement in reading has more than doubled between the time periods 1945–1980 and 1981–2011. A comprehensive search through the major databases from 1981–2011 resulted in more than 2,000 hits compared to the 700 cited studies in the previous 30 years. Publications on gender and reading varied greatly over the past three decades including qualitative and quantitative methodologies ranging

from large-scale analyses to in-depth case studies. In a goal to make educational opportunities equitable, concern for girls' achievement in mathematics and science was a focus in gender studies in the 1990s which promulgated the growth in girls' only classes (Blair & Sanford, 1999). Mounting evidence over the past decade, shifted to a focus on boys. Publicized national and international studies indicate that boys are lagging behind girls in reading and writing competencies (Wilhelm, 2009) and have thus created a hype that boys are at risk of school failure. Although a number of researchers contend that boys are not underachieving in reading (Hogrebe, Nist & Newman, 1984; White, 2007), reports from large-scale investigations (Organization for Economic Co-operation and Development [OECD], 2001, 2004a, 2007, 2010) have been compelling—suggesting that girls' superiority in reading continues to grow without drawing equal attention to the groups of boys and girls doing well and not so well.

Concern for the underachievement of boys in schools is part of a larger issue of creating equitable and inclusive education. In the pursuit of accepting diversity as a fundamental value and an essential component in building a cohesive society, governments and school boards have created policies that foster safe and caring schools for all students around issues of gender, racism, religious intolerance, bullying, homophobia, and the inclusion of special needs students. It is the goal that all students in publicly-funded schools have an opportunity to reach their highest potential (Ontario Ministry of Education, 2009).

Research over the past 30 years on gender and reading has witnessed the interplay of theoretical perspectives. Investigations that focused on socio-cultural

and cognitive positions examined the "root cause" of boys' underachievement in reading, citing gender stereotyping and schema theory. This explanation was used in investigations on motivation, sex of the teacher, use of technology, and boy-friendly book choices (Sokal, Katz, Adkins, Gladu, et al., 2005, Sokal, Katz, Adkins, Grills, et al., 2005; Sokal, 2010; Sokal & Katz, 2008; Sokal, Thiem, Crampton, & Katz, 2009; Steiner, Steinen & Newman, 1981). Other investigations have taken a biological stance referring to brain research that claims that males and females have different brains. These studies support the need for boys'-only classes, a growing trend in recent years that address ways to use brain-based learning theory and strategies to engage boys' learning styles in the classroom (Basilo, 2008; Vrooman, 2009). It is interesting how this "nature-nurture" discussion around gender differences has recycled over the past 100 years or more, although at times temporarily abandoned; the topic resurfaces but within a slightly different context (Blakemore et al., 2009). The conundrum is that effective strategies that have been addressed to support boys are also good for all students. At the heart of the achievement gap is good quality teaching (Klinger, Shulha, & Wade-Woolley, 2009).

My goal in this chapter was to complete a comprehensive search through all the major databases on the topic of gender and reading during the time period 1981–2011. This exploration rendered approximately 110 potentially relevant works. A further extensive appraisal of the 110 works limited the search to 41 qualitative and quantitative investigations, ranging from large-scale analyses to in-depth case studies. After multiple readings of the investigations and analysis

of the findings, the studies were sorted into key constructs. Five major themes emerged from this process and are described next.

1. Nine studies indicate small, or no sex differences in reading achievement throughout the last three decades at the elementary, (Flynn & Rahbar 1994; Harper and Pelletier, 2008; Lummis & Stevenson, 1990; Phillips, Norris, Osmond, & Maynard, 2002; Quinlan 1996; Wargacki, 2008), middle (MacFarlane, 2001), and high school levels (Hogrebe et al., 1984; White, 2007). In addition, issues around how large-scale assessments can be misleading are addressed because of their relevance to the points made in many of the large-scale analyses (Hogrebe et al., 1984; White, 2007).
2. Claims that boys underachieve in reading have been reported in international studies including the Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA) and the National Assessment of Educational Progress (NAEP). Gender differences reported in PISA, PIRLS and NAEP are examined, and although the gender gap favouring girls has increased on PISA from 1992, other national and international tests (NAEP, PIRLS) do not reveal this trend. Ten studies are enclosed including one national (Rampey, Dion, & Donahue, 2009), seven international studies (Elley, 1992; Mullis, Martin, & Gonzalez, 2004; Mullis, Martin, Kennedy, & Foy, 2007; OECD, 2001, 2004a, 2007, 2010) as well as two analyses of the PISA data in this part.

3. Four studies focus on the validity of large-scale international tests. Three include a meta-analysis of the major international studies up to 2003 and indicate that two-thirds of the gender differences reported were due to the time period of the large-scale assessments and the scaling procedures (Keeves, Lietz, Gregory, & Darmawan, 2006; Lietz, 2006a, 2006b). Although effort was reported to account for a small increase in female superiority in reading, overall across the majority of countries, effort did not invalidate the PISA results (Butler & Adams, 2007).
4. Eight studies, including six longitudinal studies, identified that high performing boys have maintained their status in reading but low performing boys, particularly from low socio-economic groups and ethnic groups, are doing poorly (Becker & Forsyth, 1990; Entwisle, Alexander, & Olson, 2007; Husain & Millimet, 2009; Martin & Hoover, 1987; Matthews, Kizzie, Rowley, & Cortina, 2010; Robinson & Lubienski, 2011). Indices suggest that the gap is widening between high achieving and low achieving boys. In Canada, Willms (2004), and Edgerton, Peter, and Roberts (2008) report on regional disparities and explore a number of factors that contribute to low achievement, including SES.
5. Ten studies examine the cause for the reported phenomenon of boys' underachievement in reading by surveying boys from within and across countries. Studies explored the effects the gender of the teacher

has on boys' perception, motivation and reading performance and whether students view reading as a feminine activity. All of these studies point to an increase in achievement that was unrelated to the gender of the teacher (Sokal, 2010; Sokal & Katz, 2008; Sokal, Katz, Adkins, Gladu, et al., 2005, Sokal, Katz, Adkins, Grills, et al., 2005; Sokal et al., 2009; Steiner et al., 1981). In order to address boys' needs there has been an increase in the number of single-gender classes as of late, however case studies that investigated boys-only classes had mixed reviews (Basilo, 2008; Stotsky, Denny & Tschepikow, 2010; Vrooman, 2009). In collaboration with researchers from Queen's University, the Education Quality and Accountability Office (EQAO) completed an in-depth study in Ontario schools that compared low gender vs. high gender gap schools and found that when teachers expect that all students produce quality work, there was little difference in the gender gap (Klinger e al., 2009). They concluded that the underperformance of boys in reading is complex and not one that can be solved by simple solutions.

These five themes will be explored in detail in the following sections. The reported studies that provide evidence that boys and girls do not differ greatly in reading achievement will be examined first. The widely accepted perspective is that girls are superior to boys in language and literacy skills. Findings from studies that challenge this belief are documented next. These studies purport that girls and boys are more alike than different.

Theme I: Small or No Gender Differences

Widespread reports over the past decade have emphasized the underachievement of boys, and research that challenges these findings is often underreported. In this section, nine evidence-based research studies are analyzed. The findings from each investigation question the taken-for-granted position that girls are superior in reading achievement. Gender differences in reading achievement from K–12 are reviewed. The over-identification of boys in remedial reading classes is discussed, and how reading achievement changes for both boys and girls over the elementary school years is explored. Studies that challenge the methodology of previous large-scale international assessments and contrary findings from additional investigations are examined.

An example where early identification revealed important instructional information critical to meeting the needs of all children was completed by Harper and Pelletier (2008). Gender and early literacy was assessed using the *Test of Early Reading Ability (TERA)* (Reid, Hresko & Hammill, 1981), specifically the updated TERA-2 (Reid, Hresko & Hammill, 1989) and TERA-3 (Reid, Hresko & Hammill, 2001) with English first language students (L1) and English language learners (ELL). Although there were no significant differences between gender on the subtests that measured alphabet or conventions, differences were noted between L1 and ELL students on the meaning subtest of the TERA-3 ($p < .01$). Early detection of children's emergent literacy difficulties may provide educators with knowledge of groups of children that require specific reading programs. This research is an example where ELL is a major factor in early reading delays

and not gender.

How gender and chronological age at school entry impact future reading achievement is a recurring theme from previous decades. Lois Quinlan (1996) in her masters' thesis confirmed that chronological age and gender do not predict Grade 3 reading success. Her study followed 119 students who entered kindergarten in the Hillside School System in New Jersey between January and December 1986. Students were divided into three groups: late, medial and early entrance depending on the students' birthdays. The students were further divided by gender. Records on each student were gathered using the composite reading scores from the *Metropolitan Achievement Test* (no citation nor version used was provided). Results exhibited a low or negligible correlation between a child's age at the start of kindergarten and the child's overall reading achievement at the end of Grade 3. No significant differences were found in the mean scores between genders at each age group (early, medial, late).

Lisa MacFarlane (2001) in her masters' thesis also studied the relationship between early literacy experiences and reading achievement in middle school. The overall reading achievement of 103 randomly selected students, 53 boys and 50 girls, in Grades 6, 7, and 8 in a large suburban community were assessed using the *TerraNova* in April 1996 (No citation for the *TerraNova* or specific subtests used in the assessment battery was reported). No significant differences between the boys' and girls' reading achievement averages were found. Percentages derived from students' surveys (e.g., Do you like to read? On average how often

do you read for enjoyment?), revealed that although girls reported being more involved in literacy activities, the differences were not significant.

Lummis and Stevenson (1990) completed three cross-cultural studies (two reading and one mathematics, only reading is of relevance here) at the elementary level on gender differences in achievement and beliefs about gender and achievement. They examined students at kindergarten, Grade 1, and Grade 5 in three countries the United States, Taiwan, and Japan. The first study analyzed reading achievement of students in kindergarten. A total of 1,975 students were selected from 24 classes in three cities, Minneapolis-St. Paul (United States), Taipei (Taiwan), and Sendai (Japan). The reading assessment required that the child match and recognize letters, words, simple sentences, and meanings from sets of pictures.

The second study investigated the reading achievement of students in Grades 1 and 5. Two thousand, one hundred and eleven students from Grades 1 and 2, 155 students from Grade 5 were selected randomly from 10 schools in Minneapolis-St. Paul, Taipei, and Sendai. Since standardized tests of achievement were inadequate for such a cross-cultural study, "teams of bilingual and trilingual researchers . . . constructed batteries of achievement and cognitive tests that we believe are reliable, appropriate, and culturally unbiased" (Lummis & Stevenson, 1990, p. 255). No information was provided to indicate whether the tests were piloted before being administered. Reading vocabulary and reading comprehension tests were constructed in three identical versions from K to Grade 3 and translated into each respective language. After Grade 3, parallel versions

were constructed that were comparable in grammar, subject matter and vocabulary. Results from the reading test in the first and second study showed that girls in kindergarten and Grade 1 achieved higher scores in reading than boys (at the $p < 01$ level). However, by Grade 5 there were no significant differences in boys' and girls' reading scores.

Lummis and Stevenson (1990) also interviewed the mothers regarding their children's achievement in reading and mathematics. In addition, students were asked to rate their own achievement and attitudes toward reading and mathematics, and how they perceived their parents' happiness with their performance. Since it was not possible to interview all mothers and children, equal numbers of boys and girls and their mothers were selected at random from each class. Based on interviews, the researchers detected similar beliefs among the mothers in all three cultures (United States, Taiwan, Japan) regarding their children's abilities. Mothers believed that girls were better readers than boys and that boys outperformed girls in mathematics ($p < .001$, Table 7, p. 260). These biases were conveyed to their children who in turn also rated girls as better readers ($p < .001$ in Grade 1 and Grade 5, p. 260) and boys superior in mathematics ($p < .01$ in Grade 5, not significant in Grade 1, Table 6, p. 260). Lummis and Stevenson (1990) noted that beliefs regarding gender differences were greater than the actual achievement differences and these beliefs were more prevalent for reading than for mathematics. They conclude that if "girls continue to accept the superiority of boys in mathematics and boys continue to believe girls are better readers, gender differences in these two core subjects are likely to

widen as children progress through their academic careers" (p. 263).

Flynn and Rahbar (1994) discovered that more boys are in remedial reading classes when teacher criteria are used to identify students than when standardized reading tests are administered. Standardized tests (*Iowa Tests of Basic Skills*, *The California Achievement Test*, the *Stanford Achievement Test)* (citations or editions were not included) were administered at the end of Grades 1 and 3 to 708 children from 13 mid-western school districts in the United States. Students with reading failures were identified and divided into two groups: severe and moderate. The ratio of boys to girls in the severe reading disability group (below the 10^{th} percentile) was 1.4:1 at Grade 1 and 1.3:1 at Grade 3. There were equal numbers of boys and girls in the moderate reading disability group (between the 11^{th} to 30^{th} percentile). The reading results of students were not significant at either Grade 1 or Grade 3, "indicating an absence of a gender effect on test-based reading categories" (Flynn & Rahbar, 1994). A second group of Learning Disabled (LD) students were examined based on teacher criteria. There were twice as many boys as girls in these remedial classes. When gender differences in reading achievement were assessed for the entire Grade 3 populations (including the students with reading disabilities), the findings showed no significant differences between boys and girls. The high number of boys in programs for learning disability classes, when boys and girls performed equally on standardized tests, revealed the under-identification of girls and the over-identification of boys for reading support when teacher referrals are used. The authors stressed the need to use standardized tests in order to monitor reading failures that are equitable for

both boys and girls.

Since more boys than girls are enrolled in remedial reading programs in the early grades, Phillips et al., (2002) were interested in studying if reading achievement remained static or changed over the elementary grades. The reading achievement of 87 boys and 100 girls from a rural school district in Eastern Canada were studied. Based on achievement results of the *Gates-MacGinitie Reading Tests* (MacGinitie, Kamons, Kowalski, MacGinitie, & MacKay, 1980) students were grouped into three categories: below average, average, and above average, and tracked from Grades 1 to 6. In Grades 1 to 3 a greater percentage of boys were in the below average group, however after Grade 4 this percentage decreased. Although half of the students remained at the same level, half of the students from the above and below average group moved to the average group. Compared to the static results reported by Juel (1988), the researchers found a higher probability for children to raise and lower their reading achievements in elementary grades. About half the children assessed as below average in reading in Grade 1 raised their performance to an average standing by Grade 6. Some students who were reported as above average in reading in Grade 1, decreased their performance to an average standing by Grade 6. No student in the below average category in Grade 1 moved to the above average group by Grade 6. Similarly, no student in the above average category in Grade 1 moved to the below average group by Grade 6. The notion that reading achievement is largely immutable (Juel, 1988) was challenged by Phillips et al. (2002) and gives promise for children with poor reading achievement. This research provides hopeful news

not only for boys targeted as underachievers in reading in the primary grades, but also for girls struggling in reading but whom were overlooked. The researchers stressed the need for early identification and early reading intervention while continuing to challenge students who are proficient readers. Another important finding from this study was that no gender differences in reading were noted after Grade 4. "Performance at the end of fourth grade marked a qualitative difference in the comparison between boys' and girls' reading achievement. There were no systematic relationships between gender and reading category at the end of fourth grade" (Phillips et al., 2002, p. 5).

Jennine Wargacki (2008) in her master's thesis reported similar findings that indicate no gender differences by Grade 4 with a sample of 88 students (56 males and 32 females). Data from the 2007 *Ohio Achievement Test* (Ohio Department of Education, 2007) showed no differences on the overall reading performance (significant differences were not reported). However, Wargacki noted that the interpretation of average can be misleading. Average is determined by the sum of the total score divided by the number of participants. If there is a great range in the scores, (e.g., outliers falling outside the normal range) these results can affect the average. If one student had a very high or very low-test score, the average would be skewed. In Wargacki's study, although the overall average for girls and boys was similar, there was a greater range between the boys' scores than the girls'. Therefore, Wargacki concluded that when comparing two groups using only the average is not always the best indicator. Different interpretations from statistical data were also explored in the next two studies

from Hogrebe et al., (1985) and White (2007).

Hogrebe et al., (1985) contested the findings of large-scale studies, such as those conducted by Gates (1961) and the National Assessment for Educational Progress (NAEP, 1982). (The United States Department of Education funds the Institute of Education Sciences and the National Center for Educational Statistics [NCES] publishes the National Assessment of Educational Progress [NAEP] known as the Nation's Report Card.) They argued that the statistical significances reported from these large-scale studies are due to large sample sizes that have overemphasized differences. When the number of items answered correctly was analyzed, Hogrebe et al. (1985) noted that the differences between boys and girls were not large. To address why previous research has reached such conflicting conclusions, they conducted their own investigation using data from the *High School and Beyond* (HS&B) national survey (NCES, 1982). The NCES sponsored the National Opinion Research Centre to complete the longitudinal study known as the HS&B. From the HS&B database 48,040 students were assessed, 23,362 seniors (last year of high school) and 24,678 sophomores (a term used for a student in their second year of high school). Reading achievement scores were calculated separately and reported as a combined score using vocabulary and reading comprehension tests. The reading comprehension test consisted of short passages from which students were to answer a series of multiple-choice questions. Typical passages consisted of 140 words. Students were divided into three categories—those who scored one standard deviation above the mean, those who scored one standard deviation below, and those in the

middle. The results show that the magnitude of the gender difference accounted for less than 1% of the variance, with several instances where there were no statistically significant differences in the reading test scores. Although the senior and sophomore males had a higher mean score on reading comprehension than the females in both the middle and the one-standard-deviation-above groups, the difference was not significant (significant difference was noted only for males in the sophomore year in the middle group). In the group one standard deviation below, females had a higher average score than the males (significant difference for both the senior and sophomore year). The authors noted a number of limitations in their study. Although the study indicated a weak gender difference, it did not address differences in the processes and strategies used while reading. Passages were short and limited in number and thus may not have yielded an adequate sample of the skills needed to be successful in reading across the content areas in high school. In addition, the authors noted that the study was restricted to students in their final years of high school and that many students who were not successful in school may have already dropped out. In conclusion, Hogrebe et al. (1985) questioned the presence of a large gender gap in reading achievement at the high school level and suggested that researchers should put their energy and focus on successful reading processes and strategies that can be taught to either gender (p. 723). A key finding was the nature in which significant differences in large-scale studies are calculated and interpreted. Mathematically, small differences in a large sample will show up as significantly different, even when the actual differences in the raw scores are small. In contrast, the same difference

in a small sample will not show up as significant. The word significant to a statistician means that differences could not be explained through errors of measurement, sample size, or chance. However, the dictionary describes significant as large and meaningful. When a reader who is not familiar with statistics reads that a finding is significant it could be interpreted that the difference is noteworthy and substantial, when in actuality it could be a very small difference but from a large sample size. Thus, examination of the methodology is warranted before conclusions can be drawn.

Also critical of the methodology used to analyze large-scale assessment data, Bozena White (2007) completed a more recent master's study using the *Ontario Secondary School Literacy Test* (OSSLT) (no reference was provided), a compulsory reading and writing test for all Grade 10 students in Ontario schools. The EQAO is comprised of 100 questions and includes multiple-choice, short-answer, and short-answer requiring explanations. In order to attain a secondary school diploma, students must achieve a score of at least 60% on the test. She used the data from 2002. A total of 113,050 students were analyzed, 90,185 from the Academic English, and 22,865 from the Applied English levels. Three text types were assessed (informative, graphic, and narrative) as well as three reading skills (indirectly stated ideas and information, directly stated ideas and information, and connections between personal experiences and information) for a total of nine variables. Results from the test indicated that gender accounted for less than 1% of reading achievement. The overall mean for the students enrolled in the Academic stream was significantly higher ($p < .001$) than that of students in

the Applied program on all nine variables. When the reading performance of students within each stream was compared, White (2007) found small differences between boys and girls. Although the girls in the Academic program outperformed the boys on seven of the nine variables, the magnitude of these differences was close to zero ($0.02 < d < 0.10$). Similarly, small ranges ($0.11 < d < 0.13$) and close-to-zero ranges ($0.01 < d < 0.06$) were found between boys and girls in the Applied program. Reading skills assessed within each text type did not favour either boys or girls, which is in contrast to previous studies that indicated that boys perform better in documents and non-fiction, and girls perform better in narrative and prose (Elley, 1992; Gambell & Hunter, 2000). In addition, White (2007) conducted a series of separate analyses for each gender in order to compare the girls in the Academic program to the girls in the Applied program, and the boys in the Academic program to the boys in the Applied program. Effect sizes ranged from large ($0.73 < d < 1.0$) to very large ($1.0 < d < 1.14$) in both groups. The results revealed that the within gender differences are greater than the between group differences.

> The small effect size associated with gender (less than 1%) suggests that there is not a homogenous group of successful reading behaviours 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 gender specific strategies that have been recommended to remediate the purported gender

gap (White, 2007, p. 575).

Although there is a group of boys who are at-risk readers, similarly there is a group of girls who are also at risk. White (2007) concluded that the under-achievement of boys has been overstated and to some extent misrepresented.

The findings from the nine evidence-based investigations demonstrate the absence of a gender gap in reading achievement. No significant differences in reading achievement were found by the time students reached Division II (end of Grades 3, 4, or 5) (Flynn & Rahbar 1994; Lummis & Stevenson, 1990; Phillips et al., 2002; Quinlan, 1996; Wargacki, 2008). Moreover, no gender differences were found in the studies that examined the reading achievement of students in middle school (MacFarlane, 2001), and high school (Hogrebe et al., 1984; White, 2007). Lummis and Stevenson (1990) found that regardless of the finding that boys and girls read equally well, biases exist that have been perpetuated from parent to child, namely that boys are better in mathematics and girls in reading. Flynn and Rahbar (1994) found that when standardized tests were used to identify reading disabilities, no significant differences in gender were detected. However, when teacher judgment was used there was a clear bias against boys, and girls with reading problems. Phillips et al. (2002) challenged the notion that reading achievement does not change during the elementary grades and stressed the need for early identification of reading difficulties. Harper and Pelletier (2008) also found no gender differences in early literacy but found significant differences between ELL and L1 students when students were assessed on meaning. Three studies challenged the statistical analyses from previous studies (Hogrebe et al.,

1984; Wargacki, 2008; White, 2007) and found contrary findings indicating that gender differences accounted for less than 1% of reading achievement (Hogrebe et al., White, 2007). Clearly, from these nine studies the case for boys underachieving in reading is weak and misleading.

Theme II: Gender Differences in Reading Achievement

Competency in literacy has been measured by a number of national and international studies over the last 30 years. Variation in student performance has been reported both between, and within, nations. Among the observed gaps in student performance are gender differences in reading achievement. Some research groups that report a persistent gender gap favouring girls in reading achievement are detailed in the next section.

One collaborative assessment effort that measures and compares international results in student achievement is the Programme for International Student Assessment (PISA). PISA grew out of the Organisation for Economic Co-operation and Development (OECD). The major aim of PISA is to measure and compare learning outcomes for educational policy purposes (OECD, 2001, p. 17). PISA assesses performance in reading literacy, mathematics literacy, and science literacy on a 3-year cycle. Every 3 years a detailed analysis is provided in one of the three domains, with a minor focus on the other two. Reading achievement was the major domain of focus in 2000, mathematics in 2003, and science in 2006, with reading revisited again in 2009.

> The OECD's Programme for International Student Assessment (PISA) is a collaborative effort among the Member countries of the OECD to measure

> how well young adults, at age 15 and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's knowledge societies. (OECD, 2001, p. 14)

Students who are 15 years of age range from Grades 9 to 11, however, most are in Grade 10. Gender differences were noted in reading achievement reports published by OECD in 2001, 2004b, 2007, and 2010.

Thirty-two nations participated in PISA 2000, including Canada, the United States, and Mexico. Due to worldwide interest, the study was extended in 2001 to include 11 additional non-OECD countries for a total of 43 countries (28 OECD nations and 15 non-OECD countries) and approximately 315,000 students (OECD, 2004a, p. 4). PISA 2000 reported on a wide range of factors such as individual student variables (reading performance, engagement, home background, gender differences), school related attributes (school climate and resources, infrastructure, discipline, teacher commitment), and features related to the structure of school systems (school groupings, school autonomy). Differences in the PISA reports are labeled as statistically significant at the .05 level, denoting that 95 out of 100 replications would give the same results on the same population (OECD, 2001, p. 51, p. 237). Unless otherwise stated, when the word significant is used to describe differences it is understood that it is at the .05 level. This level of significance applies to the PISA 2000, 2003, 2006, and 2009 results.

The PISA reading assessment was designed to have an overall mean of 500 points and a standard deviation of 100. The maximum score a student could achieve on the test was 1,000 points. PISA 2000 used a mixed model (item

response and population model) to scale international, national and student scores (Adams & Wu, 2002). In order to better understand the test scores, the reading scale was divided into cut points and students' overall averages were converted into proficiency levels. Proficiency levels were placed on a continuum of reading skills ranging from Level 1(very basic ability to read and understand a simple text) to Level 5 (competencies in information management in unfamiliar texts, inferences, and critical evaluation of information). Three types of reading skills were evaluated: *retrieving information, interpreting texts, and reflecting on texts.* Within the 43 participating countries, two-thirds of the students averaged between 400 and 600 points out of a total possible score of 1,000, and the remainder scored either higher or lower. From the overall results performed above the OECD average (534 points), the United States within the OECD average (504 points), and Mexico below (422 points) (OECD, 2003).

A number of student characteristics contributed to the range in reading scores. The strongest factor was socio-economic (contributing to more than an 80-point difference and accounting for 20% of the variance), followed in order by factors such as: arrives on time for school, interested in reading, confidence in learning ability, controls own learning, being female, and high sense of belonging (OCED, 2004a, pp. 6–7). All OECD countries report that they seek to reduce educational disparities among students. In PISA 2000, equity in achievement between boys and girls was one of the many factors that was analyzed in detail.

Overall, girls scored on average 32 points higher than boys in reading on PISA 2000. In contrast, the boys scored significantly higher than the girls in

about half of the participating countries on mathematics (11 point difference in favour of males) with no difference in science (OCED, 2001, p. 124). Although there is a disparity between boys and girls in reading ranges, on average more girls read at higher levels (OECD average for girls performing at Level 5 was 11.9% and for boys 7.2%) (OCED, 2003, pp. 319–320). On the PISA results, the weaker readers on average were boys. Fifteen-year-old boys, on average, are 1.7 times more likely than 15-year-old girls to perform at a Level 1 reading proficiency or less (OECD, 2003, p. 320).

On the PISA 2000 survey, 45% of the females reported that they spent more than 30 minutes a day reading compared to 30% of the males who reported that amount of time (OECD, 2001, p. 131). Also noted on the survey was that males and females read different kinds and lengths of material. Females reported that they prefer to read fiction (37% of the females, 19% of the males) and males prefer to read the newspaper (68% of males vs. 60% of females), comics (35% of the males compared to 24% of females), and emails and web pages (50% males and 40% females) (OECD, 2003, p. 155). It is important to note that within each group on each reported reading measure and genre, significant numbers of both males and females reported reading less than those in the highest percentages reported.

Gender differences were also reported for self-concept. When rating their self-confidence in reading and mathematics, the girls rated themselves higher in language arts (0.15) than the boys (-0.14), and lower in mathematics (-0.13) than the boys (0.12). These self-confidence ratings are reported to be strongly related

to success in learning. These differences in self-concept tend to mirror the boys' and girls' overall performance (OECD, 2001, p. 133). Why do boys have such low self-concepts in reading and girls in mathematics? What home and school factors are needed to raise the self-concept of boys in reading and girls in mathematics?

Results from the report on school characteristics indicated that the socio-economic composition of schools was an important factor in overall student performance. It is not surprising that schools with more resources and better student performance tended to have students from higher socio-economic backgrounds. The challenge for all schools is to narrow the gap between student achievement and to accommodate for ranges in academic levels. Students from integrated schools in heterogeneous groupings performed better than students who were faced with streaming at a young age. In addition, countries where schools were segregated according to socio-economic status did not perform as well (OECD, 2005).

Douglas Willms is the author of a number of articles pertaining to youth literacy and assessment of national reforms. He is a professor and director at the Canadian Research Institute for Social Policy at the University of New Brunswick, (Canadian Research Institute for Social Policy website http://www.unb.ca/crisp/willms_cv.html, retrieved December 5, 2011). Under the direction of the participating PISA countries, Willms completed comprehensive analyses of student-reported responses on engagement from the PISA 2000 results (2003). Several main findings emerged. Willms concluded that over half of all

students in school reported being engaged in learning and have strong or average literacy skills. The remaining half fell into three groups. About 20% of the total population of students have a low sense of engagement but maintain high literacy levels. A second group, (about 10% of all students), demonstrated below average literacy skills due to regular absenteeism from school. These students scored on average 50 points below the OECD average. The third group consisted of students with poor literacy skills who tended to come from families with low SES, but reported average levels of engagement and attendance (Willms, 2003, pp. 33–34).

Willms (2003) also examined the relationship between student engagement and school factors. From students' ratings of their feelings of acceptance at school, three relevant findings surfaced. On average, there were no differences between how boys and girls rated their sense of belonging at school. However, foreign-born students were more likely than other youths to have a low sense of belonging in school. In addition, students from low socio-economic families were more likely to be disaffected by school, as were students who attended schools with a high percentage of students from low SES (Willms, 2003). "Students are more likely to be engaged at school if they attend schools that have a high average socio-economic status, a strong disciplinary climate, good student-teacher relations and high expectations for student success" (Willms, 2003, p. 48). Willms stressed that the results do not support the popular belief that an increase in student engagement will lead to an increase in literacy achievement. Disengaged students exist in every school, and even some who are

performing well in school. He concluded that students who lack a sense of belonging in school need a different intervention than those who are often absent from school (p. 34). However, students most at risk for underachieving in literacy are those from low SES attending schools and predominantly serving low SES students. Willms (2003) described this phenomenon as a *double jeopardy* because these students have two factors working against them, being both poor themselves and going to a school with other students who are also poor (p. 48). In PISA 2000, although gender differences were noted, socio-economic conditions remained the largest factor contributing to more than an 80-point difference and accounting for 20% of the variance (OCED, 2004a, pp. 6–7) in reading differences.

PISA 2003 assessed over a quarter of a million students representing 23 million 15-year-olds in 41 countries, (30 OECD countries and 11 non-OECD countries) (OECD, 2004b, p. 24). Since the statistics reported in PISA 2003 represent estimates of national performances and are not calculated on every student on every question, unless specifically indicated the confidence errors lie within the 95% level of chance (OECD, 2004b, p. 58). Scores were scaled based on a Balanced Repeated Replication (BRR) model and Fay's method (OECD, 2004b). This statistical technique estimates discrepancy in the data based on differences between full samples and half samples. Schools are paired according to size and demographics and a set of estimated scores are calculated. Half samples from the total population between the two (or more) schools are weighted and analyzed. This process is repeatedly done at two levels, the school and

student levels. Scores are then calculated based on the repeated analyses to make the total score for a country (Adams, 2005, p. 114). The major focus of PISA 2003 was on mathematics, however reading, science and problem solving were also appraised which allowed researchers to compare the results between 2000 and 2003 and to evaluate changes over time in student knowledge and skills (OECD, 2004b, p. 24). The same framework used in PISA 2000 was applied in PISA 2003. Reading was assessed on five levels of proficiency based on different kinds of texts and reading tasks. Rather than report on all three tasks separately, as was done in 2000, PISA 2003 combined the results into one single report (reading was not the main focus for PISA 2003). The average OECD reading score was 494 points—similar to PISA 2000 (average 500 points). Some countries increased their reading performance while others experienced decreases. Results between PISA 2000 and PISA 2003 showed no significant differences in performance for Canada and the United States. Canada remained above the OECD average (528 points) and the United States (495 points) within the OECD average. Mexico however, showed a drop between PISA 2000 (422 points) and PISA 2003 (400 points) significant at the 99% confidence level (OECD, 2004b, p. 282).

In all countries, except Liechtenstein, the girls had a significantly higher average performance in reading by 34 points overall. This difference in reading scores was similar to the gender gap found in PISA 2000 (32 points) (See Table 7.1). The magnitude of the difference between boys' and girls' reading scores ranged from 58 points in Iceland to less than a 20-point gap in Korea and Macao-

China. Canada and the United States had an average of 32 points difference in reading scores between the boys and girls, and Mexico reported a 21-point gap.

In most countries, there was a widening in the distribution between the scores, with males showing a greater likelihood of being amongst the lowest performers (Level 1 and below). The average OECD ratio of girls to boys performing at and below Level 1 was 1:1.8: In Canada this ratio was 1:2.5, 1:1.2 in Mexico and 1:1.7 in the United States (OECD, 2004b, p. 446). However, the extent of within country disparities in reading performance reflected a wide range in student abilities for both boys and girls in reading. Results from PISA 2000 and 2003 revealed the persistent number of students who achieved at basic reading levels. Students who achieved below 400 points (at and below Level 1) on the PISA 2003 averaged 17.3% across OECD countries, (8.4% in Canada, 17.5% in the United States and 49% in Mexico) (OECD, 2004b, p. 446). These results mean that 1:12 students aged 15 years of age in Canada, 1:2 in Mexico, 1:6 in the United States, including boys and girls, struggled to acquire knowledge and skills to further their learning. The challenge for all countries, as well as those who have high results overall, (such as Canada) is to address low literacy levels for both boys and girls. Rather than focusing on gender differences, the issue is low literacy regardless of the sex of the student. These findings have implications for policy-makers in education who seek equitable learning opportunities for all students (OECD, 2004b, p. 298).

The focus in PISA 2006 was on science although data were collected on reading and mathematics. Around 400,000 15-year-olds from 57 nations around

the world were randomly selected for this study (OECD 2007, p. 19). The OECD average on reading was 492 points out of a total possible 1,000 points. Again, PISA 2006 used the BBR model for scaling international, national and student scores (Adams, 2009). Canada's performance was reported by country and by provinces. All participating provinces in Canada performed at or above the OECD average (527 points) and Canada ranked third behind Korea (556) and Finland (547). The reading literacy results were not reported for the United States because of an error in printing the test booklets. Mexico scored below the OECD average for a total of 410 points. Even though a number of countries experienced increases from 2000 to 2006 (in descending order, Chile, Korea, Poland, Liechtenstein, Indonesia, Latvia, and Hong Kong-China), a number of countries had a decline in their reading scores (in descending order, Spain, Japan, Iceland, Norway, Italy, France, Australia, Greece, Mexico, Argentina, Romania, Bulgaria, the Russian Federation, and Thailand) (OECD, 2007, p. 301). Again, in all of the OECD countries, females performed significantly better ($p < .05$) on average by 38 points than the males. Canada's gender gap remained the same between PISA 2000, 2003 and 2006 at 32 points while Mexico's gender gap widened from 20 points in PISA 2000 to 21 points in PISA 2003, and to 34 points in PISA 2006 (see Table 7.1). Gender differences were summarized in a special report, *Equally prepared for life? How 15-year-old boys and girls perform in school* (OECD, 2009). A précis of the major findings follow.

Trends in gender differences in reading have been reported from PISA 2000 to PISA 2006. The average reading performance of all OECD countries

declined by 6 points, showing a slight but not significant negative difference in achievement. "This increased difference is largely due to the fact that between 2000 and 2006 the performance of males decreased (statistically significant) by 10 score points" (OECD, 2009, p. 17) (significant difference at the .05 level). The proportion of boys and girls performing at the highest level (Level 5) has remained similar between PISA 2000 (7% of males and 12% of females) and PISA 2006 results (6% of males and 11% of females). The number of boys in the lower proficiency level (Level 1 and below) has increased (from 22% in PISA 2000 to 26% in PISA 2006), while the girls have remained relatively stable (13% in PISA 2000 and 14% in PISA 2006) (OECD, 2009, p. 17). This increase in the number of boys in the lower levels of reading (Level 1 and below) has contributed to the overall lower OECD reading average by 6 points.

Table 7.1

Gender Differences Favouring Female Performance in Reading from PISA Results (2000–2009) for Canada, Mexico, and the United States

	PISA 2000 *	PISA 2003 *	PISA 2006 *	PISA 2009*
OCED average	32 points	34 points	38 points	39 points
Canada	32 points	32 points	32 points	34 points
Mexico	20 points	21 points	34 points	25 points
United States	29 points	32 points	no results reported	25 points

(* results are all statistically significant at the .05 level)

These reading results are an anomaly when gender differences are examined in relation to the results on mathematics and science. There were no

changes in gender differences in mathematics between PISA 2003 and PISA 2006. The boys performed better (by 11 points) than the girls on both PISA 2003 and 2006. The gender gap favouring girls on reading continued to increase (32 points in 2003, 34 points in 2003, and 38 points in 2006) (OECD, 2009). In addition, PISA 2003 developed a problem-solving assessment to measure students' cross-disciplinary skills independent of mathematics and science. The intent of the assessment was to examine whether boys perform better than girls in mathematics because they have better knowledge or have better problem solving skills that help them solve mathematical problems. The assessment was designed with an OECD average of 500 points, a standard deviation of 100 and three proficiency levels. "Students were required to identify problems in various settings, choose relevant information or constraints, represent possible alternatives or solution paths, develop solution strategies, solve the problem and communicate the solutions" (OECD, 2009, p. 21). Small gender differences indicated that neither sex had superiority in analytical reasoning. "On average in OECD countries, 18% of male students and 16% of female students were below Level 1, while 19% of male students and 18% of female students reached Level 3" (OECD, 2009, p. 21). It does seem perplexing then that boys would underachieve in reading, since reading involves problem solving and analytical reasoning and since both science and mathematics require sophisticated reading skills. The critical question here is, what is it about the PISA reading assessment that is showing such disparities worldwide between boys and girls?

When gender differences in the overall science results are examined in detail, small differences were noted in the PISA 2006 report. An analysis of the field-tested questions revealed that some questions favoured boys, while others favoured girls. In order to generate an unbiased test, a balance of questions that favoured both boys and girls were included (OECD, 2009, p. 24). The fact that the PISA 2006 science test was carefully designed to yield a fair test raises the question about how the reading test was field-tested and designed? Gender differences attributable to the design of the large-scale assessments were examined in the study completed by Petra Lietz (2006a, 2006b). Dr. Petra Lietz presently works for the Australian Center for Educational Research (ACER) as a Senior Research Fellow (personal communications, December 6, 2011). When she was an assistant professor of quantitative research methods at the International University of Bremen in Germany, she conducted a meta-analysis using a systematic approach to assessing research findings in order to determine the contradictory evidence around the topic of gender differences in reading. Her analyses of various cross-national assessments revealed that gender differences in reading after 1992 are related to the design of the test (2006a, p. 140). A synthesis of this study is reported in the next section.

PISA 2009 assessed reading as a major focus for a second time in a decade. It included for the first time digital texts, as well as continuous, and non-continuous texts. Students were assessed on how they reflected on, and evaluated what they read. The proficiency levels were extended to six from five levels reported in PISA 2000, 2003 and 2006 included a new set of reading items for

more basic reading levels (divided level one into two, 1a and 1b). New questions were added to determine students' engagement in reading activities including the techniques they used to learn, and how they used libraries, the internet and new technologies. In all, 470,000 students representing 26 million 15-year-olds in 65 countries and economic regions (some countries report their findings by economic regions rather than by country, for example, Hong Kong-China) including 34 OECD countries and 31 partners participated in PISA 2009. A second round was conducted in 2010 with an additional 50,000 students in 10 countries and economic regions for a total of 75. Overall, Canada scored above the OECD 500 average (524 points), the United States within the OECD average, (500 points) and Mexico scored below the OECD average (425 points) revealing similar results and trends in reading as PISA 2000 (OECD, 2010, p 15). Similar to PISA 2000, international, national and student scores were scaled using a mixed model (from item response and population model) (OECD, 2010, Chapter 9, p. 4). On average, across all OECD countries, the girls scored better than the boys by 39 points, an increase from 2000 by 7 points showing a widening in the gender gap. The girls scored significantly higher than the boys ($p < .05$ level) in Canada, Mexico, and the United States. Although the gender gap did not change between PISA 2000 and PISA 2009 for the United States, in Canada and Mexico there was a slight increase (2-point increase for Canada and a 5-point increase for Mexico) (refer to Table 7.2). The reading proficiency levels, across OECD countries on average, revealed that about half as many girls as boys scored below Level 2. Additionally, about twice as many girls as boys scored in the upper level of

reading (Levels 5). Gender differences were also noted in the different aspects of reading. When the combined reading score as well as the three subscales (*access and retrieve, integrate and interpret, and reflect and evaluate*) were examined gender differences were noted. The girls significantly ($p < .05$) outperformed the boys on the overall reading scores as well as on the three subscales (with the exception of Columbia on the one subscale *access and retrieve*) in every country and economic region. Smaller variations were noted on the *integrate and interpret* subscale (with girls on average scoring 36 points higher than the boys) with a larger disparity detected in the *reflect and evaluate* subscale (with girls scoring on average 44 points higher than the boys) (OECD, 2010, p. 71). Comparisons were also made on the text-formats subscales. Again, the girls scored significantly better ($p < .05$) than the boys on the *continuous* subscale (OECD average of 42 points). However, the gap narrowed slightly between boys and girls on the *non-continuous* subscale (OECD average of 36 points) (OECD, 2010, p. 88). The fact that boys do better on non-continuous text may be associated with the kinds of reading materials that boys and girls prefer to read. Even though a large number of boys and girls reported that they do not read for pleasure at all, the girls that do, prefer to read longer texts while the boys report that they spend more time reading newspapers and comics (OECD, 2010, p. 90).

Differences in reading preferences raise the question of test characteristics and how the framework of the PISA design may affect the disparity noted in the reading test. The same trends in gender differences have been reported since PISA 2000. Although the design of the science literacy test in PISA 2006 was

field tested to balance questions that favoured boys and girls, the reading assessment has purposefully kept the same framework in order to measure trends over time. "Any major change in the distribution of item types in print reading might also impact on the measurement of trends" (OECD, 2010, p. 46). Although studies have suggested that the response format of the PISA reading assessment significantly favours girls (Lafontaine & Monseur, 2009), to ensure a valid assessment, PISA 2009 used the same framework as previous assessments (PISA 2000, 2003, 2006) (OECD, 2010, p. 46). It stands to reason then, if the reading test is designed to favour girls, a reported gender-gap will continue to be noted. Lafontaine and Monseur (2009) completed a study on the gender gap increase between PISA 2000 and the previous IEA reading comprehension assessment completed by Elley in 1992. The fact that there was a small gender gap in 1992 and such a large one within a decade is startling. This finding was also documented in Lietz's (2006b) meta-analysis "the gender gap in favour of girls is even more pronounced for the assessment programs that have been conducted since 1992" (p. 140). Over the past decade, PISA results have indicated that there is a significant and growing gender gap in reading. These findings continue to perpetuate the spurious belief that boys are in trouble in reading and the self-fulfilling prophecy that boys are underachieving. The results call into question the continued use of a test that favours girls and thus continues to bolster the differences in reading achievement. Findings from Lafontaine and Monseur's study (2009) revealed that differences between 1992 results and the present PISA results are the result of a change in the PISA framework. Boys perform better on

multiple-choice questions and poorer on open-ended questions and prefer non-continuous rather than continuous texts. However, PISA purports the importance of balancing multiple-choice and open-ended responses where the quality of the responses is measured rather than the conclusion (OECD, 2010, p. 46). Although Lafontaine and Monseur (2009) see the advantages in studying trends over time, they caution test developers of the PISA reading assessment to

> be careful to guarantee a similar balance of the various components of the reading framework in successive assessments; otherwise, the validity of the trends indicator is likely to be jeopardized. If, for instance, the PISA 2009 reading were to include more "reflect" items or, especially, more continuous texts or more open-ended items, it would clearly result in an increased gender gap compared to the PISA 2000 assessment. Of course, some powerful considerations or arguments could or should lead to revisions of the framework. In conclusion, we want to stress the importance of clearly arbitrating the advantages and disadvantages of changes and adaptations between successive assessments. (p. 77)

Although updates were made to the PISA 2009 reading assessment by extending the reading proficiency levels, including digital texts, and elaborating on the survey to understand reading preferences and student engagement, no revisions were made to the response format (OECD, 2010, pp. 38–48), even though there have been "some powerful consideration or argument . . . (that) should lead to revision of the framework" (Lafontaine & Monseur, 2009, p. 77). It is apparent that the gender gap reported in the PISA data from 2000 to 2009 is

not as straightforward as it appears. Caution is warranted in how the gap is identified, interpreted, and the claims made.

The International Association for Evaluation of Education Achievement (IEA) also conducts cross-national comparisons in reading achievement. The IEA is an independent international cooperative of national research institutions and government agencies (Mullis et al., 2007, p. 16). Following the initial project that was carried out by Thorndike (1973), (refer to Chapter 6 for more details), the IEA conducted a second major survey of reading achievement under the direction of Warwick Elley (1992) from the University of Canterbury in Christchurch, New Zealand. Elley's study reported on 32 countries from all continents of the globe for two populations, 9-year-olds (Population A) and 14-year-olds (Population B). However, not all countries submitted data for both populations. Elley's 1992 study was the precursor to the present day *Progress in International Reading Literacy Study* (PIRLS). Results from Elley's study revealed that amongst 9-year-olds, the girls' overall reading achievement average surpassed the boys' in all countries. The differences were statistically significant in 19 countries ($p < .05$) but were small and not significant for eight nations. Gender comparisons between 14-year-olds, indicated that the girls achieved significantly higher reading scores in 11 countries ($p < .05$), were not significant in 15 countries, 2 countries reported no differences, and 3 countries had results that showed that boys out-performed girls in reading, (of which two scores were significant) ($p < .05$ level). As Elley reported, "the gender gap, however, is not uniform and the differences among countries deserve exploration" (p. 104). Elley hypothesized that differences

between boys' and girls' reading achievement were possibly related to the gender of the teacher, although the data did not support this suggestion. In addition, a reading readiness hypothesis was proposed but not verified. Overall, the results of Elley's study indicated stronger support for girls' superior performance amongst 9-year-olds, but were not conclusive amongst 14-year-olds.

Unfortunately, many researchers and others cite Elley's 1992 IEA study as support for the claim that girls outperform boys in reading (This study has been cited in 452 related articles according to Google Scholar, December 3, 2011). However, it is a false claim to use Elley's findings as evidence that boys are underachieving in reading. Elley established that although girls achieved higher levels in all countries, the differences were not very large: "There are very good male readers in every country, and many boys achieve well above the average girl in every country" (1992, p. 57). Elley's study shows no conclusive evidence to support the claim that boys are underachieving in reading. Therefore, when researchers and others quote Elley's findings to support their claims of the underachievement of boys, they are misrepresenting the facts with concomitant serious implications for both boys and girls.

From the initial reading surveys conducted by Thorndike (1973) and Elley (1992), further developments in reading assessments were made by the IEA. The IEA formed PIRLS, and in 2001 initiated a cyclical approach to reading assessment in order to measure reading trends over time. Every 5 years, students in their fourth year of schooling are assessed in reading (around 9 to 10 years of age). Countries vary in the age at which students enter school, so the age at which

students write the test differs. The fourth year of formal training was targeted, as this is about the time that students make the transition from learning to read to reading to learn (Mullis, Martin, Gonzalez, & Kennedy, 2004). PIRLS has completed 3 cycles of international assessments: 2001, 2006, and 2011. The overall results from the 2001 and 2006 studies are summarized next. The findings from the 2011 study are not reported because the results have not been released at this point.

PIRLS 2001 (Mullis, I., Martin, M., & Gonzalez, E., 2004) measured reading processes of literary and informational texts as well as reading behaviours and attitudes of students at the fourth grade in 35 countries. Four major reading processes were assessed: 1) retrieval of explicitly stated information, 2) straightforward inferences that required a student to fill in the gap based on information that is contained in the text, 3) interpret and integrate ideas and information beyond the text, and 4) examine and evaluate content, language, and textual elements. After reading the passages, students were expected to answer multiple-choice questions or construct written responses. Questionnaires were given to students, parents, teachers, and school principals to determine whether factors from home and school influence students' reading habits and attitudes. The PIRLS scales were developed using the Item Response Theory (IRT), a psychometric method based on item scores and a mathematical relationship between the response given and each item in the test. Nine different booklets were developed and the IRT method was adopted to equate students' scores since students responded to different passages depending on the test booklet they

received. The "IRT methodology produces a score by averaging the responses of each student to the items that he or she took in a way that takes into account the difficulty and discriminating power of each item" (Mullis, Martin, et al, 2004, p. 10). A common scale was developed from which all countries were compared. Since countries vary in size, each are weighted to develop the average and standard deviation. The test was designed for an average score of 500 and a standard deviation of 100 (Mullis, Martin, et al, 2004, p. 10). Although the exact number of students who participated in PIRLS 2001 was not provided, each country agreed to have a sample size of at least 3,700 students (Mullis, Martin, et al., 2004, p. 3). Achievement differences across countries are reported as an average or percentage. The standard error or measure of uncertainty was calculated to a 95% confidence level ($p < .05$) (Mullis, Martin, et al., 2004, p. 183).

PIRLS 2001 included data from Canada (data collected from Ontario and Quebec only), and the United States, but not Mexico (Mullis, Martin, et al., 2004). Both Canada and the United States achieved an average score higher than the international average (500), with Canada placing 6th overall (mean score of 544) and the United States 9th (mean score of 542 points). Gender differences were noted. In all countries the girls had significantly higher achievement in the order of 20 points ($p < .05$) than the boys (average score for the girls was 510 and 490 for the boys). In Canada, the average score for the girls was 553 and the boys 536, a 17-point difference. In the United States the average score for the girls was 551 and 533 for the boys, an 18-point difference (Mullis, Martin et al., 2004, p.

9). Girls also performed better than the boys on both literary and informational text. The international average revealed that on literary passages, the girls scored 511 points and the boys 490, a difference of 21 points ($p < .05$). In Canada, the girls scored 554 points and the boys 535, a 19-point difference ($p < .05$). In the United States, the girls scored 558 points and boys 542 points, a 16-point difference ($p < .05$). On informational texts the international average was 500 points with girls scoring on average 509 points and boys 491 ($p < .05$). In Canada the girls scored 549 points and boys 534 points that marked a 15-point difference ($p < .05$). In the United States on informational text the girls scored 541 points and the boys 525 points, a 16-point difference ($p < .05$). The girls consistently did better than the boys on both the literary and information passages (Mullis, Martin et al., 2004, p. 14) (Refer to Table 7.2).

Mullis et al. (2004) extended the analyses presented in the *PIRLS 2001 International Report* and completed an additional investigation into the four major reading processes assessed by PIRLS 2001. The four reading processes were grouped into two categories. Girls significantly performed better ($p < .05$) than the boys on both categories: 1) retrieve and straightforward inference, and 2) interpreting, integrating and evaluating processes. Internationally, the girls exceeded the boys on the retrieval and straightforward inference by 18 points (average for the girls was 509 points and the boys 491 points). For Canada, the girls scored 544 points and the boys 529 points a 15-point difference ($p < .05$). In the United States the girls scored 545 points and the boys 526 points, a 19-point difference ($p < .05$).

Table 7.2

Gender Differences Favouring Female Performance in Reading from PIRLS 2001 Results for Canada and the United States

Variables	Can. Girls' points	Can. Boys' points	Point diff.	U.S. Girls' points	U.S. Boys' points	Point diff.
Overall Average	553	536	17	551	533	18
Types of Text: Literary	554	535	19	558	542	16
Types of Text: Information	549	534	15	541	525	16
Reading Processes: retrieve & straightforward inference	544	529	16	545	526	19
Reading Processes: interpreting, integrating and evaluating	558	540	18	557	539	18

On the higher level thinking processes of interpreting, integrating and evaluating, internationally there was a 20-point difference in favour of the girls (average score of the girls was 510 and 490 for the boys). In Canada, the average score of the girls was 558 points and 540 for the boys, an 18-point difference ($p < .05$). A similar 18-point ($p < .05$) difference was noted in the results from the United States (an average of 557 for the girls and 539 for the boys) (Mullis, Martin et al., 2004, p. 28). Interestingly, in both Canada and the United States, overall the boys surpassed the international average by more than 30 points, indicating that the boys are not underachieving comparatively speaking internationally. Refer to Table 7.2 for specifics.

PIRLS 2006 recorded the results from 45 countries and jurisdictions, including 38 countries, 5 provinces from Canada, and two languages from Belgium: Flemish and French. Canada chose to report by province, as well as by country. The provinces included: Alberta, British Columbia, Ontario, Nova Scotia, and Quebec and these combined accounted for 88% of Canada's population. The average scores on the PIRLS achievement are reported for each of the 45 participants including standard error, and a 95 % confidence level ($p < .05$) (Mullis et al., 2007, p. 36). Differences in the overall results between 2001 and 2006 were noted for the countries that participated in both studies. No significant differences ($p < .05$) in the overall reading achievement were reported for Canada (Ontario +6 points, Quebec -4 points) and the United States (-2 points) between 2001 and 2006 (Mullis et al., 2007, p. 44). Girls had statistically higher scores ($p < .05$) in every country and province, with the exception of Spain and Luxembourg. Internationally, the difference between the girls' and boys' scores on average was 17 points ($p < .05$). Canada (with the exception of Nova Scotia), and the United States were below this international average of 17 points (although all scores were statistically different at ($p < .05$). In increasing order, Alberta showed the smallest difference in average reading achievement scores between girls and boys of 8 points (favouring girls), followed by British Columbia with 9 points, United States, 10 points, Quebec and Ontario 13 points each, and Nova Scotia 21 points (Mullis et al., 2007, p. 48) (Refer to Table 7.3). This drop in difference between 2001 and 2006 is considerable suggesting that the boys are closing the gap.

Amongst the 45 participating countries and jurisdictions, on average, girls reported spending more time during a day reading books and magazines (1.5 hours) than boys reported (1.3 hours). In reality this works out to be a difference of 12 minutes more a day for the girls (no significant differences were reported). Boys report spending slightly more time reading for information on the internet than the girls (1.0 hours per day compared to 0.9), about 6 minutes more (Mullis et al., 2007, p. 153). Relatively speaking within the 45 participating countries and jurisdictions boys and girls in Grade 4 spent the same amount of time reading outside of school.

The average achievement in reading for literary and informational purposes was also reported by gender. The international average revealed that the girls outperformed the boys in both literary and informational text by 17 and 16 points respectively. However, the differences between the girls' and boys' scores in Canada and the United States were not as large as the international averages. For literary text, differences favouring the girls ($p < .05$) include the following average score differences: Alberta (11), British Columbia, Ontario, Quebec, and the United States (12), and Nova Scotia (18). Even smaller gender differences were noted on informational texts ($p < .05$) including average score differences in the following: British Columbia (6), Alberta (7). United States (9), Ontario and Quebec (11), Nova Scotia (20) (Mullis et al., 2007, p. 56) (Refer to Table 7.3).

Results from the reading processing revealed that girls performed significantly better ($p < .05$) than the boys in the retrieving and straightforward inferencing processes (15 points) and the interpreting, integrating and evaluating

category (17 points). Although the girls scored significantly better than the boys ($p < 0.5$) in both Canada and the United States, there was a drop in the gender gap.

Table 7.3

Gender Differences Favouring Female Performance in Reading from PIRLS 2006 Results for Canada and the United States ($p < .05$)

Variables	Alberta	B.C	N.S.	Ontario	Quebec	U.S.
Overall Average	8 points	9 points	21 points	13 points	13 points	10 points
Types of Text: Lit	11 points	12 points	18 points	12 points	12 points	12 points
Types of Text: Infor.	7 points	6 points	20 points	11 points	11 points	9 points
Reading Proc: retr & stratfwd inference	6 points	7 points	17 points	11 points	9 points	10 points
Reading Proc:interp, integr & evaluating	11 points	9 points	21 points	13 points	16 points	12 points

In the retrieving and straightforward inferencing processes the following average point differences were reported: 6 points in Alberta, 7 points in British Columbia, 9 points in Quebec, 10 points in the United States, 11 points in Ontario, and 17 points in Nova Scotia. Similarly, but not as great, was a decrease in gender differences in the interpreting, integrating and evaluating reading processing category: British Columbia showed a 9-point difference, Alberta 11 points, the

United States 12 points, Ontario 13 points, Quebec 16, and Nova Scotia 21 points (Mullis et al., 2007, p. 64). Compared to 2001 where differences ranged from 16 to 19 points, the 2006 results in all categories dropped by almost half. Refer to Table 7.3 for specifics.

What caused this drop in gender differences between 2001 and 2006? Are these changes due to reading activities in the home or in the schools resulting in an actual increase in boys' reading performance, or due to a design change of the PIRLS? To further explore these questions, analyses of the data that reported trends in student literacy activities within the home and school between 2001 and 2006 were examined. Trends were reported for only Ontario, Quebec and the United States, because the remaining Canadian provinces did not participate in PIRLS 2001. When students were asked how often they engaged in reading stories and novels outside of school, although Ontario, Quebec and the United States scored above the international average, no significant differences ($p < .05$) were found between the results reported in 2001 and 2006 except for Quebec which showed a significant drop, (4 points, $p < .05$) (Mullis et al., 2007, p. 147). With regards to informational texts read outside of school, students' responses indicated a significant drop ($p < .05$) from 2001 to 2006 for Ontario (3 points), Quebec (2 points), and the United States (4 points) (Mullis et al., 2007, p. 150). This data on students' reading activities indicates there was no increase in the amount of time students reported reading outside of school between the years 2001 and 2006.

Since no increases in literacy activities during leisure time were noted, the reading practices in schools between 2001 and 2006 were analyzed to determine whether there were any major pedagogical shifts that could have accounted for the changes in gender differences in reading achievement in Ontario, Quebec, and the United States. No differences ($p < .05$) were noted in class size (Mullis et al., 2007, p. 188), or in the number of hours reading was taught in the classroom (Mullis et al., 2007, p. 182). There was a drop in the average number of years of teaching experience ($p < .05$) in Ontario and the United States but not in Quebec (Mullis et al., 2007, p. 201). Quebec reported a significant increase ($p < .05$) in the kind of information students were expected to read (literary and informational) but this trend was not detected in Ontario and the United States (Mullis et al., 2007, p. 213). Students were required to complete a variety of comprehension activities after reading. There were no differences in the number of expected written formats or oral summaries teachers requested of students, however, there was a significant increase ($p < .05$) in the practice of having students talk to each other after reading. The strategy of oral discussions to enhance comprehension was noted in Ontario and the United States (Mullis et al., 2007, p. 224). A significant increase ($p < .05$) was reported on the availability of internet access in Ontario and United States (Mullis et al., 2007, p. 233) with a drop in reading assigned for homework in United States and Ontario ($p < .05$) (Mullis et al., 2007, p. 236). Since 2001 there has been a significant increase ($p < .05$) in the use of diagnostic reading tests in Ontario and the United States (but a significant decrease in Quebec ($p < .05$), and a significant increase ($p < .05$) in the

use of multiple-choice tests in the United States for student assessment. It is well known that the teacher plays a central role in how literacy activities are organized and the kinds of materials that are used. Reading instruction in the classroom also depends on the skill, knowledge and expertise of the teacher. Although some trends have been reported between PIRLS 2001 and 2006, it is not possible to determine which specific practices, if any, led to the lowering of gender achievement in reading. An examination of the design of the test was needed before any definitive conclusions could be made.

PIRLS 2006 endeavored to construct a reading assessment using all of the "state-of-the-art methods" (Mullis et al., 2007, p. 16) in order to understand educational practices and trends across the globe. A new framework for PIRLS was established in 2001and PIRLS 2006 continued to extend assessment practices and the manner of reporting achievement. In PIRLS 2006, 10 passages were assessed, 4 of which were kept from the original PIRLS 2001. Six new passages were developed cooperatively involving the participating countries. "In PIRLS 2001, girls had significantly higher achievement than boys in every country so efforts were made to make the passages equally interesting to both genders" (Mullis et al., 2007, p. 18). A primary aim amongst the participating countries was to develop a reading assessment that would be motivating to all students and to search for texts that would be of interest to all students, both boys and girls, in Grade 4. Over 100 texts were reviewed and 6 were selected and added to the 4 original PIRLS 2001 passages. The reading passages and questions were field-

tested and reviewed by reading and measurement specialists from the International Study Center in Boston College.

Based on the evidence presented it is doubtful that the decrease in gender gap between 2001 and 2006 was attributable to changes in school or home practices. However, there is evidence that a change in test design that addressed the reading interests of both boys and girls contributed to the decline in the gender gap in reading comprehension in Canada and the United States.

It is interesting that the PIRLS assessment of students' reading in fourth grade in Canada and the United States showed a decline in the gender gap reported. This decline is contrary to the PISA reading assessment of 15-year-old students that purported a continued increase. PIRLS and PISA work together as complementary assessments. PIRLS reading assessment was designed to measure literacy in the fourth year of formal schooling with a focus on how to improve instruction for the purposes of future improved achievement. PISA reading focuses on 15-year-olds and examines literacy as an indicator of employability and citizenship and collects little information on curriculum and instructional factors related to schooling. However, both assessments define literacy in terms of similar processes and skills (Mullis, Kennedy, Martin, & Sainsbury, 2004, p. 105). "Both PIRLS and PISA view reading as an interactive, constructive process and emphasize the importance for students' ability to reflect on reading and to use reading for different purposes" (Mullis, Kennedy et al., 2004, p. 103). Both assessments include multiple-choice and written responses built to an average of 500 and a standard deviation of 100. This standardization makes it possible to

make comparisons. PIRLS 2006 reported a gender difference between 8 and 13 points in Canada and 10 points in the United States while PISA 2009 reported a gender difference of 34 points in Canada, and 25 each in Mexico and the United States.

Although inconsistencies in a gender gap have been reported between PIRLS and PISA, in the United States the National Center for Education Statistics (NCES), (which takes the data from the NAEP for reporting purposes—also known as the *Nation's Report Card*), has followed long-term trends in reading from 1971 to 2008 for students ages 9, 13, and 17 and found that gender gaps have remained relatively unchanged over time and between ages. The 2008 scores were used as a benchmark, and both the boys and girls at age 9 scored significantly better than in 1971 ($p < .05$). Similarly, the gap between the boys and girls from 2008 has significantly changed with the gap narrowing from 13 points in 1971 to 7 points in 2008. Students aged 13 and 17 showed no significant difference in the gender gap or reading performance between 1971 and 2008 ($p < .05$). The gender gap for 13-years-olds, although not significant, dropped from 11 points in 1971 to 8 points in 2008. For 17-year-olds the gap has remained relatively steady with a 12-point difference reported in 1971 and an 11-point difference in 2008. "Across all three age groups, female students continued to score higher on average in reading than male students in 2008" (Rampey et al., 2009, p. 18), with a slight decrease in the gender gap between 1971 and 2008.

The NAEP gender gaps are similar to the gaps reported for the United States in the PIRLS. What explanation is there for more than twice the gender

gaps reported between the results of the PIRLS and those for PISA? One explanation is that PIRLS has made an attempt to include test items that are fair to both boys and girls. PISA on the other hand has purposefully kept the same framework for reading assessment since 2000 in order to ensure validity, despite the criticism that the test favours girls (Lafontaine & Monseur, 2009). It stands to reason then that the test design of the PISA contributes to the growing reported gender gap. Test design of national and cross-national reading assessments will be further explored in the theme that follows.

Theme III: The Issue of Test Design of National and Cross-national Studies

Policymakers rely on evidence-based studies, such as the PIRLS, PISA and NAEP, to make informed decisions. Because of the potential influence large-scale studies have on the development of educational policies, it is critical that these assessments report valid results (Butler & Adams, 2007). In the case of gender differences some researchers, after further analyses of the data from national and cross-national tests, have challenged the degree to which gender differences have been reported in reading. In particular, the analyses of researchers such as Lietz (2006a, 2006b), Keeves et al. (2006), and Butler and Adams (2007), need further examination.

Petra Lietz (2006a) carried out a meta-analysis of the results of 139 large-scale studies conducted both nationally and internationally between the years 1970 to 2002 on reading achievement at the secondary level (p. 336). Her motive for this meta-analysis was to make sense of the vast array of contradictory evidence on gender differences in reading. She wanted, in a systematic and

quantitative way, to verify the extent of the gender gap in reading. Lietz (2006a) used a two-level hierarchical linear modeling (HLM) for this meta-analysis. The HLM is a two-step process that examines variations from within-study (Level 1) and between-study (Level 2) variables. Examples of Level 1 variables include gender, effort, and time spent doing homework, while Level 2 variables include school conditions such as size of school, or number of resources. The first step in the HLM is to determine if each variable is homogeneous or heterogeneous. If the results are heterogeneous second level factors are examined to see what characteristics have caused the difference in the results. However, due to tremendous variation in the design, size, scope, and scale of each reading assessment, Lietz (2006a) first standardized the results into a metric-free effect size (ES) (p. 329). She was then able to conduct a Level 1 analysis on the gender differences reported in the 139 studies. After Lietz (2006a) determined that the results were not due to chance, she conducted a Level 2 analysis to examine which variable(s) resulted in the discrepancy between the boys' and girls' achievement scores in reading. Several variables were examined (e.g., age, language of administration, calculations of effect size). Each variable was analyzed and those that did not contribute significantly to the difference of the ES were removed. Overall, Lietz (2006a) identified that over half of the differences (59.46%) between gender differences were not due to chance but could be explained by differences in the design of the large-scale assessment programs and in how the effect size was calculated (p. 336). Differences were most pronounced in the recent large-scale studies conducted by NAEP, the Australian Studies, and

PISA. The greatest gender difference was noted in the PISA 2000 assessment. Although Lietz's (2006a) analyses explained more than half of the gender gap observed, she concluded that other reasons may account for a recent increase in reading achievement that favoured girls. These included, "item selection procedures, contextual changes that surround reading in society and at the school, or the scaling of reading scores" (p. 337). Among the three reasons that may account for an increase in gender differences, Lietz (2006a) hypothesized that the scaling of reading scores had the greatest probability since reading in society for the most part did not change drastically and the items selected for large-scale assessments underwent rigorous field testing. Leitz (2006a) then conducted a second analysis that compared national and international reading assessments administered before and after 1992. Her goal for this investigation was to verify if gender differences in large-scale assessments were attributable to the methods in which reading scores were scaled in order to eliminate the effects of measurement errors.

From her initial meta-analysis, Lietz (2006b) noted that there was a difference in the gender gap between large-scale assessments conducted in recent years compared to earlier studies carried out before 1990–1991. Lietz (2006b) chose 1991 as a cut-off date because after this date large-scale assessment programs used the advanced Bayesian estimation procedures to scale scores (p. 131). Using the same two-level hierarchical linear modeling (HLM), Lietz (2006b) first examined 147 primary studies from national and international reports to determine if they had similar or different effect sizes (ES). She then

further analyzed variables that would contribute to the ES, adding the variable of "time" in the analyses (studies conducted before 1992 and those after 1992). Of the possible predictors (e.g., age, language of test-administration) only time had a positive (G = 0.24) and highly significant (p = 0.00) relationship on the effective size (Lietz, 2006b, p. 138). In order to examine whether there was a possible systematic impact on how performance scores were scaled, a dummy code was created to identify studies completed before 1991 (dummy code 0) from studies conducted after 1992 (dummy code 1). "In particular" Lietz (2006b) discovered that,

> it became interesting to examine whether the differences may not be so much stemming from the different testing programs per se but be a consequence of different procedures for calculating test scores that were introduced in the early 1990s. (p. 137)

The gender difference in studies completed before 1992 showed a positive but small (0.06) coefficient and was not significantly different from zero. In contrast, gender differences in studies completed after 1992 revealed a positive and sizeable difference (0.25). Although overall girls performed better than boys before 1992, the difference was small. In contrast after 1992 the gap widened considerably. Overall, a gender difference that favoured girls' achievement in reading was detected in the 147 studies analyzed that was not related to chance, however roughly two-thirds of the difference was associated with the time period the assessment was conducted. In conclusion, Lietz (2006b) proposed two possible explanations for her results: 1) the change in using the new Bayesian

estimation in scaling after 1992 may have introduced a bias into the effect size that may have resulted in an increase in the gender gap reported, or 2) the way in which gender differences were calculated before 1992 were inappropriate and therefore a gender gap was under-reported. Furthermore, according to Lietz (2006b), cross-national studies need to resolve how data is scaled before any discussions continue around trends in gender and achievement.

Even though ongoing advancements have been made to improve the quality of assessment procedures, problems have existed in the analyses of cross-national data since the 1980s (Keeves et al., 2006). Improvements in statistical methods in educational research have addressed some of the difficulties in assessing the range of data within and between countries that participate in large-scale studies, however continued problems remain.

> Anomalous results have been found from secondary data analyses that would appear to stem from the procedures that have been employed during the past 15 years for the estimation of educational achievement. (Keeves et al., 2006, p. 110)

Three persistent issues regarding the analyses of cross-national studies over the past 40 years have been identified and examined by Keeves et al. (2006, p. 112). Firstly, in order to assess school curriculum without making the test too long for students, large assessments use a rotating system (balanced incomplete block (BIB) for test items and questionnaires (up to 9 in the PISA and PIRLS). In order to compare the data between various tests booklets, procedures are used to equate the scores. Secondly, when data is missing from students or schools, procedures

are put in place to estimate the level of achievement to compensate for the missing data. Thirdly, estimations are put in place from prior population distributions in order to improve the accuracy and reduce measurement errors in future assessments. However, Keeves et al. (2006) proposed that the overall effect of such estimations inflates and distorts the results. Since the distributions of scores vary greatly between countries, and populations within countries, if mis-specification were made to form the prior distribution there would be errors in the estimates of measurements (Keeves et al., 2006, p. 114). The Bayesian estimation procedure, essential for multivariate and multilevel models, has been widely used since the 1990s to address the issue of missing data, the rotating system of tests (BIB), and in estimating populations. However, according to the analyses completed by Keeves et al., the Bayesian estimation procedures should be replaced by other methods that are better suited for secondary data analyses, such as bootstrapping or jackknifing (2006, p. 125). Based on the results from the two meta-analyses conducted by Lietz (2006a, 2006b), this is a valid argument. Higher estimated effect sizes were reported for PISA 2000 compared to previous cross-national assessments due to "procedures used for scaling and compensating for missing data and improving the accuracy of the national estimates of performance" (Keeves et al., 2006, p. 113). It is a challenge to find the optimal model to account for widespread disparities in school characteristics and achievement between 75 nations and economic regions worldwide. Realistically, large-scale tests cannot replicate ideal conditions to assess reading, particularly since there is such a range in students' backgrounds and experiences globally.

Another identified factor that challenged the reported gender differences from PISA, and threatened the validity of the assessment, is the widespread concern for the amount of effort students expend when writing the test (Butler & Adams, 2007). Data from PISA 2000 and 2003 were used to examine the impact between student effort reported on the survey and the overall achievement in reading. A second goal of the study was to determine if effort had an influence on reading performance and gender. The first step was to construct an effort variable and then determine the relationship between effort and reading achievement for all countries involved in PISA. An Effort Thermometer was administered at the end of the PISA 2003 test based on a 10-point scale. The percentage of students who did not respond to the Effort Thermometer was 17.5%. Those students who did respond did so accurately (92.8%). Typically most countries scored between seven and eight with a difference of 2.61 between the highest rating country (Denmark) and the lowest rating country (Japan) on the Effort Thermometer (no significant differences were reported). Contrary to popular belief, students in all participating countries were motivated to do their best on the PISA test (Butler & Adams, 2007, p. 286).

Secondly, the researchers completed a correlation between the average national achievements in reading with the mean effort score determined from the first part of the study. They found that the effects were not large and accounted for about only 0.5% of student variation. When compared to economic, social, and cultural characteristics that account for one fifth of the student variation in PISA 2003, these effects are not large enough to invalidate the cross-national

comparisons (Butler & Adams, 2007, p. 290).

Thirdly, the researchers explored the influence of gender, effort, and reading achievement. Since girls reported that they were more engaged in reading on the PISA 2000 survey, the researchers wanted to determine if there was a gender bias in effort that was related to reading achievement. They investigated the difference between males' and females' reported motivation to complete the PISA 2003. An examination between gender and effort by country showed that female students reported a higher effort. Multiple regressions for gender, effort and reading were completed revealing low correlations for some countries (Japan, $r = 0.199$) and modest for others (Norway, $r = 0.364$). However, the correlations were not reported for either Canada, Mexico, or the United States. In all, when adjustments were made for effort there was a reduced gender difference. The middle range for all countries was about 5 points. From the bar graph, approximations were made. The reduced gender difference was less than 1 point for Canada, 2 points for Mexico, and 7.5 points for the United States. In summary, the study revealed that effort does explain some of the increase in female superiority in reading on the PISA 2003. In addition, it was noted that effort is related to reading achievement and that "expenditure of effort is fairly stable across a majority of countries . . . countering the claim that differential effort invalidates international comparisons" (Butler & Adams, 2007, p. 303).

One main purpose for cross-national testing is to follow achievement trends in order to monitor progress and set new educational goals. However, in order for policymakers to make the best decisions they need trustworthy results.

Therefore, those who design international tests must be responsive to findings from current research that call for improved accuracy of their measures. Overall, test design has been shown to be a major factor in gender differences. Although small by comparison, a reduction in gender gap was noted when effort was calculated revealing that an improvement in effort on the part of boys may also improve their reported reading performance.

Theme IV: Gender Gap: Low Achievement and Low Socio-economic Status

Six longitudinal studies and the analyses of two PISA reports are examined in this section. The longitudinal studies investigated the reading achievement of boys and girls over a range of consecutive years and developmental stages: Becker and Forsyth (1990), Grades 3–12; Entwisle et al. (2007), Grades 1–5 with a follow up at age 22; Husain and Millimet (2009), K–Grade 3; Martin and Hoover (1987), Grades 3–8; Matthews et al. (2010), K–Grade 5; Robinson and Lubienski (2011), K–Grade 8. Researchers used longitudinal data to determine trends and distributions in gender gaps in reading comprehension from school populations in the United States. All six American studies reported a small gap in reading comprehension that favoured females. Three of the longitudinal studies revealed a pattern of widening difference among low achieving boys over time (Becker & Forsyth, 1990; Martin & Hoover, 1987; Robinson & Lubienski, 2011). The remaining three investigations revealed a correlation between low achieving boys and socio-economic status (Entwisle et al., 2007; Husain & Millimet, 2009; Matthews et al., 2010).

In Canada, disparities in reading achievement were also reported. Factors

related to under-achievement in reading among 15-year-olds reported from PISA 2000 and 2003 included SES, gender, and regional differences (Edgerton et al., 2008; Willms, 2004). The specifics of each of these investigations are featured next.

Martin and Hoover (1987) followed 4,875 females and 4,497 males from Grades 3 to 8 from 1978 to 1984. Students were assessed during each of the 6 years using the *Iowa Tests of Basic Skills (ITBS)* (Hieronymus, Lindquist, & Hoover, 1978). The ITBS is an achievement battery of 11 subtests (Vocabulary, Reading Comprehension, Spelling, Capitalization, Punctuation, Language Usage, Visual Materials, Reference Materials, Mathematics Concepts, Mathematics Problem Solving, and Mathematics Computation). Data for each student were collected to provide a summary mean, standard deviation and percentile rank for each year. Overall, the standardized differences (female mean minus male mean divided by the total sample SD) were small (Martin & Hoover, 1987, p. 68). A positive standardized difference indicated a higher female score while a negative difference reflected a higher male score. The scores on the Reading Comprehension subtest ranged from .21 in Grade 3 to .10 in Grade 5 (no significant differences were reported in the entire study). The greatest differences were noted in the language subtests, namely Spelling, Capitalization, Punctuation, and Language Usage (Martin & Hoover, 1987, p. 69). Boys had marginally higher scores on the Vocabulary test, (ranging from a -.02 in Grade 5 to a -.08 in Grade 7). When gender differences were analyzed at each stipulated percentile rank (above 90th percentile, between 25th and 75th percentile, and below the 10th

percentile), a noteworthy pattern emerged. In Grade 3 the girls had higher reading achievement scores than the boys at each specific percentile rank by 2 and 3 units (each unit equaled a tenth of a year). However, from Grades 4 to 8 the difference in reading achievement between the boys and girls above the 50th percentile was small and at times the gender gap disappeared entirely. Below the 50th percentile the gap widened in favour of the girls by 3 to 6 units (more or less equivalent to 3 to 6 months). Even greater differences in favour of girls were noted in Spelling, Capitalization and Punctuation below the 50th percentile (up to 12 units/one full school year). In general, the researchers concluded that the reported differences in reading comprehension are small, and although the girls performed better, the differences were most pronounced below the 50th percentile. Above the 50th percentile, the boys did as well as girls on reading comprehension but an increased difference was detected among the boys achieving below the 50th percentile. The trend of a widening gap for boys in the lower percentile ranks supports Martin and Hoover's (1987) concern that examining only gender gap averages can be misleading. The analyses of specific percentiles that comprise the average score provide a much different picture, one that shows the underachievement of the low achieving boys rather than all boys.

Becker and Forsyth (1990) extended Martin and Hoover's (1987) study and examined students' reading achievement from Grades 3 to 12. They investigated gender differences in achievement in five areas (Vocabulary, Language Usage, Reading, Mathematics Problem Solving, and Using Sources of Information) over a period of 10 consecutive years between 1978–1979 through

to 1987–1988 with 1,642 females and 1,360 males. The *Iowa Tests of Basic Skills (ITBS),* (Hieronymus et al., 1978) was used for Grades 3 to 8 and the *Iowa Tests of Educational Development* (Feldt, Forsyth, & Lindquist 1979; Feldt, Forsyth, & Alnot, 1986) for Grades 9 to 12. For each grade the mean, standard deviation, grade equivalent scores, standard scores, and percentiles (90^{th}, 75^{th}, 50^{th}, 25^{th} and $10^{th)}$ were calculated for males and females. Differences between female and male scores were standardized which provided the indices from which the longitudinal trends were compared. Generally, the gender gap in Reading was small to nonexistent at the upper percentiles but widened in favour of females in the lower percentiles. At the 90th percentile the differences in some grades favoured the girls and in other grades the boys (.18 in Grade 3; -.17 in Grade 5; .12 in Grade 12). Similar trends were noted at the 75th percentile. At the 50th percentile the scores varied from a very small gap in favour of the girls in Grade 7 (.04) to a minimal but wider gap by Grade 12 (.27). At the 25th percentile the scores increased in favour of the girls from .12 in Grade 5, to .55 in Grade 12. Scores at the 10th percentile were similar to those at the 25th percentile (.11 in Grade 5 to .59 in Grade 12). Across all grades, although the gap was not the same, the trend was the same. The girls outperformed the boys in reading achievement in Grade 3. From Grades 4 to 8 the gap narrowed, revealing a very small difference and then continued to increase from Grades 8 to 12 in favour of the girls. The gap was widest at the lower percentiles (25th and 10th). Although Becker and Forsyth's (1990) study was not designed to identify reasons for gender differences in reading achievement, it does reveal trends in ability levels

and developmental periods. The authors conclude that gender patterns in reading achievement point to the need for further examination of possible causes of these trends.

In an attempt to understand the reason for gender differences in reading, researchers Entwisle et al., (2007) in their longitudinal study proposed and tested a model that showed that boys from low SES are at most risk for underachieving in reading. Data were collected from a random sample of 403 students from 20 public elementary schools in Baltimore City (276 from meal subsidy programs and 127 from no meal subsidy) who started school in the fall of 1982. Meal subsidy was used as an indicator of low family income. Student scores on the *California Achievement Test* (CAT), (edition was not reported) over the first 5 years in school (Grades 1 to 5) were analyzed. Information from the children's parents regarding their education, occupation and expectations for their children's success in reading and mathematics was gathered throughout the 5 years of the study (1982–1987). Teacher questionnaires regarding attitudes, professional backgrounds, and school records on test scores and retention rates were also collected. An additional survey was conducted 16 years later when students were 22 years of age. This survey provided additional information on employment and school enrollment of the students involved in the study. About 80% of the original group was interviewed.

Results from the CAT reading scores (Entwisle et al., 2007) over the first 5 years of school showed a significant ($p < .05$) gender gap in reading achievement in favour of females among students receiving subsidy, while no

significant differences were noted in reading achievement between boys and girls not on subsidy. Boys from low SES had lower reading scores than girls from low SES in Grades 2, 3, 4, and 5 ($p < .05$). A positive difference indicated a higher female score while a negative difference reflected a higher male score. The gender gap scores were 5.3 in Grade 1, 23.4 in Grade 2, 14.2 in Grade 3, 23.7 in Grade 4, and 17.3 in Grade 5. (The total possible score on the CAT was not disclosed.) For students not receiving meal subsidies, in some years the girls' overall scores were higher (3.3 in Grade 1 and 2.9 in Grade 3), while in other years the boys had an overall higher score (-5.9 in Grade 2, -4.2 in Grade 4, and -6.0 in Grade 5). However, all differences between boys and girls not receiving meal subsidies were small and not significant (Entwisle et al., 2007, p. 124).

A multivariate model, adapted from a study conducted by Entwisle and Alexander (1996) from a previous study, was proposed and tested. Collectively seven predictors were added to the regression model including race, family SES, parents' psychological support, teacher marks, reading achievement (CAT), and retention rates. The results revealed that the girls on subsidy scored 17.3 ($p < .05$) more points on the reading achievement test (CAT) than boys on subsidy over the first 5 years of school. By comparison, the gender gap was not significant for non-subsidy children with a very small difference in favour of the boys (-0.8) (Entwisle et al., 2007, p. 122). The full model explained 44% of the variance in gains in test scores among children on subsidy (Entwisle et al., 2007, p. 122).

> The model thus illustrates how, for subsidy students, the female advantage in parental expectations, reading marks, classroom behaviour, and

> retention . . . together translate into lower reading comprehension scores for boys at the end of five years of elementary school (Entwisle et al., 2007, p. 124).

The researchers then compared the predictor variables (race, family SES, parents' psychological support, marks, reading achievement, and retention) with the descriptive data collected from the questionnaires and other data from related studies to support the conclusion that the combination of the "poor fit" between boys in classrooms and low SES parents' lower expectation of boys resulted in higher rates of retention and placement of boys in low track programs that set the course for the rest of their lives (Entwisle et al., 2007, p. 127). Although some statistics are disclosed from the questionnaires, sample questionnaires are missing, which makes it difficult to determine the full range of questions or data that was collected. However, this longitudinal study does reveal the critical issue of underachievement of low SES boys in reading comprehension in the first 5 years of schooling. Entwisle et al. (2007) provided a strong case for "Early Schooling: The Handicap of Being Poor and Male."

Husain and Millimet (2009) over a 4-year period studied whether boys were lagging behind girls in mathematics and reading during the primary years (K–3). The data used for this study were from the Early Childhood Longitudinal Study-Kindergarten (ELS-K) Class from 1998–1999 to spring of third grade, 2001–2002 (no citation provided). Data from a total of 17,565 students from 994 schools across the U. S. were analyzed. Although the authors report using IRT (Item Response Theory) in their reading measures and analyses, the specific

reading achievement test was not reported. Additional data were collected on each student's age, birth weight, SES, number of books in the home, and mother's age at first birth (not all figures and tables were included in the article). A regression analysis was completed on mathematics and reading comparing race, SES quartiles, regions in the United States, urban vs. rural, and private vs. public schools. Only the reading achievement results are relevant here. The reading comprehension results indicate that overall the girls scored higher than the boys at each grade level ($p < 0.001$) with increases over the 4 years from K to third grade (-0.139 at Kindergarten and -0.202 at end of the third grade, negative scores indicate higher female scores). The total possible score on the reading test was not included in the report.

Husain and Millimet (2009) then examined the gain in the reading scores by boys relative to girls from Kindergarten to Grade 3. When the reading gains were examined by race, SES quintile, region, urban-rural, and school type, a negative gain in reading was detected for boys in the lowest SES quintile and for school type (all other sub-samples were not statistically significant). Low SES quintile scores were reported by race [- .271 (White students), - .185 (Black students), - .389 (Hispanic students)]. School type scores for reading revealed all positive values for private schools [0.146 (White), 0.195 (Black students), 0.159 (Hispanic)] but negative values for public schools [-0.041 (White), -0.154 (Black students), -0.206 (Hispanic students)]. The values represent ground gained by boys relative to girls, so a negative number showed no gain. (All values were reported as statistically significant at the conventional level for both the lowest

SES and school type). However, what is considered conventional was not described but is likely ($p < .05$).

> [F]or children attending private school, boys begin kindergarten substantially behind girls in reading but reduce the gap by over 50% by the end of third grade. In contrast, boys attending public school begin kindergarten relatively closer to girls . . . but then lose ground over the first 4 years of school. The fact that boys lose ground relative to girls in public, but not private schools as well as in the bottom quintile of SES is surely not coincidental since few students from the bottom SES quintile attend private school . . . the widening gender gap in reading . . . is a 'low SES' phenomenon. (Husain & Millimet, 2009, p. 43)

The authors conclude based on the evidence from their study that a widening gender gap is related to SES and not gender per se. The authors challenged the mythical boy crisis and questioned why a gender gap in reading achievement exists in sub-populations (boys from low SES) while not among other populations.

Matthews et al. (2010) also used the Early Childhood Longitudinal Study-Kindergarten Cohort 1998–1999, (NCES, 2002, 2005), and examined the literacy achievement of 12,385 students composed of only African American boys (1,257) and girls (1,237) and non-Hispanic White boys (5,086) and girls (4,805). Literacy achievement data were analyzed using the data from the NCES (kindergarten to the fifth grade). Data were collected from 870 public and private schools selected at random nationwide. Additionally, telephone interviews and teacher surveys

were analyzed. The researchers were primarily interested in the relationship between social skills, specifically learning-related skills (LRS), and how these skills influenced literacy development for African American boys from K to Grade 5. In addition, the researchers wanted to examine how race and gender differences could be reduced among African American students who displayed positive social and behavioural factors, including LRS (attentiveness, persistence, organization, learning, and independence), externalizing behaviours (degree to which children argued, got involved in fights, disturbed activities), and interpersonal skills (friendships, helped other students, sensitive to others' feelings). Each of the social and behavioural factors for each child was determined using a 4-point Likert scale that ranged from one (never) to four (very often). Correlations were calculated between the reading achievement from K to Grade 5, social and behavioural skills, LRS, and background variables such as SES, and home literacy environment. The authors specifically wanted to ascertain which factor influenced reading growth the most. Since the researchers hypothesized that LRS was a significant factor, they wanted to determine the actual effects of LRS on reading growth. Three models were hypothesized and tested: 1) the effects of race and gender on reading achievement 2) the influence of LRS on race, gender and reading achievement, and 3) all social and behavioural factors and their relationship between race, gender and reading achievement. Model 1 indicated a reading achievement gap between African American boys and White boys (K =14.0, Grade 1 = 18.3, Grade 3= 22.7, Grade 5 = 19.1) ($p < .001$). Also noted was a gender gap between African American boys

and African American girls (K= 3.4, $p < .01$; Grade 1= 5.4, $p < .001$; Grade 3 = 4.0, $p < .001$; Grade 5 = 3.0, $p < .01$). Model 2 indicated, "as hypothesized, LRS proved to be the strongest predictor of reading performance and development" (Matthews et al., 2010, p. 763). The details of all results are discussed here because of my interest in gender and reading performance. Ways to mitigate a gender gap are mentioned at this point because they are an extension to the primary goal of the research. Therefore, to avoid confusion between all the background variable scores, only the significant scores are noted. In Kindergarten, LRS was a significant predictor of reading achievement and reduced the size of the race effect by almost half ($p < .001$). Over time, between Kindergarten and Grade 5, LRS decreased the effect of race, although race remained significant ($p < .001$). In Model 3, of all of the variables, only LRS was significantly related to growth in reading achievement through to the fifth grade ($p < .001$). The authors concluded that although SES and literacy environment at home were influential in relation to reading achievement, the factor that proved to have the greatest positive impact was learning-related skills (LRS). Although the study reveals the under-performance of African American boys in reading achievement, the intent of the study was to move beyond identifying the gap and to move toward ways to influence and mitigate differences. Since Learning Related Skills are not innate but learned, then if students acquire these skills they can overcome other risk factors that limit academic success. However, the study clearly demonstrates that although SES is a factor in reading achievement for boys and girls, boys from low SES underachieve relative to girls from low SES.

Robinson and Lubienski (2011) studied gender achievement gaps in mathematics and reading by also using the ECLS-K data collected by the NCES (no citation provided), *The National Report Card.* This study involved 7,075 students from Kindergarten in 1998–1999 to eighth grade. Two broad categories of data were collected—assessments from the cognitive achievement tests from the NCES and teacher evaluations. All assessments were converted to a standardized unit for comparisons. Achievement scores were analyzed at the 10th, 50th, and 90th percentiles for both males and females. In this study, negative scores indicate higher female scores. Although the girls had higher reading achievement scores in all grades and percentiles, between Kindergarten and Grade 5 the gap narrowed between boys and girls. In the spring of Grade 5, the smallest difference in reading achievement was at the 90th percentile (-.019 and not a significant difference) and 75th (-.061, no significant difference), with a larger gap at the 50th (-.186, $p < .01$) and 25th (-.147, $p < .001$) percentiles. By Grade 8, the gender gap in reading achievement widened again (in favor of the girls) ranging from -.243 at the 10th percentile ($p < .001$) to -.105 at the 90th percentile ($p < .01$). An overall trend revealed that at the upper percentiles (90th and 75th) the boys and girls did not differ much in their reading achievement (small difference in favour of the girls but not significant), however the gap continued to widen (in favour of the girls) below the 50th percentile (Robinson & Lubienski, 2011, p. 285). By Grade 8, of the bottom 5% of readers, 67% were boys (Robinson & Lubienski, 2011, p. 288). Interestingly, an examination of the teacher ratings of reading progress compared to the achievement from the

cognitive assessments revealed that teachers consistently rated the girls higher than the boys. The teacher ratings and reading achievements are similar in spring of Grade 1, however the difference widens after Grade 1. Results for subsequent grades indicate that approximately a 0.2 standardized difference existed in Grade 5 and Grade 8 between the teacher rating and the cognitive reading assessment (NCES, nd). The authors questioned whether teacher assessments were influenced by behaviour, with teachers mistaking compliance for achievement. The authors imply that the under-reporting of boys' and over-inflating of girls' achievement could have a negative influence on their actual achievement. Overall, the authors conclude that greater attention is needed in elementary and middle schools on the lowest-achieving males.

Disparities in performance are reported to exist among Canadian school populations. Large variations between regions of Canada, with New Brunswick scoring 33 points below the Canadian average, were reported in PISA 2000 (Willms, 2004, p. 42). Although the range in reading achievement among the different provinces is attributable to students' SES, there was a wide range in scores at all levels of SES (Willms, 2004, p. 18). Other family factors, (number of siblings, parents' occupation and education, educational resources, foreign born) were also found to contribute to the variation (40%) in provincial mean scores (Willms, 2004, p. 42). Edgerton, Peter and Roberts (2008) used regression analyses to analyze the PISA 2003 results for SES, educational aspirations, sex, and province for educational inequalities across Canada. They found four statistically significant predictors ($p < .001$) (SES, expected years of schooling,

sex, province), while one, (SES-Education Aspiration interaction) was not.

> More specifically, (a) for every 1-unit increment on the SES index, a student's reading score increased by 24 points; (b) a 1-year increment in expected number of years of schooling was associated with an increase in reading score of 12 points; (c) girls display a 25-point advantage over boys on the reading score; and (d) students in all provinces except Quebec and British Columbia score significantly below their Alberta counterparts, with Prince Edward Island showing the largest gap at 43 points. (Edgerton et al., p. 874)

Although the study highlighted factors that contributed to underachievement in reading, the researchers did not combine predictors, other than SES and educational aspiration. One hypothesis is that boys from low SES are underachieving more than girls from low SES, since the individual predictors (SES and gender) contributed to the greatest difference in reading achievement, however such a hypothesis is speculative based on the findings reported.

Although research on gender differences has a long history, few studies examined differences over time (Becker & Forsyth, 1990, p. 4). Concerned that most studies report only mean differences between boys' and girls' reading performance, Martin and Hoover (1987), Becker and Forsyth (1990), Robinson and Lubienski (2011) focused on gender differences at various achievement levels (e.g., above average, average, and below average) and found differences in the distribution of achievement gaps. There is a small gap between boys and girls in the average and above average reading achievement levels, however low achieving

boys underachieve more than low achieving girls. A general trend in reading performance between boys and girls was noted across the grades. The girls outperformed the boys in the primary grades but the boys close the gap by Grades 4 to 8. However, the differences between boys and girls continue to increase from Grades 8 to 12 with girls outperforming boys. In all grades, the gap between boys and girls reading achievement is widest at the lower percentiles. Robinson and Lubienski (2011) conclude that greater attention in reading is needed early and with the lowest-achieving males. Entwisle et al. (2007) sought to understand the origin of gender differences and discovered that boys from low SES were retained more often due to lower expectations from parents and teachers. Husain and Millimet (2009) analyzed reading achievement from K–3 from the Early Childhood Longitudinal Study—Kindergarten and reported similar findings—that only boys from low SES do not close the gap in the first years of reading achievement. Matthews et al. (2010) also used the Early Childhood Longitudinal Study—Kindergarten over a 6-year period and found that race and gender differences persist from Kindergarten to fifth grade for African American boys, but a focus on Learning-Related Skills (LRS) support reading achievement for students from low SES and impoverished home literacy environments. From the PISA 2000 results, Willms (2004) reported regional differences in Canada related to SES and family background. Edgerton et al. (2008) investigated additional factors, using the results from PISA 2003. Four predictors were correlated (SES, educational aspiration, sex, and province) with poor achievement in reading. Although four factors were identified, the relationships between the factors were

not analyzed, so it is difficult to determine the compounding effect of these factors.

All six longitudinal studies revealed small but not significant differences in the gender gap in reading achievement for average and above average students but a growing and significant gap for low achieving students (Becker & Forsyth, 1990; Entwisle et al., 2007; Husain & Millimet, 2009; Martin & Hoover, 1987; Matthews et al., 2010; Robinson & Lubienski, 2011). Evidence from the six American longitudinal and two Canadian studies also revealed SES (Edgerton et al., 2008; Husain & Millimet, 2009; Willms, 2004), race (Matthews et al., 2010), teacher (Entwisle et al., 2007; Robinson & Lubienski, 2011), and parent expectations (Entiwisle et al., 2007; Willms, 2004) as potential contributors to the prevalent and growing difference among the lowest achieving boys. It is evident that possible solutions to such a phenomenon of the underachievement of low achieving boys is complex, and one that will require coordination between in and out of school initiatives.

Theme V: Towards an Understanding of Boys Underachieving in Reading

Ten investigations are synthesized and interpreted in this section. Several studies sought to explain why boys were underachieving in reading. Others have examined the effectiveness of interventions in order to address the gap between boys' and girls' performance in reading. This section commences with studies that explore whether boys view reading as a feminine activity thereby explaining why some boys lack motivation to engage in reading (Sokal, 2010; Sokal, Katz, Adkins, Gladu et al., 2005; Steiner et al., 1981). Next, inquiries as to whether the sex of the teacher has an effect on boys' achievement in reading are analyzed

(Sokal & Katz, 2008; Sokal, Katz, Adkins, Grills et al., 2005; Sokal et al., 2009). One investigation looked into the possible use of computers to teach reading and improve boys' reading achievement (Sokal & Katz, 2008). In response to the persistent claims that boys are underachieving, gender-based reforms have seen an increase in the number of single-gender classes in public schools in Canada and the United States. Four case studies investigated the benefits of single-gender classes and the effect these initiatives have on reading achievement (Basilo, 2008; Klinger et al., 2009; Stotsky et al., 2010; Vrooman, 2009). In order to determine which school-level factors influence boys' literacy, the last case study to be discussed compares the approaches used in schools between high-gender gap schools and low-gender gap schools (Klinger et al., 2009). The reasons for boys underachieving in reading, and possible solutions to deal with boys' literacy needs were found to be multi-dimensional and complex.

Steiner et al. presented their study, "A Current Investigation of Sex Differences in Reading Achievement and the Sex-Typing of Reading" at the American Educational Research Association in Los Angeles in 1981. Unfortunately, the study omitted critical details. Notwithstanding, the purpose of the study was to confirm previous findings from studies conducted by Downing and Thomson (1977) and Downing et al. (1979) that claimed that male students viewed reading as a female activity, which was a contributing factor to the underachievement of boys in reading. Steiner et al.'s (1981) study addressed five questions: 1) Are there significant differences between girls' and boys' reading achievement scores from Grades 1 to 4? 2) What is the relationship between the

perception that reading is a female activity and reading achievement for gender, IQ, grade level, age, reading attitude, and SES? 3) Do more boys consider reading to be a feminine activity than girls? 4) Is there a change in the perception between Grades 1 and 4 in the numbers of students who perceive reading to be a feminine activity? and, 5) What is the relationship between each of the independent variables (grade level, reading achievement, IQ, age, reading attitude, SES) and the dependent variable (perception that reading is a feminine activity)? Students (sample numbers for grades and gender were not provided) from Grades 1 to 4 in two schools were assessed using three measures: 1) the *Downing Object-Activity Opinion Survey*, 2) the *Estes Reading Attitude Scale*, and 3) *Iowa Tests of Basic Skills* (editions of tests were not included). Multiple linear regressions were used to analyze the relationship between the independent and dependent variables. A chi-square analysis was completed to determine whether reading was perceived to be a female activity. However, no data tables or statistics were included in the study. The main findings include: 1) There were no significant differences between the girls' and boys' reading achievement scores in all grades; 2) The perception that reading is a feminine activity did not affect reading achievement; 3) Reading was perceived to be an appropriate activity for both boys and girls; 4) and there were no significant changes ($p < .05$ level) in how boys and girls perceived reading between Grades 1 and 4, and 5) There was a positive correlation (no statistics provided) between having a positive reading attitude and the perception that reading is for both boys and girls, and having a negative reading attitude and the perception that reading is a feminine activity. The

authors concluded that there is a trend away from the perception that reading is a feminine activity and a greater emphasis on acquiring reading skills by students. As a result, Steiner et al. (1981) challenged the findings from prior research and instead emphasized that both boys and girls see value in reading. However, all results need to be considered with caution since the data were not available to verify the findings and conclusions purported.

Sokal, Katz, Adkins, Gladu et al. (2005) examined the interplay between boys' intrinsic motivation to read and their perception of reading as a male, female, or gender-neutral activity. The primary goal of the investigation was to determine whether boys' views of reading could be modified using exposure to boy-friendly books and male reading tutors. Sixty-nine boys from four diverse elementary schools in Grade 2 in a large Canadian urban center participated in the study over a 10-week period. They examined students from Grade 2 since the literature shows that boys at age 7 have already established what is a male and female activity (Sokal, Katz, Adkins, Gladu et al., 2005). The boys were randomly assigned into four reading groups: 1) female tutors reading typical books; 2) female tutors reading boy-friendly books; 3) male tutors reading typical books; and 4) male tutors reading boy-friendly books. The tutors worked with the boys once a week for 10 weeks for 20–40 minutes. Boy-friendly books included stories about animals, adventure stories with boys as the main characters, sports, series such as *Captain Underpants*, informational texts, and natural disasters. A modified *Children's Academic Intrinsic Motivation Instrument* (Gottfried, 1985) based on a 5-point Likert scale was administered before and after the 10-week

reading period. Added to the self-rated assessment were statements that asked the students to rate reading as a feminine, and/or masculine activity. Therefore, the instrument measured both intrinsic motivation and perceptions regarding reading as an activity. Analyses of the data revealed several findings. Overall, boys who listened to boy-friendly books exhibited a significant reduction in their views of reading as a feminine activity (-1.16 on the 5-point Likert scale) compared to boys who listened to typical books (.17 on the 5-point Likert scale), ($p = .02$). Regardless of the sex of the tutor, the difference in attitude toward reading was related to the type of books that were read (boy-friendly books were preferred).

Further analyses of a) intrinsic motivation and b) views of reading were completed with subgroups. To assess intrinsic motivation, two subgroups of boys were examined; boys who liked reading, and those who didn't like reading. Fifty-four of the 69 boys liked reading. Boys who liked reading and who listened to boy-friendly books showed a greater reduction in their views of reading as feminine (-1.18) compared to boys who liked reading but who listened to typical books (-.19), ($p = .03$). Of the 15 boys who did not like reading, within this group the boys who experienced boy-friendly books showed greater gains in their interest in reading (3.73) over the boys who listened to typical books (1.04), $p = .05$). To assess views on reading other subgroups were analyzed; boys who viewed reading as a feminine activity, and another group who viewed reading as not feminine (neutral or masculine). Of the 52 boys who did not view reading as feminine, although not significant, the boys who were read typical books by female tutors increased their intrinsic motivation (liked reading more) (.23 on the 5-point Likert

scale), while the boys who were read typical books from male tutors reduced their intrinsic motivation (decreased the amount they liked reading) (.31 on the 5-point Likert scale). Boys' interest to read was not positively influenced by having a male tutor or reading boy-friendly books, which implies that it is the quality of the instruction that matters. Among the subgroup of boys who viewed reading as feminine (17), neither the genre of the book, nor sex of tutor had any influence on these boys' perceptions of reading as a feminine activity (Sokal, Katz, Adkins, Gladu et al., 2005). The authors concluded that merely offering boy-friendly books and hiring more male teachers in the classroom will not serve the needs of boys. Such oversimplified solutions cannot solve the complexities of boys' intrinsic motivations and perceptions of reading. The authors cautioned against drawing implications from a small 10-week study. However, they pointed out the need to offer choice as well as a wide range of books to address the needs of a range of boys from varied backgrounds in most classrooms (Sokal, Katz, Adkins, Gladu et al., 2005).

In addition to choice and a wide range of genres, proposals for the use of computer-based books as a way to engage boys and improve their reading achievement have been made. In response to suggestions to include more boy-friendly practices, Sokal and Katz (2008) investigated the effects technology and male teachers would have on boys' self-perceptions of reading and reading achievement. The study included 119 boys in Grades 3 and 4 from 12 schools in Winnipeg. Demographic information was obtained from a parent and principal survey and revealed that 76 % of the schools were located in the inner city. Fifty-

six percent of the children's parents were from a variety of non-Canadian ethnic groups and approximately 60 % of the families lived in poverty (determined by their annual salary). All boys were considered "at risk" in reading achievement, ranging from 1.4 to 2.4 grades below their grade placement. Classroom teachers recommended the boys for the study. Three assessments were administered before and after participating in the 22-week reading program: 1) *The Alberta Diagnostic Reading Program* (Alberta Education, 1986), an informal reading inventory; 2) *Gendered activities Q-sort* (Sokal, Monette, McBey, & Wocjik, 2006) where children sorted pictures of activities (e.g., reading) into three categories (female activity, male activity, gender-neutral activity); 3) *Readers' self-perception scale* (Henck & Melnick, 1995) that included 33 statements (on a 5-point Likert scale) on readers' self-perceptions. The students were randomly assigned to 2 groups, half to a technology-based reading program and half to the no-technology reading group. Male and female research assistants were divided between the 2 reading groups and instructed each student using a paired reading strategy for 30 minutes a week for 22 weeks. Paired reading is duet reading where a student and tutor read together and upon an agreed signal the student moves into solo reading. When the student makes an error, the tutor corrects the student and begins reading simultaneously until the student signals to read independently again (Sokal & Katz, 2008).

The findings indicated that overall significant gains were made for all the boys in both reading performance (0.67 grade equivalent gain, $p = .00$), and general self-perception as readers ($t = 3.38$, $p = 0.001$). A multivariate analysis of

variance was completed on the fixed factors (sex of reading teacher, use of computers or books), and dependent variables (boys' reading scores, self-reported gender view of reading and reading self-perceptions). Results showed that the sex of the teacher had no significant effects on any of the dependent variables (p ranged from 0.15 to 0.81). Boys who read using computers viewed reading as less feminine (0.12 points on a 3-point scale) while boys who read from books increased their views that reading was a feminine activity by 0.09 points, ($p = 0.02$). Lastly, the researchers analyzed the effects of the sex of the tutor combined with computer-based reading. The boys taught by female tutors using computer-based reading did not show any significant differences on achievement scores, gendered views, or reading self-perceptions. Boys who were taught by male tutors and who used computers demonstrated a de-feminized view of reading by 0.17 ($p = 0.03$) while boys who worked with male tutors who did not use computers showed an increase in their view of reading as a feminine activity (0.11 on a 3-point scale and no significant differences were reported). Overall, being taught by male or female tutors, using technology or regular books to teach reading did not reveal a difference in the boys' reading achievement or in their reported self-perceptions. The combination of having a male tutor and using computers as a method to teach reading changed only the boys' view that reading was not a feminine activity. From their findings from this study Sokal and Katz (2008) challenge the claim that, "a gendered view of reading is the root of most boys' reading difficulties" (p. 89) and instead cast

> suspicion on the wisdom of interventions aimed at 'masculinising' reading

> practices as a way to deal with underachievement. A feminine view of reading, while evident in some boys, is not the basis of reading difficulties in all or even most boys in our study and generalised intervention based on this belief is misdirected (p. 89).

Sokal et al. (2009) completed a similar study with 173 third- and fourth-grade boys identified as struggling readers. The design and questions of the investigation were identical to the study Sokal completed in 2008 with Katz, except computers were not included as an independent variable. Their findings were similar to the previous study. For boys who viewed reading as a masculine or gender-neutral activity, the gender of the tutor had no effect on the boys' reading achievement or reported self-perceptions. However, one interesting new finding came from a small group of boys who viewed reading as a feminine activity. Interestingly, they responded better to female tutors than to males on the self-perception assessment. Boys struggle in reading for multiple reasons, therefore Sokal et al. (2009) caution school boards against the practice of hiring teachers based on gender. The diverse needs and interests of children in reading require a complexity of approaches that cannot be solved by one variable such as the gender of the teacher.

Sokal (2010) completed a cross-national comparison on the "Prevalence of Gender Views of Reading in Thailand and Canada." Reading results from PISA 2000, 2003, and 2006 of 15-year-olds reveal that Canada has one of the lowest gender gaps in reading worldwide while Thailand has one of the largest. Sokal (2010) sought to compare the views on reading by Canadian and Thai boys. The

purpose of this study was to verify if there was a similar or different gendered view (that reading is a feminine activity) in Thailand and Canada. Similar findings for both countries would suggest that the root of a gender gap in reading achievement lies in the fact that boys view reading as a feminine activity. A total of 168 boys participated in the study with a subgroup from a large central Canadian city and another subgroup from Bangkok. The boys were matched on grade (K–7), age, and SES (family income). The boys were from middle to upper income families whose parents were well educated. The researchers completed the *Gendered Activities Q-sort* (Sokal et al., 2006) with each participant from Canada and Thailand. The assessment required that the boys sort a number of pictures depicting different activities and categorize each picture as masculine, feminine or gender-neutral (reading was one of the pictures). In addition, the boys' families completed a parent survey. Examination of the data revealed that few boys regarded reading as a feminine activity (no significant difference was found, $p = .35$). Since boys from both countries (Canada and Thailand) do not regard reading to be a feminine activity, the authors concluded that the differences in the gender gap as reported in PISA 2000, 2003, and 2006 are related to factors, other than the boys' perception of reading. Since this study investigated middle and upper SES boys, Sokal (2010) cautioned that it is possible that the findings would be quite different if boys from low SES were assessed. "The findings of the current study highlight the importance of studying boys' reading underachievement in more sensitive and sophisticated ways that consider the variety of experiences that contribute to "being a boy"' (Sokal, 2010, p. 5). Sokal

considers that differences in males' views of reading may also be the result of family attitudes which can only be examined by studying the interrelationship between home and school. Since Sokal (2010) examined middle and upper SES, it would be interesting to study the perception of reading from a cross-national comparison with boys from lower SES.

In order to address the underachievement of boys in literacy, some public school districts in North America have considered boys-only classes, a relatively new development in single-gender education programs. Historically, girls' and boys' private and in some cases, religious schools were available for those who could afford it. Single-gendered private schools have provided exceptional preparation for girls in both academics and leadership (Blair & Sanford, 1999). Girls-only public programs in Canada and the United States grew in popularity in the 1990s. Primarily, the purpose of girls-only classes in public schools was to allow more opportunities for young women to focus on academics and to gain skills in science, mathematics and technology typically dominated by males. However, the emergence of boys-only classes in public schools is a relatively new phenomenon (Blair & Sanford, 1999).

In a paper presented to the American Educational Research Association in 1999, Blair and Sanford outlined three single-gender programs in Western Canada. The primary purpose of their case study was to examine how the teachers viewed their practice in relation to the goals of the program. Of the three programs, two included single-sex classes for both Junior High School boys and girls, and the third offered programming for adolescent girls only. The

researchers conducted interviews with students, teachers, administrators, and parents (although the numbers interviewed and an example of the interview questions were not disclosed in the report). From the interviews, Blair and Sanford established that although the purpose for offering a single-gendered class varied between the three schools, all schools indicated that the main reason for offering single-gender classes was for the purpose of addressing gender issues with adolescents; in particular, equal opportunities for girls and concerns for boys' literacy. The researchers discovered that the school boards that offered these single-gender classes did not have any policies outlining either the rationale or goals for such an alternative program. Further there was very little research that supported single-gendered programs in public schools (at the time of their report). Overall, the policies and reasons for offering girls-only classes were clearer than for the boys' only classes. The teachers, upon reflection of their teaching practice in single-gendered classes, indicated that they needed to look more closely at the curricula and consider additional practices that would fit the needs of their students better especially for the boys-only classes than the girls-only. To improve upon programming to meet the needs of boys, teachers considered pedagogical changes that would involve increased time in physical education, outdoor education, and a redesign of the language arts program (Blair & Sanford, 1999) (although not clearly specified, it is assumed that the teachers thought these features would improve and engage boys more in their learning). From such an exploratory investigation in single-gendered classes, there has been a growing rise in the number of programs offered, particularly in boys-only

classes. Evidence from the case studies presented next however, is mixed.

Eric Basilo completed three case studies on the reading performance of boys from single-sex classrooms for his PhD dissertation (2008). The focus of the study was to examine if boys from single-sex classes outperformed boys from coeducational classes in reading achievement. Data were collected from state reading tests, written questionnaires, and personal teacher interviews. Schools were selected using the database from the National Association for Single-Sex Public Education (NASSPE).

Case One involved students in Grade 3 in a school of 800 students located in Central Florida. About 60% of the students received free and reduced lunch signaling that more than half of the students were from low-income families. The single-sex class consisted of 20 students and the coeducation group of 80 boys and 37 girls. After a full year of teaching and learning (2006–2007), the *Florida Comprehensive Assessment Test* (FCAT) (neither the edition nor the authors were identified) was administered for math and reading. When the reading results, at the end of Grade 3, from the single-sex class were compared to the results from the coeducation class there was no statistical difference ($p < 0.05$). However, when the boys from the coeducation class were compared with the all-boys class, the male students from the single-sex class scored significantly better ($p < .05$) than the males from the coeducation class. Data were not provided from the school district to compare the learning gains over the course of the school year between the coeducation class and the all-boys class (Basilo, 2008, pp. 100–102). The initial and end-of-year reading levels provide information on students'

reading growth regardless of the grade of achievement the students started their year. It is possible that the students in the single-sex class started Grade 3 with higher reading levels.

Case Two involved students from Grade 5 in an elementary school of 684 students located in South Florida. The single-sex class had 12 boys and the coeducation classes consisted of 100 students (52 males and 48 females). Results from the FCAT (edition and authors were not disclosed) in reading revealed no significant differences ($p < .05$) when the class results were compared. Results from the boys' reading scores between the single-sex class compared to the boys from the coeducation classes, no significant differences were noted ($p < .05$). Although not significant, the boys in the coeducation classes had a higher mean than the boys from the single-sex class. A chi-square test was performed to determine if there was a significant relationship between learning gain and class structure (single-sex vs. coeducation). Results uncovered no statistical significance ($p < 0.05$). "In other words, no significant relationship could be determined between single-sex and coeducational classes regarding learning gains" (Basilo, 2008, p. 101).

Case Three included students from Grade 6 from a rural school of 624 students in Southern Louisiana. Over half of the students (58%) were eligible for free or reduced lunch, signifying that more than half of the students were from low SES. The school had two single-sex programs, a male group of 17 boys, a female group of 15 girls, and a coeducation group of 98 students (52 boys and 47 girls). Test scores from the state test in reading (specific name, edition, or author

was not provided), implied that although, on average, the students in the coeducational classes (both boys and girls) scored higher than those in the single-sex classes, the results were not statistically significant ($p < 0.05$). When the boys' reading performance from the single-sex class was compared to the boys' reading performance in the coeducation classes (not significant $p < 0.05$), the boys in the coeducational classes had a higher mean on the state reading achievement test. To determine gains in reading performance, a chi-square test of independence was performed comparing the relationship between the gains in reading with the type of class. Only the male students from each class were analyzed. Although the boys in the coeducational classes, on average, had higher learning gains than the boys from the single-sex class, the gains were not statistically significant ($p < 0.05$) (Basilo, 2008, pp. 104–109).

Teacher responses to the questionnaires and interviews regarding the single-sex classes were mixed. Some teachers reported that all boys' classes were more competitive and that the boys had poor social skills. Other teachers appreciated the smaller classes, took advantage of computer-based learning, and used a guided discovery approach to successfully engage boys in learning. One common finding from the qualitative data was the lack of training the teachers received in professional learning to support the strategies needed to adapt teaching practices for boys-only classrooms. The author suggested that further research was needed in order to determine if single-sex classrooms would improve the achievement gap in reading between boys and girls. Based on the three case studies investigated, only the boys in the Grade 3 (20 males) showed an

increase in reading achievement. The boys in the Grades 5 (12 males) and 6 groups (17 males), even though they were in a smaller class size, did not perform better in reading achievement than the boys in the coeducation class. It would be interesting to study single-sex classes over time; possibly a longitudinal quasi-experimental study would provide greater insight into the benefits and drawbacks of boys-only groupings.

Title IX of *The Education Amendment Act* was passed in 1972 in the United States in order to promote equivalent educational opportunities for both males and females and this act gave rise to the growth of coeducation classes in public schools throughout America. When the policy, *No Child Left Behind* (NCLB, 2002) was passed, annual standardized testing across the nation was mandated. In addition, further assessments from national and international level testing (NCES, OECD) revealed consistent achievement gaps between boys and girls in reading and mathematics. In response to the reported gender-gap, policymakers and administrators sought alternative ways to educate boys and girls, drawing on brain-based theory that holds that boys and girls learn differently. Single-gender classes as a result, continue to grow in popularity across the United States (Vrooman, 2009).

In order to determine if boys-only and girls-only classes make a difference in student achievement, Marilyn Vrooman (2009) completed her doctoral dissertation on a comparison of reading and mathematics performance of single-gender classes (boys and girls) with coeducation classes in three middle schools in an urban school district. Only the results from reading are of interest here, and

thus were analyzed and synthesized. Data from 4064 students, using the *Oklahoma's Core Curriculum Test* (CCT) (no editions or authors were provided) in Grades 6, 7, and 8 were collected. A General Linear Model of a six factorial Analyses of Covariance (ANCOVA) was used with two independent variables (classroom type and gender) and three dependent variables (reading scores for Grades 6, 7, and 8) for the years 2006, 2007, and 2008. Two main results from the reading data were found. First, there was no significant difference between the students in the single-gender classes (both boys and girls) and the coeducation classes in reading in Grade 6 (sig = .015), Grade 7 (sig = .935) and Grade 8 (sig = .009), (Vrooman, 2009, p. 75). Second, for both the single-gender classes (boys and girls) and the coeducation classes, the boys had higher average scores in reading in all three grades, however the results were not significant (Grade 6 [sig= .127], Grade 7 [sig= 124], Grade 8 [sig = .788]) (Vrooman, 2009, pp. 75–76). When the reading scores from the male and female students in the coeducation classrooms were compared with the male and female students in the single-gender classrooms, there were no significant differences (Grade 6 [sig = .450], Grade 7 [sig = .080], Grade 8 [sig = .028]) (Vrooman, 2009, p. 75). Although both boys and girls performed better on mathematics for single-gender classes than those in coeducation classes, no differences were found on reading. In the case of reading achievement, there was no advantage to being in a single-gender class or a coeducation class. From this study, factors other than reading would need to be analyzed and studied before definitive conclusions could be made declaring the benefits of single-sex classrooms. Also, Vrooman (2009) opposed the findings

cited in national and cross-national testing that indicate boys are falling behind girls in reading achievement.

Even though more investigations have been completed with middle- and high-school-aged students from single-sex classes, fewer research studies have been conducted in public elementary schools. Particularly absent from the literature are investigations that demonstrate that boys-only classes increase reading achievement. To advance this research gap, Stotsky et al. (2010) conducted a study with two Arkansas Elementary Schools. They compared the reading achievement of students in single-sex classes to students from a coeducation class in Grades 5 and 6. In both schools data were collected and compared from three classes: all boys, all girls, and coeducation. Data were collected from the annual state (Arkansas) *Assessment of Literacy* (edition is not included), and interviews were conducted on the benefits and disadvantages of single-sex classes with the principals, teachers, several parents, and two students who happened to be with their parents at the time of the interviews.

In Elementary School A, 60 students in Grade 5 (27 girls and 33 boys) were divided into three balanced groups; (e.g., similar number of ELL, special needs, and high achieving) 21 in a boys'-only class; 23 (16 girls and 12 boys) in a coeducation class; and 16 in a girls'-only class. Results from the state assessment showed that students in all three classes made significant gains ($p < .05$) in their reading from Grade 4 to Grade 5. No significant differences ($p < .05$) were reported in reading achievement between the coeducation class and girls-only class. The reading achievement was significantly better ($p < .05$) for the boys-

only classes than for the coeducation class. Interview comments regarding the benefits and disadvantages of an all-boys class and an all-girls class were synthesized according to teacher, parents, and student comments. The teacher in the boys-only class reported having an advantage to choose books tailored to boys and that the boys got along better with fewer distractions, however the teacher found that it was difficult to maintain the energy level needed to teach young boys. The teacher of the all-girls class found that the girls were delightful to teach and noted that the girls not only enjoyed being together but also demonstrated more confidence. Parents of children in single-sex classes commented that the girls were able to talk about girls' issues and boys were less self-conscious. However, some parents were concerned that the girls completed fewer math and science lessons and talked too much in class. Parents of the all-boys' class made no negative comments. Only one girl was interviewed and she commented that she enjoyed not having loud and distracting boys in the class. No boys' comments were included.

About 700 students from pre-school to Grade 6 attended Elementary School B. There was a total of 70 students in Grade 6; 23 in the all-girls' class; 24 in the all-boys' class; and 23 in a coeducation class. Similar to School A, the teachers placed students into three classes with an even spread of abilities and demographics. Data were provided for the number of absences, discipline referrals, and scores from the state's reading test. Gains in reading were determined from the difference in the scores between Grade 5 and Grade 6. Group means were compared with a repeated ANOVA, using the .05 level of

statistical significance. Overall, the girls and boys made significant gains between Grade 5 and Grade 6 ($p < .05$). The girls and boys in the coeducation class made greater gains in reading achievement than the girls and the boys from the single-sex class and gains were significant for the boys but not for the girls ($p < .05$). The comments from the principal, teachers, parents, and students were mixed. Benefits were mentioned in terms of reading choice and writing achievement for the boys, however, the teachers and parents noted that the boys were louder, more active, and competitive than normal. No significant differences were observed in daily attendance or discipline referrals between the single-sex classes and coeducation class ($p < .05$). Stotsky et al. (2010) concluded that there was neither an academic downside to experimenting with single-sex classes nor an advantage either.

The *Education Quality and Accountability Office* (EQAO), an agency funded by the Government of Ontario, worked collaboratively with researchers from Queen's University in Kingston to better understand gender differences in reading and writing achievement (Klinger et al., 2009). EQAO assessments are administered to all students in Ontario in Grades 3, 6, 9, and 10. The assessment measures the reading, writing and mathematics skills from the Ontario Curriculum. The goal of this investigation was to explore the most effective strategies in reading and writing that would close the widely-reported gender gap. Three sets of analyses were completed: 1) Reported reading and writing data from EQAO for Grades 3 and 6 for the years between 2004 and 2007 were used to tag schools that had high and low gender gaps in literacy; 2) From 400 schools (200

with high gender gaps and 200 with low gender gaps), 13 schools were identified and the students from these schools were surveyed to establish student factors related to reading and writing; and 3) Two schools from the 13 identified in the second analyses, served as pilot schools to refine the research protocols for the remaining case studies. In-depth case studies were completed with all 13 marked schools.

Based on the EQAO data, schools were ranked on the differences between the girls' and boys' scores on reading and writing. A second set of analyses was completed on 200 schools that were identified as having the lowest gender gaps and 200 schools considered to have the highest gender gaps. Raw reading and writing scores were converted into four levels (Level 4: above the provincial standard; Level 3: high level of achievement; Level 2: approaching the provincial standard and Level 1: below the provincial standard) (EQAO, 2007). Results revealed that the boys with high reading and writing scores attended schools with a low gender gap in achievement (reading = 3.22 for the boys, and 3.07 for the girls; writing = 3.13 for boys and 3.25 for girls). By comparison, the boys from high gender gap schools tended to have an *Individualized Educational Plan* (IEP), and attended schools with poorer overall performance for all students (reading = 2.75 for boys and 3.33 for girls; writing = 2.83 for boys and 3.37 for girls).

Based on the EQAO 2007 results, seven schools identified as having a low gender gap in reading and writing were compared to six schools with high gender gaps. A total of 1,019 students were surveyed to identify student and school factors related to literacy. Three main significant differences ($a < .05$) were found

between the boys from low-gap and those from the high-gap schools. Boys from low-gap schools indicated on the survey that they did more homework, played a musical instrument, and read more at home. All other descriptive results revealed no significant differences between high and low-gap schools (other questions included: attitude toward reading and writing, number that played sports, used the computer, library use, and whether they volunteered at school). Although schools were reported as having similar demographics (grade, number of boys and girls, school size, urban-rural), corresponding SES was not reported.

A more in-depth case study was completed with the 13 identified schools. Data for the case studies were collected from the school principal, teachers, and students. The qualitative analyses from the case studies revealed three recurring themes: 1) Teachers and administrators from both low- and high-gap schools held a variety of beliefs regarding boys' and girls' literacy. Many of the widely held beliefs are not supported in the research and are in fact contrary to the research evidence (e.g., male teachers will increase boys' reading achievement; boys' achievement in reading will increase if offered "boys' friendly" books such as non-fiction, comics and computer reading). Therefore, researchers need to publish accurate evidence and disclose practices that are possibly detrimental to boys' learning. 2) It is critical to identify good teaching practices that benefit all students. Teachers from low-gap schools demanded high standards and made efforts to ensure that all students, including boys, produce quality work. In contrast, teachers from high-gap schools found boys to be a challenge to teach, and focused their attention on ways to help boys concentrate and learn. 3)

Differences in school culture were reported between low- and high-gap schools. Low-gap schools were able to create a community that was caring and collaborative, valued learning, and accepted all students. The administrator was identified as the key instructional leader, who supported a positive atmosphere and a culture that focused on learning in the school. In conclusion, the researchers encouraged educators to refrain from focusing on the underachievement of boys. Rather, they advised educators to concentrate on what groups of students are underachieving. School administrators and teachers need to identify and use classroom practices that address the variability of learners within the classroom.

In response to the increased numbers of claims that indicate there is a boys' crisis in education, the findings from the 10 studies cited provide some clarity to the range of conflicting information regarding boys and reading achievement. Although the research from Downing and Thomson (1977) and Downing et al. (1979) indicated that the "root" cause for the underachievement of males in reading is that boys view reading as a feminine activity, Steiner et al. (1981) countered this claim. Results from Steiner et al. (1981) not only indicate that there is no significant difference between the girls' and boys' reading achievement scores in Grades 1 to 4, but that reading is perceived to be an appropriate activity for both boys and girls. Sokal (2010) completed a cross-national comparison of the "Prevalence of Gender Views of Reading in Thailand and Canada" and arrived at the same conclusion. Although Sokal's study primarily focused on boys from middle and high SES, Sokal, Katz, Adkins, Gladu

et al. (2005) investigated a diverse school district in Canada and discovered that most of the boys in Grade 2 enjoyed reading (about 80%). Solutions to deal with a small subgroup of boys who did view reading as feminine (17/69 = 24%) could not be resolved by simple remedies such as offering boy-friendly books or hiring more male teachers. Sokal et al. (2009) conducted a similar study with 173 third- and fourth-grade boys identified as struggling readers, and found similar findings that the gender of the tutor had no effect on the boys' reading achievement or self-perceptions of reading. Additionally, Sokal's and Katz's (2008) investigation reported that technology and male teachers also did not change boys' reading achievement or self-perceptions and suggested other factors such as home and socio-cultural factors need to be explored further. A return to the former days of same-sex classes has not shown to improve reading achievement for boys either. In all case studies reviewed, the boys from same-sex classes did not perform better in reading achievement than boys from coeducation classes. Qualitative reviews from the teachers and students were also mixed (Basilo, 2008; Blair & Sanford, 1999; Stotsky et al., 2010; Vrooman, 2009). An in-depth study completed between low-gender gap schools and high-gender gap schools in Ontario, revealed that teacher attitudes towards their students influenced how all students achieved in their classrooms including boys. When teachers demanded high standards and made efforts to ensure that all students produced quality work, there was little difference in the gender gap (Klinger et al., 2009). In addition, the presence of a strong instructional leader was deemed critical in providing a positive learning culture. In conclusion, Klinger et al. (2009) encouraged

educators to refrain from focusing on the underachievement of boys and to engage in concentrating on which sub-groups of students are underachieving. Findings from the ten studies presented are contrary to many of the views in the mass-media and popular non-research-based literature.

Synthesis and Interpretation of Findings

Publications on the topic of gender and reading achievement have more than doubled between 1945 and 1980 (about 700 studies were previewed) and 1981 and 2011 (approximately 2000 investigations were screened), revealing an explosion in the number of published pieces during this time period, and an increase in interest in the topic of boys and literacy. From the vast array of studies, 41 were deemed topical and robust. These 41 investigations include a range of research methodologies: descriptive, explanatory and experimental. After a critical analysis of the findings from the 41 studies, five unifying and recurring themes were identified.

Firstly, nine studies indicate no significant differences in reading achievement between boys and girls in elementary (Flynn & Rahbar, 1994; Harper & Pelletier, 2008; Lummis & Stevenson, 1990; Phillips et al., 2002; Quinlan, 1996; Wargacki, 2008), middle school (MacFarlane, 2001), and high school (Hogrebe et al., 1984; White, 2007). However, such findings have not influenced the perceptions of parents and educators. A prevailing view from parents, as noted in the study completed by Lummis and Stevenson, (1990) is that girls are superior in reading, even when the evidence indicates no gender differences. The over-identification of boys in remedial classes was noted by

Flynn and Rahbar (1994), who discovered that more boys than girls were identified as learning disabled when teacher judgments were used rather than standardized tests. Researchers also stress how gender differences can be overemphasized in experimental and quasi-experimental studies. In large-scale assessments, slight differences in raw scores show up as being statistically significant. If significant difference is interpreted as meaning sizeable rather than unlikely to have occurred by chance, this misunderstanding will lead the reader into thinking there is a wide difference in scores. Large-scale tests that show a significant difference between boys' and girls' reading achievement have been interpreted as prodigious when in actuality the difference in real scores is quite small (Hogrebe et al., 1984; White, 2007). Understandably, the case for boys' underachieving in reading needs to be challenged.

Secondly, PISA 2000, 2003, 2006, and 2009 report worldwide a growing gender gap that favours girls in reading achievement amongst 15-year-old students. In contrast the PIRLS reading assessment, a similar international test with students in fourth grade (approximately 9–10 years of age), shows a decreased gender gap in Canada and the United States. Interestingly, the gender-gap in reading performance on the PIRLS for the United States is similar to the results reported from their *National Report Card* (NAEP). Findings show that although girls, on average, have higher scores in reading than boys, the gap is about half that reported in PISA, and has decreased for students aged 9, 13, and 17 between the years 1971 and 2008. Canadian results from the PIRLS also show a decline in the gender gap in reading from 2001 and 2006. One explanation for

the difference between PIRLS and PISA is that although PIRLS has made an attempt to include test items that are fair to both boys and girls, PISA has purposefully kept the same framework in order to ensure validity, despite the criticism that the test favours girls and is biased against boys (Lafontaine & Monseur, 2009).

Thirdly, to further examine test bias a meta-analysis was completed on 139 large-scale studies between the years of 1970 and 2002 (Lietz, 2006a, 2006b). The results showed that although overall, girls performed better than boys before 1992, the difference was small and that after 1992, the gap widened considerably. Approximately two-thirds of the difference is associated with the time period the assessment was conducted. Proposed as a possible explanation for the widening gap after 1992, has been the introduction of the Bayesian scaling procedure that has produced a bias into the effect size (Keeves et al., 2006). Overall, the test design of PISA has been shown to be a major negative factor in gender differences. Thus changes to redesign PISA are necessary in order to make it more equitable and trustworthy that policymakers can rely on. Widespread concern over the amount of effort that students put into the PISA were not considered to be a major factor in the validity of the test (Butler & Adams, 2007).

Fourthly, further examination of the reported trend from PISA indicates that boys' performance at the upper levels in reading (Level 5) has remained the same even though there has been an increase in the number of boys achieving at the low levels of reading. Similar findings are revealed in six longitudinal studies that show small but no significant differences in the gender gap in reading

achievement for average and above average students, but a growing and significant gap between boys and girls for low achieving students (Becker & Forsyth, 1990; Entwisle et al., 2007; Husain & Millimet, 2009; Martin & Hoover, 1987; Matthews et al., 2010; Robinson & Lubienski, 2011). Furthermore, boys from low SES (Edgerton et al., 2008; Husain & Millimet, 2009; Willms, 2004), and racial groups (Matthews et al., 2010) are at risk for underachieving in reading. Teacher (Entwisle et al., 2007; Robinson & Lubienski, 2011) and parent expectations (Entwisle et al., 2007; Willms, 2004) are shown to be potential contributors to the prevalent and growing difference among the lowest achieving boys. Why boys from low SES and from minority groups are not doing as well as the girls from low SES and minority groups needs further study. Solutions are complex and require home and school collaboration.

Lastly, a number of investigations have looked at the "root" cause for the underachievement of males and have experimented with possible solutions. The claim that boys view reading as a feminine activity has been countered by a number of studies (Sokal, 2010; Sokal & Katz, 2008; Sokal, Katz, Adkins, Gladu et al., 2005; Sokal et al., 2009; Steiner et al., 1981), however these studies focus on younger boys in elementary school. Presently, studies with older students were not found. Investigations into the gender of the teacher (Sokal & Katz, 2008; Sokal, Katz, Adkins, Gladu et al., 2005; Sokal et al., 2009), the approach to teach reading using technology (Sokal, & Katz, 2008) and the grouping of same-sex classes, do not increase reading achievement (Basilo, 2008; Blair & Sanford, 1999; Stotsky et al., 2010; Vrooman, 2009). A comparison between low-gender

gap schools and high-gender gap schools in Ontario revealed that when teachers demand high standards and make efforts to ensure that all students produced quality work, there was little difference in the gender gap (Klinger et al., 2009). A strong instructional leader was also considered to be critical in providing a positive learning culture.

The research studies presented clearly demonstrate that educators need to refrain from focusing on the underachievement of boys as a group, and instead concentrate on the sub-groups of students who are underachieving. There are low achieving boys and girls, and all need specific instructional support in order to achieve their potential.

Implications for Boys and Reading Comprehension (1981–2011)

Results from the 41 studies analyzed, synthesized and interpreted from 1981 to 2011 do not support the claim that **all** boys are underachieving in reading. In fact, many studies revealed that boys were not underachieving in reading. International studies such as PISA claim that there is a gender difference but have debatable test designs and questionable scaling procedures. Studies that have teased out achievement levels rather than attend only to overall averages have determined that boys from low SES and racial groups are underachieving. Middle and high performing boys continue to achieve comparatively to their female counterparts. The gap is widening between high achieving and low achieving boys. Although socio-economic status (SES) has been identified as a major factor in poor literacy performance in the past, girls from low SES and racial groups

perform better than boys from similar SES and racial groups. Therefore, the issue is not only SES and racial groups, but rather boys in low SES and racial groups.

There is a persistent perception by teachers and parents that girls are superior in reading, even when the data suggests differently. Boys and girls from low SES and some racial groups are generally underachievers. Influenced by media hype that continues to announce that boys are underachieving in literacy, and failure to access studies that report contrary findings contribute to the false construction that boys are not as proficient as girls in reading. Unfortunately, underachieving girls are rarely discussed because girls' achievement is used in an inclusive sense. Is there a relationship between teacher and parent perceptions and the underachievement of boys from low SES and racial groups? If so, what home and school factors need to be considered as we continue to wrestle with the issue of low achievement amongst boys from low SES and racial groups?

We live in a data rich society. The sheer volume of studies completed on the topic of boys' underachievement in literacy makes it difficult for teachers and policymakers to know which studies are reliable and trustworthy. There is a danger in taking studies at face value. Although statistical information is extremely valuable, the varied ways statistics are reported (e.g., percentages, rankings, significant differences, correlations) and interpreted can be misleading. One example is the nature of large-scale results, and the term "significant difference." Significant is often misconstrued to mean large when in actual statistical terms it means the likelihood of happening. In order to make good judgements when using data, it is critical to know how the data were collected and

analysed in the context of the study. When PISA 2000 reported that girls surpass boys on a reading test by 32 points in Canada, although this may appear large, further analysis revealed that this difference was out of 1,000 points. A 3.2 % variance renders a totally different interpretation than 32 points especially when people think it is out of 100 points. In addition, when analysis of PISA 2000 (Leitz, 2006a, 2006b) revealed that the procedures used to scale student scores produced inflated results that could account for two-thirds of the reported difference, the actual variance between girls and boys reading achievement is approximately 1%. Yet PISA 2000 leaves the reader with the impression that we are experiencing a boy crisis in reading. Such awareness of the manner in which statistical information is reported is important for policymakers, who rely on and engage in data-driven decisions.

Over the past 30 years of research in reading and gender, there has been a greater interplay between gender theory and reading research. This interplay is especially noticeable in recent studies that have focused on the root cause for the underachievement of boys in reading. Studies that have focused on role models, boy-friendly books, how boys view reading, brain-based activities, and the increased interest in same-sex classes are a few examples. Rather than the traditional disciplinary research of yesteryears, there has been an increase in the number of studies that use an interdisciplinary approach to advance research investigations of boys and reading. These studies make it clear that school boards need to refrain from simple solutions and policies, such as hiring more male teachers or buying more boy-friendly books. The underachievement of students

in reading requires complex solutions and cannot be solved by single solution scenarios.

With respect to the underachievement of boys, although seen in headlines in the newspapers, magazines and some journals, boys generally are not underachieving. With the exception of low achieving boys from low SES and visible minorities the issue of the underachievement of boys has been blown out of proportion and misrepresents the facts. However, if the re-emergence of boys' underachievement in reading has put a focus on good teaching practices in the classroom, then all students regardless of gender stand to benefit as educators realize that all children deserve good teaching and the best teaching and learning opportunities. Nonetheless, the importunate construction of and focus on the underachievement of boys in reading denies well-achieving boys their rightful standing and diminishes attention on low-achieving girls.

Chapter Eight:

Review of the Study, Summary of Results, Implications, and Limitations

Review of the Study

The persistent debate around the topic of the underachievement of boys in reading was my motivation to make sense of the discrepancies in the literature that date back more than a century. The background of the research findings on the perceived girl-boy gap in reading achievement had not been studied. Although research reviews were completed, no known historical synthesis and interpretation of the available research was available. Thus, my main purpose was to determine if there was a sex difference in reading achievement for the time period 1890 to present in North America. Since approximately half of the school population is comprised of boys, the undesirable consequences of boys underachieving in reading are extensive. If there is a girl-boy gap in reading, it is critical that we understand the extent of the gap and the reason(s) for the difference in order to provide recommendations for parents, classroom teachers, policymakers, and researchers. Before the publicity around a "boys' crisis" in reading takes an even stronger hold, I undertook a comprehensive examination of sex differences in reading achievement.

A review of the literature revealed basically two camps: researchers who claim that girls are superior in reading comprehension, and those who claim there is no difference (Leitz, 2006a, 2006b). Studies that take the stance that there is a sex difference in reading achievement refer either to biological or socio-cultural factors. Some researchers dispute the claim that boys are underachieving in

reading because they have found little evidence to verify a girl-boy gap. Other researchers have discovered discrepancies in test design and scaling procedures that have inflated the reading achievement differences in favour of girls.

Historically, studies on sex differences focused on biological differences between anatomical and developmental patterns between males and females (Havelock, 1904). Comparative studies on student achievement in school evolved from these discernable physical differences. The impact from the environment and life experiences was stressed in gender differences after the post-war years, and more recently there has been another return to a biological position. Studies on the brain purport that boys and girls have different brain structures and therefore learn in distinct and unique ways (James, 2007). Changes to the way literacy activities should be taught have been made including suggestions for brain gyms and the use of physical movement as well as technology to increase boys' engagement with, and interest in, reading (Millard, 1997; Smith & Wilhelm, 2002). These sorts of claims seem to be oversimplified and misrepresentative of the neuroscience research (Ansari, 2008).

In contrast, researchers from a sociological and cognitive perspective have disputed biological causes for gender differences and refer to environmental reasons for sex differences. The perspective that sex is biological and gender socially-constructed gained popularity in the 1950s and 1960s. In the active meaning-making of their worlds, individuals are influenced by what it means to be a boy or girl in the culture they live in (Bussey & Bandura, 1999; Martin & Ruble, 2009). Influenced by gender theory some researchers claim that the

difference in reading achievement between boys and girls is related to the difference in attitude toward reading and school engendered by societal and home factors (Dwyer, 1973; Gambell & Hunter, 1999; Millard, 1997). More recently, researchers have regarded the reciprocal nature between biological conditions, cultural contexts and personal choice, and the dichotomous variable of sex has since been debated for persons who do not fit into one biological category (Martin & Ruble, 2009).

Even though many studies have indicated a sex difference in reading achievement over the past century, other studies have found no or small differences between boys and girls. Discrepancies in the balance and design of large-scale international assessments that favour girls have been noted (Lafontaine & Monseur, 2009). The introduction of new sampling procedures and scaling methods used in national and international studies after 1992 inflated the gender gap between boys' and girls' reading achievement results (Leitz, 2006a, 2006b). The manner in which statistical information is reported, ranging from averages, correlations, and different levels of analyses is confusing to many readers. Misinterpretation of data has led to the overstatement of results and incorrect conclusions (Leitz, 2006a, 2006b; Mead, 2006; White, 2007). From the plethora of inconsistent research findings and range of methodologies and perspectives, it was confirmed that a systematic and critical appraisal of all the studies on the topic of gender and reading was needed in order to make sense of the controversial issue of the underachievement of boys in reading.

I chose to employ a systematic evidence-based historical interpretive (SEHI) methodology, an efficient, effective and logical way to assemble, and critically appraise the key findings from such a diverse and massive number of studies. Critical to this methodology was setting up criteria for the inclusion of research investigations. Based on the work of Lincoln and Guba (1985), and Creswell (2008), four criteria were considered essential for a trustworthy study: 1) evidence-based, 2) transparent and clearly articulated methodology, 3) data substantiated, and 4) a primary investigation. All relevant search engines were used to systematically examine the major databases. In addition, I used the references from retrieved studies to locate other investigations. Snowballing was particularly necessary in historical searches, as keywords were not consistently considered in the existing electronic databases. I kept track of the number of hits and studies that were previewed and created a spreadsheet of all of the synthesized studies, and organized them by study type (descriptive, explanatory and experimental). I read through each research category to determine major themes from both 1) within a study, and 2) cross-study analysis and synthesis to determine the major interpretations and implications for the study.

A systematic review of over 3,000 studies from over 12 decades was completed. This search rendered 78 studies that met the criteria outlined in Chapter One. Further analysis of the identified studies revealed that 17 investigations, although included in this study, had conflicting conclusions that could not be corroborated from the data provided. An analysis, synthesis, and interpretation of the remaining 61 studies identified as being trustworthy covered

four time periods (1890–1920; 1921–1945; 1946–1980; 1981–2011). The findings are discussed in the preceding chapters with a summary of the results, implications, and limitations presented in subsequent sections.

Summary of Results

Persuasive and powerful-sounding arguments have been made in the news that report boys are lagging behind girls in reading achievement, and that boys are being short-changed. This manufactured hype is riddled with unsubstantiated theories for ways to deal with the "boy crisis" that are based either on superficial evidence or misinterpretations. This girl-boy debate in reading achievement has not benefited either gender, and in fact has distracted educators from other issues such as poverty and race (Mead, 2006). Based on my comprehensive identification, examination, and interpretation of the evidence-based studies on the topic of gender differences in reading over the past 120 years, I have come to the conclusion that the publicity around boys' poor results in reading comprehension overstates reality. For the most part, boys are doing as well as girls in reading achievement and have done so for over the past century.

The concern for boys underachieving in school was first announced around the turn of the 20th century when laws were instituted that required every child to attend school. Three studies procured between 1890 and 1920 revealed sex differences in achievement. Girls were superior in spelling tests (Earle & Thorndike, 1903), and boys in mathematics and motor skills (Pressey, 1918). Also, it was noted that boys were not progressing as well as girls in school (Ayres, 1909). Assessment of achievement in school prior to WWI was

predominately conducted from a behaviourist perceptive where observable outcomes were quantified. Since reading comprehension involves cognitive processes, the internal workings of the mind were not visible or open for inspection, and therefore, were not measured. One of the first attempts to measure reading comprehension was conducted by Thorndike (1917) who tried to characterize what was going on in the brain. In a quest to measure reading comprehension, researchers realized that reading was a complex process and they fell short on valid indices to measure the actual process. It was not until after WWI that the first reading comprehension assessments were conducted (Sarroub & Pearson, 1998). However, indices used to measure the processes of reading comprehension "all interpose some other cognitive and/or motor task (marking, writing, speaking, or reflecting) between the act of comprehension and the evidence of its occurrence" (Sarroub & Pearson, 1998, p. 98). Gender differences in achievement in the early years of compulsory schooling were measured, however reading comprehension was not measured.

Following the preliminary work of Earle and Thorndike, (1903), Lincoln (1927) completed one of the first comprehensive studies on sex differences. In terms of reading achievement, Lincoln concluded that although girls could read faster, there was no conclusive evidence to support the claim that girls were superior to boys in reading comprehension. The perception amongst educators during the 1930s and 1940s was the notion that girls were superior to boys in achievement in elementary school and that boys caught up to girls by junior high school and high school. Suggestions were made that boys should start school

later in order to give them a level playing field with girls. This perceived superiority of girls in achievement in the elementary grades was credited to girls' earlier maturation rate. Although the relationship between physical maturity and mental growth was hypothesized, in later years this widely believed explanation was disputed and abandoned (Maccoby, 1966, p. 38). More definitive was the evidence that supported the relationship between IQ and reading achievement (Millard, 1940). Sex differences were studied from a variety of perspectives, yet there was no evidence to support the conclusion that boys underachieved in reading comprehension between 1920–1945.

From the 26 studies out of the hundreds cited during the years 1945–1980, only 9 met the robust standards outlined in my methodology. Two studies challenged the assumption that boys underachieve in reading in the primary grades. Even though girls performed better than boys on many of the pre-reading subtests, the differences were often small or inconsistent (Dykstra & Tinney, 1969; Prescott, 1955) and did not measure reading comprehension. Five studies indicated that when differences in reading achievement were correlated with intelligence and/or chronological age, no sex differences in reading achievement were found (Anderson, Hughes & Dixon, 1956, 19567; Clark, 1959; Parsley et al., 1963; Sinks & Powell, 1965). These studies implied a biological cause for differences in reading achievement unrelated to sex, but were tied to an innate individual difference. In contrast, two cross-cultural studies were conducted and established that the variances in reading achievement were linked to socio-economics rather than sex or biological factors (Finn, 1980; Thorndike, 1973).

These findings suggested environmental differences between countries. These studies typify the varied positions on the nature-nurture debate of human development. The extent of how genetics and culture interacted with human development was not yet reported in the literature. In summary, the results of these nine studies overwhelmingly revealed no evidence to support female superiority over males in reading during the time period 1945–1980.

Results of 41 studies between 1980 to present were analyzed, synthesized and interpreted. Nine studies indicated no significant differences in reading achievement between boys and girls in elementary (Flynn & Rahbar 1994; Harper & Pelletier, 2008; Lummis & Stevenson, 1990; Phillips et al., 2002; Quinlan, 1996; Wargacki, 2008), middle (MacFarlane, 2001), and high school (Hogrebe et al., 1984; White, 2007).

Ten national and international studies were examined. Reported results from the PISA studies (OECD 2001, 2004b, 2007, 2010) that indicate girls are superior in reading comprehension were contested on the basis of test design (Lafontaine & Monseur, 2009). Four studies focused on the questionable validity of large-scale international tests. Findings indicate that the introduction of the Bayesian scaling procedure in PISA after 1992 produced a bias in the effect size in the reported gender differences in favour of girls (Lietz, 2006a, 2006b).

Eight studies, including six longitudinal studies, identified that although boys from average and high reading achievement groups have similar results to girls, low performing boys, particularly from low socio-economic and ethnic groups, are underperforming (Becker & Forsyth, 1990; Entwisle et al., 2007;

Husain & Millimet, 2009; Martin & Hoover, 1987; Matthews et al., 2010; Robinson & Lubienski, 2011). Indices suggest that the gap is widening and that boys from lower SES and boys of colour are most at risk.

Ten investigations sought to explain why boys were underachieving in reading. Three studies explored whether boys conceived reading to be a feminine activity and found no relationship between views of reading and motivation. Most boys regard reading as a gender-neutral activity (Sokal, 2010; Sokal, Katz, Adkins, Gladu et al., 2005; Steiner et al., 1981). Inquiries as to whether the sex of the teacher had an effect on boys' achievement in reading were analyzed and found to be moot (Sokal, Katz, Adkins, Grills, et al., 2005, Sokal & Katz, 2008; Sokal et al., 2009). One investigation looked into the possible use of computers to teach reading to improve reading achievement and found no differences in boys' or girls' achievement (Sokal & Katz, 2008). Other studies examined the benefits of single-gender classes and found they had no effect on reading achievement for boys (Basilo, 2008; Blair & Sanford, 1999; Stotsky et al., 2010). A case study (Klinger et al., 2009) conducted in Ontario compared the approaches used in schools between high-gender gap schools and low-gender gap schools. Their results revealed that when teachers demand high standards and make efforts to ensure that all students produce quality work, boys achieve equally as well as girls. Good teaching practices and high expectations benefit all students, both boys and girls. In summation, the reasons that boys underachieve in reading, and possible solutions are multi-dimensional and complex. The only select groups of boys underachieving in reading comprehension borne out by the research

evidence are those from low SES backgrounds and boys of colour, and the combination of both is serious. The overall findings from these 61 evidence-based studies and their implications for parents, teachers, researchers, and policymakers, are explained in more detail next.

Implications and Suggestions for Further Research

Analyses of the collected data from over a century on the issue of a perceived boy-girl gap in reading comprehension achievement has generated some significant conclusions and recommendations for further research. First, the evidence is quite clear that not all boys are underachieving in reading. Male and female are not universal constructs. Boys and girls constitute a diverse population and many factors play a role in understanding how they vary. Subgroups of boys from low SES and boys of colour have lower achievement in reading achievement than girls. Albeit SES has been identified in the past as the single most important contributor to poor performance in literacy (OECD, 2001; Thorndike, 1973), girls from low SES and colour perform better than boys from low SES and colour. Both boys and girls from low SES and minority groups perform less well than those from high SES and majority populations. Evidence points to teacher (Entwisle et al., 2007; Robinson & Lubienski, 2011) and parent expectations (Entwisle et al., 2007; Willms, 2004) as potential contributors to the prevalent and growing gender gap in reading comprehension achievement. There is no evidence to support either the claim for a complete overhaul of our school system or to hire more male teachers (Sokal & Katz, 2008; Sokal, Katz, Adkins, Gladu et al, 2005; Sokal et al., 2009); to offer boys' only classes (Basilo, 2008; Blair & Sanford,

1999; Stotsky et al., 2010; Vrooman, 2009); to include more technology (Sokal & Katz, 2008); and, to include more boy-friendly titles (Sokal et al., 2009; Sokal & Katz, 2008; Sokal, Katz, Adkins, Gladu et al., 2005; Sokal, Katz, Adkins, Grills, et al., 2005). These findings confirm the shortcomings of one variable considerations and over-simplistic solutions to a complex phenomenon.

Second, to shine the spotlight on boys as a group presupposes that all boys are underachieving in reading and that all girls are doing well, thereby overstating the problem for boys while ignoring girls. Claims that there is a gender gap in reading achievement has misrepresented the facts with serious implications for both groups. Evidence from low-gender gap schools reveal that when teachers have high expectations for their students with concomitant support for struggling learners, all students achieve (Klinger et al., 2009). In addition, there is evidence that more students have greater scholastic success when schools have strong instructional leaders who support an ethos of a positive learning culture and an optimistic outlook for all learners (Klinger et al., 2009). It is critical to use these findings in order to advance valid perceptions about boys and reading. Boys achieve as well as girls in reading, and students who experience poor reading performance can achieve with proper supports. In order for boys to reach their potential in reading comprehension achievement and ultimately school achievement in general, teachers and parents must consider and act promptly on the fact that such a perception is untenable. Otherwise, teachers and parents will continue to perpetuate a widely spread myth at the expense of their students and children. Schools must work immediately with families to foster the

understanding that boys are capable and good readers when given a chance. Continued research is needed on the ways home and schools can positively work together. Our goal should not be to quantify and label, but rather to recognize each learner's potential, to avoid stereotypes, and refrain from the pigeonhole categorizations of students.

Third, the inconsistencies in the analyses of findings, and the misuse of reported statistical information from 1890 to present have contributed to, and perpetuated the myth of a boy-girl gap. An erroneous assumption based on literacy-related subtests and not reading comprehension subtests from the turn of the 19th century, has consistently mislead subsequent researchers because they accepted without question the general conclusion that since girls were superior in word-related activities, they were also more advanced than boys on reading comprehension. As a result, the mistaken belief that girls' are superior to boys in reading has become a common opinion amongst many educators and parents, regardless of counter-evidence that shows otherwise. Boys' underachievement in reading has been over-reported, while evidence of small or no differences have been under-reported. This uneven reporting of the facts has led to the false impression that boys are at risk, which if not debunked will continue to have long-term consequences for boys as well as girls (given that some girls are not getting the attention that they need). Evidence disproves that boys are underachieving in reading. It is time to publicize, challenge, and educate all about this falsehood!

The misinterpretation of reported data and the terms used to describe findings, including averages without consideration of ranges and statistical

differences, have also contributed to the widespread belief that girls are superior to boys in reading. Misapplied reasoning that has overstated the underachievement of boys in reading comprehension is partly due to the way statistical information is understood. An "average" does not describe the variance in data. If, on average, boys are underachieving in reading comprehension, the erroneous conclusion is that all boys are in difficulty. Averages that are statistically significant are also misconstrued. The term statistical significant does not mean a large difference, but the chance that the reported information will be replicated. Unfortunately, averages that are statistically different are interpreted as being large and important when in actuality the difference can be small but replicable. In an attempt to make statistical information more available, unfortunately the overwhelming result has been even greater confusion. Moreover, in the case of boys' achievement, the use of new statistical scaling methods in large-scale assessments has continued to inflate the difference in scores between the girls and boys, even though in reality analyses of the data reveal that boys and girls differ by approximately only 1%. Misinterpretation and misrepresentation of statistical information is particularly damaging for policymakers, researchers, and administrators who widely use evidence-based studies such as national and international reports to determine trends in education to set new directions. In education, we need to press on for the best evidence and guard against using information that is false. Misinterpretations must be challenged. To build trust in research findings, educators need support to discern

valid and reliable studies and to find constructive ways to avoid being manipulated by misinformation.

Fourth, the increased volume of information and research data on any one topic makes it difficult to assess trustworthy findings from popular culture. The need to focus on evidence-based, primary studies is required when making ethical and informed educational decisions. Research and reports that quote findings from secondary sources run the chance of misrepresenting the facts since it is difficult to report the conclusions without reading the complete original study. An example of this misrepresentation of facts has been documented in Chapter 7 from the analysis of Elley's (1992) study that showed no conclusive evidence to support the claim that boys are underachieving in reading. Researchers who to refer to Elley's (1992) findings based on secondary sources to support the claim that boys are underachieving in reading have **not** read Elley's (1992) study in its entirety and have therefore misrepresented the facts. In addition, the changing nature of methodologies and the varied ways that data is reported makes literature reviews less reliable and points to the necessity for systematic research and meta-analyses in education, especially when contentious issues arise and unconscionable consequences arise, such as in this case for half of the school population.

Fifth, a prevalent and recurring theme in all the research on gender differences is the nature-nurture debate over the cause for gender differences. Inherent biological factors and environmental conditions continue to be considerations when studying gender differences in reading. Historically, we

have moved from a focus on biological factors (up to the WWII) to socio-cultural (WWII to 1980s) and to a present day view that considers the interactions between the two factors (1980s to present). Even though new theories and insights have changed how genetics and social factors influence gender differences, the degree to which innate differences and environmental factors influence the individual remain under research and debate.

Six, the interplay between the field of reading and gender development, and the growth in interdisciplinary research that combines research practices, perspectives and theories supports the need for continued integration of various fields of knowledge to solve complex problems. Psychologists initially studied reading and gender in the late 1800s and early 1900s, which contributed to a growing field of reading and gender development in the 1930s and 1940s. However, for the most part, these two disciplines remained as separate fields until quite recently. Combined reading research and gender development theory shows promise for continued interdisciplinary collaboration that is much needed in order to provide deeper understanding and interconnectedness of ongoing and future problems.

Upon reflection of the massive number of studies analyzed from over a century, I am left with the questions of how boys' achievement in reading is different today than from the past. In 1909, Ayres released some disturbing news that boys were being retained and dropping out of school more than girls. It was alarming news for a new publicly-funded school system that hoped to change society. Now a century later the dismal news of a "boys' crisis" rears its ugly

head again. Although there is little evidence to support the underachievement of boys in reading over the last 120 years or so, the obvious question that remains unanswered is why has this anxiety about boys' reading achievement resurfaced? Although suggestions were made regarding a gender gap in literacy from various research studies over the past 12 decades, not since the turn of the 19th century has there been so much urgency around boys' underachievement. Are there any connections to what was happening in North America some 120 years ago that has resurfaced again today, or is the phenomenon of failing boys the result of a backlash from a focus on girls and their stated underachievement in mathematics and science from the 1980s and 1990s? Since it has been established that the underachievement of boys in reading has been overstated, how much merit is there to the claim that we are in the midst of a boys' crisis? These are just a few of the many topics for additional research. More certain from this study is that poverty and race are not only major factors in the underachievement of all students in reading, but in particular our boys.

Limitations

This research was limited to studies researched in North America. It is possible that boys are underachieving in reading in other countries worldwide. Although international achievement studies were analyzed and synthesized, the results that pertained to North America were the focus of this dissertation.

Although I used a systematic approach to secure all relevant research studies on gender differences in reading, I do not assume that every article or study on the topic was procured, read and analyzed. It is possible that although a

thorough approach was taken to secure every investigation on the topic, there may have been studies that were missed.

The topic of the underachievement of boys in reading is part of the general topic of the underachievement of boys. This research does not presume to answer the general topic of a boys' crisis in education. I have limited the topic to gender and reading achievement.

I made every effort to remain unbiased in my representation of the findings of each study. Studies that supported the underachievement of boys as well as those that disputed the claim were analyzed and synthesized, and a fair and reasonable presentation on all studies was made. Although I tried to maintain an objective stance on the topic, I am aware that subjectivity is part of any social research. Nonetheless, every primary source was studied and in all cases citations were provided. In conclusion, the persistent and contradictory comparative claims of boys' and girls' reading achievement: A historical interpretive approach was a massive undertaking and unequivocally shows the staying power of unfounded beliefs and misconceptions despite the research evidence.

References

Adams, R. (Ed.). (2005). *PISA 2003 technical report*. Paris: OECD Publications.

Adams, R. (Ed.). (2009). *PISA 2006 technical report*. Paris: OECD Publications.

Adams, R., & Wu, M. (Ed.). (2002). *PISA 2000 technical report*. Paris: OECD Publications.

Alberta Education. (1986). *Alberta diagnostic reading program*. Edmonton, AB: Author. Retrieved from www.lrc.education.gov.ab.ca

Alexander, P., & Fox, E. (2008). A historical perspective on reading research and practice. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 33–68). Newark, DE: International Reading Association.

Allington, R., & McGill-Franzen, A. (2004). Looking back. Looking forward: A conversation about teaching reading in the 21st century. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 5–32). Newark, DE: International Reading Association.

Anasari, D. (2008). The brain goes to school: strengthening the education-neuroscience connection, *Education Canada,* 48 (4), 6-10

Anastasi, A. (1958). *Differential psychology* (3rd ed.). New York, NY: Macmillan.

Anderson, I., Hughes, B., & Dixon, W. (1956). Age of learning to read and its relation to sex, intelligence, and reading achievement in the sixth grade. *The Journal of Educational Research, 49*(6), 447–453.

Anderson, I., Hughes, B., & Dixon, W. (1957). The rate of reading development and its relation to the age of learning to read, sex, and intelligence. *The Journal of Educational Research, 50*(7), 481–494.

Ayres, L. (1909). *Laggards in our schools: A study of retardation and elimination in city school systems*. New York, NY: Charities Publication Committee.

Baldwin, D. (2008). *Teachers, students and pedagogy: Selected readings and documents in the history of Canadian education*. Markham, ON: Fitzhenry and Whiteside.

Barry, A. (2007). David Harris Russell (1906–1965). A man who loved learning and teaching. In S. Israel & J. Monaghan (Eds.), *Shaping the reading field* (pp. 375–395). Newark, DE: International Reading Association.

Basilo, E. (2008). *Case study of reading performance of male students and the single-sex classroom* (Unpublished doctoral dissertation). University of Central Florida, Orlando.

Becker, D., & Forsyth, R. (1990, April). *Gender differences in academic achievement in Grades 3 through 12: A longitudinal analysis*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Belser, D. (1926). The reading interests of boys: A committee report. *The Elementary English Review, 3*(9), 292–296.

Bem, S. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88*(4), 354–364.

Bereday, G. Z. F., & Lauwerys, J. A. (Eds.). (1963). *The education and training of teachers: The yearbook of education*. London: Evans Brothers.

Berman, I., & Bird, C. (1933). Sex differences in speed of reading, *Journal of Applied Psychology, 17*, 221–226

Binkley, M., Phillips, L., & Norris, S. (1995). Creating a measure of reading instruction. In M. Binkley, K. Rust, & M. Winglee (Eds.), *Methodological issues in comparative educational studies* (pp. 193–229). Washington, DC: Office of Educational Research and Improvement.

Blair, H., & Sanford, K. (1999, April). *Single-sex classrooms: A place for transformation of policy and practice*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, QC.

Blakemore, J., Berenbaum, S., & Liben, L. (2009). *Gender development*. New York, NY: Psychology Press.

Bohannon III, J., & Bonvillian, J. (2009). Theoretical approaches to language acquisition. In J. Berko Gleason & N. Bernstein Ratner (Eds.), *The development of language* (pp. 227–284). Boston, MA: Allyn & Bacon/Pearson Education, Inc.

Boyd, W., & King, E. J. (1965). *The history of Western education.* London: A. & C. Black.

Buchner, E. (1909). Review of the psychology and pedagogy of reading, with a review of the history of reading and writing and of methods, texts, and hygiene of reading. *Psychological Bulletin, 6*(4), 147–150.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676–713.

Butler, J., & Adams, R. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement, 8*(3), 279–304.

Canadian Council on Learning. (2007). *State of learning in Canada: No time for complacency*. Retrieved from http://www.ccl-cca.ca/ccl

Cardon, B. (1968). Sex differences in school achievement. *The Elementary School Journal, 68*(8), 427–434.

Carroll, J. (1968). Forward. In J. Huey, *The psychology and pedagogy of reading* (pp. vii–xii). Cambridge, MA: M. I. T. Press.

Carroll, M. W. (1948). Sex differences in reading readiness at the First grade level. *Elementary English, 25*, 370–375.

Clark, W. (1959). Are there significant ability and achievement differences? *The Phi Delta Kappan International, 41*(2), 73–76.

Cole, L. (1938). *The improvement of reading: With special reference to remedial instruction*. New York, NY: Farrar and Rinehart.

Comacchio, C. (2000). Foreword. In N. Sutherland *Children in English-Canadian society: Framing the twentieth-century consensus*. Waterloo, ON: Wilfrid Laurier University Press

Commins, W. D. (1928). More about sex differences. *School and Society, 28*, 599–600.

Cremin, L. (1965). *The genius of American education.* New York, NY: John Wiley.

Creswell, J. (2008). *Educational research: Planning, conduction, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson.

Criscuolo, N. (1968). Sex influences on reading, *The Reading Teacher, 21*(8), 762–764.

Cunningham, H. (1995). *Children and childhood in Western society since 1500*. London: Longman.

Degler, C. (1967). *The age of the economic revolution 1876–1900*. Glenview, IL: Scott, Foresman and Company.

Denzin, N., & Lincoln, Y. (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dewey, J. (1903). Democracy in education. *The Elementary School Teacher, 4*(4), 193–204.

Dewey, J. (1916). *Democracy and education*. New York, NY: Free Press.

Dolch, E. (1931). *The psychology and teaching of reading*. Boston, MA: Ginn and Co.

Dinnan, J. (1971). *The Oral P/S Language Inventory*, Unpublished test, Athens, GA: University of Georgia.

Downing J., Dwyer, C., Feitelson, D., Jansen, M., Kemppainen, R., Matihaldi, H., Thomson, D. (1979). A cross-national survey of cultural expectations and sex role standards in reading, *Journal of Research in Reading*, *2*(1), 8–23.

Downing, J., & Thomson, D. (1977). Sex role stereotypes in learning to read. *Research in the Teaching of English, 11*(2), 149–155.

Doyle, D. P. (1991). A study in change: Transforming the Charlotte-Mecklenburg schools. *Phi Delta Kappan*, *74*(7), 534–539.

Durrell, D. (1940). Trends in the teaching of reading. In H. Bixby (Ed.), *Trends in the teaching of reading: Summary report.* New Brunswick, NJ: Rutgers University.

Dwyer, C. (1973). Sex differences in reading: An evaluation and a critique of current theories. *Review of Education Research*, *43*(4), 455–467.

Dwyer, E. (1980). Analysis of reading achievement, listening comprehension, and paradigmatic language of selected second grade students by race and sex. *The Southern Journal of Educational Research*, *14*, 205–218.

Dykstra, R., & Tinney, R. (1969). Sex differences in reading readiness - first-grade achievement and second-grade achievement. *Reading and Realism*, *13*, 623–628.

Earle, E., & Thorndike, E. (1903). A sex difference in spelling ability, *Columbia University Contributions to Philosophy, Psychology and Education*, *6*(2), 47-49.

Edgerton, J., Peter, T., & Roberts, L. (2008). Back to the basics: Socio-economic, gender, and regional disparities in Canada's educational system. *Canadian Journal of Education*, *31*(4), 861–888

Education Quality and Accountability Office. (EQAO). (2007). *Assessment of reading, writing and mathematics, junior division (Grades 4–6): Framework.* Toronto, ON: Author. Retrieved from http://www.eqao.com/pdf_e/08/6e_Framework_07_web.pdf

Elley, W. (1992). *How in the world do students read?* The Hague, NL: International Association for the Evaluation of Educational Achievement.

Ellis, H (1904). *Man and woman: A study of human secondary sexual characters*. London: Walter Scott

Entwisle, D., & Alexander, K. (1996). Family type and children's growth in reading and math over the primary grades. *Journal of Marriage and the Family*, 58, 341-355.

Entwisle, S., Alexander, K., & Olson, L. (2007). Early schooling: The handicap of being poor and male. *Sociology of Education*, *80*(April), 114–138.

Feldt, L., Forsyth, R., & Alnot, S. (1986). *Iowa tests of educational development, forms X-8 and Y-8*. Iowa City, DE: The University of Iowa.

Feldt, L., Forsyth, R., & Lindquist, E. (1979). *Iowa tests of educational development, forms X-7 and Y-7*. Iowa City: The University of Iowa.

Finn, J. (1980). Sex differences in educational outcomes: A cross-national study. *Sex Roles*, *6*(1), 9–26.

Flaherty, R., & Anderson, H. (1966). Boys' difficulty in learning to read. *Elementary English*, *43*,471–472.

Flesch, R. (1955). *Why Johnny can't read*. New York, NY: Harper & Brothers.

Flynn, J., & Rahbar, M. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools*, *31*(1), 66–71.

Fox, W., & Thorndike, E. (1903). The relationships between the different abilities involved in the study of arithmetic. *Columbia University Contributions to Philosophy, Psychology and Education*, *6*(2), 32–40.

Gambell, T., & Hunter, D. (1999). Rethinking gender differences in literacy. *Canadian Journal of Education, 24*(1), 1–16.

Gambell, T., & Hunter, D. (2000). Surveying gender differences in Canadian school literacy. *Journal of Curriculum Studies*, 32(5), 689-719.

Gates, A. I. (1917). Experiments as the relative efficiency of mean and women in memory & reasoning. *Psychological Review*, 24(2), 139-146.

Gates, A. I. (1927a). Studies of phonetic training in beginning reading. *Journal of Educational Psychology, 18*(4), 217–226.

Gates, A. I. (1927b). *The improvement of reading: A program of diagnostic and remedial methods.* New York, NY: Macmillan.

Gates, A. I (1928). *New methods in primary reading*. New York, NY: Teachers college, Columbia University.

Gates, A. I. (1935). *The improvement of reading: A program of diagnostic and remedial methods* (Rev. ed.). New York, NY: Macmillan.

Gates, A. I. (1947). *The improvement of reading: A program of diagnostic and remedial methods* (3rd ed.). New York, NY: Macmillan.

Gates, A. I. (1961). Sex differences in reading ability. *Elementary School Journal, 61*(8), 431–434.

Gilstad, J. (1985). William S. Gray (1885–1969): First IRA president. *Reading Research Quarterly, 20*(4), 509–511.

Goodman, K., & Goodman, Y. (1980). Learning to read is natural. In L. B. Resnick, & P.A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 1, pp. 137–154). Hillsdale, NJ: Erlbaum.

Gorski, P. C. (1999). *A brief history of multicultural education*. Retrieved from http://www.edchange.org/multicultural/papers/edchange_history.html

Gottfried, A. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Education Psychology*, 77, 631–645.

Gray, W. (1929). *Summary of reading investigations (July 1, 1927, to June 30, 1928).* [Supplement No. 4, Continuation of Monograph No. 28]. Chicago, IL: University of Chicago, Department of Education.

Gray, W. (1933). *Improving instruction in reading: An experimental study*. [Supplementary Educational Monographs in conjunction with The School Review and The Elementary School Journal, Number 40]. Chicago, IL: University of Chicago.

Gray, W. (1935). Special methods and psychology of the elementary-school subjects: Reading. *Review of Educational Research*, *5*(1), 54–69.

Gray, W. (1937). A decade of progress. In G. Whipple (Ed.), *The thirty-sixth yearbook of the National Society for the Study of Education: Part 1, the teaching of reading: A second report* (pp. 5–22). Bloomington, IL: Public School Publishing.

Gray, W. (1956). *The teaching of reading and writing: An international survey*. Paris: UNESCO.

Great Britain. Committee on English in the Educational System of England. (1921). *Headmaster of England's rugby school in the board of education's Newbolt report*. London: H.M.S.O. Retrieved from http://www.archive.org/stream/teachingofenglis00greaiala

Greenwald, M. (1980). *Women, war and work: The impact of World War I on women in the United States*. Westport, CT: Greenwood Press.

Gulick, L., & Ayres, L. (1913). *Medical inspection of schools*. New York, NY: Russell Sage Foundation.

Haertel, E., & Calfee, R. (1983). School achievement: Thinking about what to test. *Journal of Educational Measurement*, *20*(2), 119–132.

Harper, S., & Pelletier, J. (2008). Gender and language issues in assessing early literacy. *Journal of Psychoeducational Assessment*, *26*(2), 185–194.

Havelock, E. (1904). *Man and woman; a study of human secondary sexual characters*. London: Scott

Heath, S. (1996). The sense of being literate: Historical and cross-cultural features. In R. Barr, M. Kamil, P. Mosenthal, & D. Pearson (Eds.), *Handbook of reading research volume II* (pp. 3–26). Hahwah, NJ: Lawrence Erlbaum Associates.

Henck, W., & Melnick, S. (1995). The reader self-perception scale (RSS): A new tool for measuring how children feel about themselves as readers. *The Reading Teacher*, *48*, 470–482.

Hieronymus, A., Lindquist, E., & Hoover, H. (1978). *Iowa test of basic skills, forms 7 & 8*. Chicago, IL: Riverside.

Hiscox, M., & Rodger, R. (n.d.). Kahlmann-Anderson tests (8th ed.). *All mental measurements yearbook.* Buros: Tests Reviews Online. Retrieved from http://buros.org/

Hobsbawn, E. (1992). Mass-production traditions: Europe, 1870–1914. In E. Hobsbawn, & T. Ranger (Eds.), *The invention of tradition*. New York, NY: Cambridge Press.

Hobsbawn, E. (1994). *The age of extremes: A history of the world, 1914–1991*. New York, NY: Pantheon.

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.

Hollingworth, L. (1914). Variability as related to sex differences in achievement. *The American Journal of Sociology*, *19*(4), 510–530.

Huey, E. (1898). Preliminary experiments in the physiology and psychology of reading. *American Journal of Psychology*, *9*(4), 575–586.

Huey, E. (1908/1968). *The psychology and pedagogy of reading*. Cambridge, MA: M. I. T. Press.

Husain, M., & Millimet, D. (2009). The mythical 'boys crisis'? *Economics of Education Review*, *28*(1), 38–48.

James, A. (2007). *Teach the male brain: How boys think, feel and learn in school*. Thousand Oaks, CA: Corwin.

Janovicek, N., & Parr, J. [Eds.]. (2003). *Histories of Canadian children and youth*. New York, NY: Oxford University Press.

Jantz, R. (1974). The effects of sex, race, IQ and SES on the reading scores of sixth graders for both levels and gain in performance. *Psychology in the Schools*, *11*(1), 90–94.

Johnson, D. (1973–74). Sex differences in reading across cultures. *Reading Research Quarterly*, *9*(1), 67–86.

Johnson, D. (1976). Cross-cultural perspectives on sex differences in reading, *The Reading Teacher*, *29*(8), 747–752.

Jordon, A. M. (1937). Sex differences in mental traits. *High School Journal*, 20, 254–261.

Judd, C. (1918a). *Introduction to the scientific study of education*. Boston, MA: Ginn. [electronic resource]

Judd, C. (1918b). *Reading: Its nature and development*. Chicago, IL: University of Chicago.

Judd, C. (1922). Relation of school expansion to reading. *The Elementary School Journal*, *23*(4), 253–266.

Judd, C., & Buswell, G. (1922). *Silent reading: A study of the various types*. Chicago, IL: University of Chicago Press.

Juel, C. (1988). Learning to reading and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, 80, 437–447.

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

Katz, M. (1976). *A history of the compulsory education laws*. Bloomington, IN: The Phi Delta Kappa Educational Foundation.

Keeves, J., Lietz, P., Gregory, K., & Darmawan, I. (2006). Some problems in the analysis of cross-national survey data. *International Education Journal, 7*(2), 110–126.

King, I. (1955). Effect of age of entrance into Grade 1 upon achievement in elementary school. *The Elementary School Journal*, *55*(6), 331–336.

Klein, H. (1977). Cross-cultural studies: What do they tell about sex differences in reading? *The Reading Teacher*, *30*(8), 880–886.

Klinger, D., Shulha, L., & Wade-Woolley, L. (2009). *Towards an understanding of gender difference in literacy achievement.* Toronto, ON: Education Quality and Accountability Office (EQAO).

Kohlberg, L. (1966). A cognitive-development analysis of children's sex-role concepts and attitudes. In E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.

Konski, J. (1951). An investigation into differences between boys and girls in selected reading readiness areas and in reading achievement. *Microfilm Abstracts*, 11, 919–920.

Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. New York, NY: Longman Publishing Group.

Lafontaine, D., & Monseur, C. (2009). Gender gap in comparative studies of reading comprehension: To what extent do the test characteristics make a difference? *European Educational Research Journal*, *8*(1), 69–79.

Lauritzen, C. (2007). William Scott Gray (1885-1960): Mr. Reading. In S. Israel & J. Monaghan (Eds.), *Shaping the reading field* (pp. 307–326). Newark, DE: International Reading Association.

Lears, T. (2009). *Rebirth of a nation: The making of modern America, 1877–1920*. New York, NY: HarperCollins.

Lietz, P. (2006a). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation*, 32, 317–344.

Lietz, P. (2006b). Issues in the change in gender differences in reading achievement in cross-national research since 1992: A meta-analytic view. *International Education Journal*, *7*(2), 127–149.

Lincoln, E. A. (1927). *Sex differences in the growth of American school children.* Baltimore, MD: Warwick and York.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Lindquist, E. (1940). *Statistical Analysis in Education Research.* Boston, MA: Houghton Mifflin.

Luke, A. (1987). Making Dick and Jane: Historical genesis of the modern basal reader. *Teachers College Record*, *89*(1), 91–116.

Lummis, M., & Stevenson, H. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, *26*(2), 254–263.

Maccoby, E. (Ed.). (1966). *The development of sex differences*. Stanford, CA: Stanford University Press.

Maccoby, E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press.

Maccoby, E. (2002). Gender and group process: A developmental perspective. *Current Directions in Psychological Science*, 11, 54–58.

MacFarlane, L. (2001). *Gender differences in reading achievement and early literacy experiences* (Unpublished master's thesis). Kean University, Union, NJ.

MacGinitie, W., Kamons, J., Kowalski, R., MacGinitie, R., & MacKay, T. (1980). *Gates-MacGinitie reading tests*. Toronto, ON: Nelson Canada.

Manzo, A., & Manzo, U. (1990). *Content area reading: A heuristic approach*. Toronto, ON: Merrill.

Martin, C., & Ruble, D. (2009). Children's search for gender cues: Cognitive perspective on gender development. In L. Liben (Ed.), *Current directions in developmental psychology* (2nd ed.) (pp. 165–172). Boston, MA: Pearson.

Martin, J., & Hoover, H. (1987). Sex differences in educational achievement: A longitudinal study. *The Journal of Early Adolescence, 7*(1), 65–83.

Matthews, J., Kizzie, K., Rowley, S., & Cortina, K. (2010). African Americans and boys: Understanding the literacy gap, tracing academic trajectories, and evaluating the role of learning-related skills. *Journal of Educational Psychology*, *102*(3), 757–771.

McBroom, M. (1938). Desirable scope of the school program in reading in practises in reading in the elementary school, bulletin of the department of elementary school principals. *National Education Association*, *17*(7), 235–246.

McGill-Franzen, A. (1993). *Shaping the preschool agenda: Early literacy, public policy, and professional beliefs*. Albany, NY: State University of New York Press.

Mead, S. (2006). *The truth about boys and girls: The evidence suggests otherwise*. Retrieved from http://www.educationsector.org/usr_doc/ESO_BoysAndGirls.pdf

Metropolitan Achievement Test. (1970). *Metropolitan achievement test*. New York, NY: Harcourt, Brace Jovanovich.

Millard, C. (1940). The nature and character of pre-adolescent growth in reading achievement. *Child Development*, *11*(2), 71–114.

Millard, E. (1997). *Differently literate: Boys, girls and the schooling of literacy*. London: Falmer Press.

Monaghan, J. (2007). Scientific research and progressive education: Contexts for the early reading pioneers, 1870–1956. In S. Israel & J. Monaghan (Eds.), *Shaping the reading field* (pp. 1–32). Newark, DE: International Reading Association.

Monaghan, J., Israel, S., & Dahl, M. (2007). May Hill Arbuthnot (1884–1969): A pioneer in the field of children's literature. In S. Israel, & J. Monaghan (Eds.), *Shaping the reading field* (pp. 199–219). Newark, DE: International Reading Association.

Monroe, M. (1932). *Children who cannot read: The analysis of reading disabilities and the use of diagnostic tests in the instruction of retarded readers.* Chicago, IL: Chicago Press.

Monroe, P. (1911/1968). *Cyclopaedia of education*. New York, NY: Macmillan.

Montessori, M. (1936). *The secret of childhood*. Toronto, ON: Longmans, Green and Co.

Mullis, I., Kennedy, A., Martin, M., & Sainsbury, M. (2004). *PIRLS 2006 assessment framework and specifications*, Chestnut Hill, MA: Boston College.

Mullis, I., Martin, M., & Gonzalez, E. (2004). *PIRLS international achievement in the processes of reading comprehension: Results from PIRLS 2001 in 35 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis I., Martin, M., Kennedy, A., & Foy, P. (2007). *IEA's progress in international reading literacy study in primary school in 40 countries*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Assessment of Educational Progress [NAEP]. (1982). *Reading comprehension of American youth: Do they understand what they read? Results from the 1979–80 national assessment of reading and literature* {Report No. NAEP-11-R-02}. Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 217 396)

National Center for Education Statistics [NCES]. (1982). *High school and beyond: Sample design report*. Washington, DC: Author.

National Center for Education Statistics [NCES]. (2002). *Early childhood longitudinal study-Kindergarten class of 1998–99 (ECLS–K):*

Psychometric report for kindergarten through first grade. Washington, DC: Author.

National Center for Education Statistics [NCES]. (2005). *Early childhood longitudinal study-Kindergarten class of 1998–99 (ECLS–K): Psychometric report for kindergarten through first grade*. Washington, DC: Author.

New, W. (Ed.). (2002). Encyclopaedia of literature in Canada (electronic source). Toronto, ON: University of Toronto Press.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Ohio Department of Education. (2007). *Highlights: Ohio achievement tests preliminary results May 2007-test administration* [Electronic version]. Columbus, OH: Author.

Ontario Ministry of Education. (2009). *Realizing the promise of diversity: Ontario's equity and inclusive education strategy*. Toronto, ON: Author. Retrieved from www.edu.gov.on.ca.

Organisation for Economic Co-operation and Development [OECD] PISA. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2003). *Literacy skills for the world of tomorrow—Further results from PISA*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA.

(2004a). *Messages from PISA 2000*. Paris: OECD publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2004b). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2005). *School factors related to quality and equity: Results from PISA 2000.* Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2007). *Science competencies for tomorrow's world: Volume 1 analysis*. Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2009). *Equally prepared for life? How 15-year-old boys and girl perform in school.* Paris: OECD Publications.

Organisation for Economic Co-operation and Development [OECD] PISA. (2010). *PISA 2009 results: What students know and can do: Student performance in reading, mathematics and science (Volume 1)*. Paris: OECD Publications.

Parsley, K., Powell, M., & O'Connor, H. (1964). Investigation of sex differences in achievement of under-, average-, and over-achieving students within five IQ groups in Grades Four through Eight. *The Journal of Educational Research, 57*(5), 268–270.

Parsley, K., Powell, M., O'Connor, H., & Deutsch, M. (1963). Are there really sex differences in achievement? *The Journal of Educational Research, 57*(4), 210–212.

Patterson, S. (1930). *Teaching the child to read: Practical studies in reading method.* Garden City, NY: Doubleday, Doran & Co.

Pauly, F. (1951). Sex differences and legal school entrance age. *Journal of Educational Research*, *45*(1), 1–9.

Pauly, F. (1952). Should boys enter school later than girls? *Journal of the National Education Association*, *41*(1), 29–31.

Percival, W. (1926). *A study of the causes and subjects of school failure,* New York, NY: Columbia University Teachers College.

Petrina, S. (2004). Luella Cole, Sidney Pressey, and educational psychoanalysis, 1921–1931. *History of Education Quarterly*, *44*(4), 524–553.

Petticrew, M., & Roberts, H. (2006). *Systematic review in the social sciences: A practical guide*. Padstow, Cornwall, UK: Blackwell Publishing.

Phi Delta Kappan. (1959). Let's give BOYS a break! *Phi Delta Kappa International*, *40*(7), 281–283.

Phillips, C. (1957). *The development of education in Canada*. Toronto, ON: Gage.

Phillips, L. M., & Norris, S. P. (2002). Literacy policy and the value of literacy for individuals. In Y. Hébert (Ed.), *Citizenship in transformation in Canada* (pp. 209–227). Toronto, ON: University of Toronto Press.

Phillips, L., Norris, S., Osmond, W., & Maynard, A. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology*, *94*(1), 3–13.

Polakow, V. (2007). *Who cares for our children? The child care crisis in the other America*. New York, NY: Teachers College, Columbia University.

Prescott, G. (1955). Sex differences in Metropolitan Readiness Test results. *The Journal of Educational Research*, *48*(8), 605–610.

Pressey, L. W. (1918). Sex differences shown by 2, 544 school children on a group scale of intelligence with special reference to variability. *Journal of Applied Psychology*, *2*, 323–340.

Preston, R. (1962). Reading achievement of German and American children. *School and Society*, 90, 350–353.

Purves, A. (1992). *The IEA study of written composition II: Education and performance in fourteen countries*. Oxford: Pergamon Press.

Quinlan, L. (1996). The effects of school entry age and gender on reading, achievement scores of third grade students (Unpublished masters thesis). Union, NJ: Kean College Rampey, B., Dion, G., & Donahue, P. (2009). *NAEP 2008 trends in academic progress* (NCES 2009–479). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Reed, J., & Meyer, R. (2007). Edmund Burke Huey (1870–1913): A brief life with an enduring legacy. In S. Israel, & J. Monaghan (Eds.), *Shaping the reading field: The impact of early reading pioneers, scientific research,*

and progressive ideas (pp. 159–175). Newark, DE: International Reading Association.

Reid, D., Hresko, W., & Hammill, D. (1981). *The test of early reading ability*. Austin, TX: Pro-Ed.

Reid, D., Hresko, W., & Hammill, D. (1989). *The test of early reading ability*–(2nd ed.). Austin, TX: Pro-Ed.

Reid, D., Hresko, W., & Hammill, D. (2001). *The test of early reading ability*–(3rd ed.). Austin, TX: Pro-Ed.

Robinson, J., & Lubienski, S. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*(2), 268–302.

Rousseau, J. (1762/1974). *Emile* (B. Foxley, Trans.). New York, NY: Dutton.

Roy, H., Roy, E., Sheffield, P., & Bollert, G. (1932). *Highroads to reading*. Toronto, ON: Ryerson Press.

Russell, D. (1941). Reading preferences of younger adolescents in Saskatchewan. *The English Journal*, *30*(2), 131–136.

Sailors, M. (2007). Arthur Irving Gates (1890–1972): Educational psychology and the study of reading. In S. Israel, & J. Monaghan (Eds.), *Shaping the reading field: The impact of early reading pioneers, scientific research, and progressive ideas* (pp. 327–346). Newark, DE: International Reading Association.

Samuels, F. (1943). Sex differences in reading achievement. *The Journal of Educational Research*, *36*(8), 594–603.

Sarroub, L., & Pearson, D. (1998). Two steps forward, three steps back: The stormy history of reading comprehension assessment. *Faculty Publication: Department of Teaching, Learning and Teacher Education,* Paper 39. Retrieved from http://digitalcommons.unl.edu/teachlearnfacpub/39

Schwandt, T., Lincoln, Y., & Guba, E. (2007). Judging interpretations: But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*, (114), 11–25. Retrieved from InterScience http: www.interscience.wiley.com

Siegel, L. (2003). Basic cognitive processes and reading disabilities. In H. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities*. New York, NY: Guilford Press.

Sinks, N., & Powell, M. (1965). Sex and intelligence as factors in achievement in reading in Grades 4 through 8. *The Journal of Genetic Psychology*, *106*(1), 67–79.

Slavin, R. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21.

Slosson, R. (1963). *Slosson Oral Reading Test* (SORT). East Aurora, NY: Slosson Educational Publications.

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

Smith, N. (1950). Readiness for reading. In *Readiness for reading and related language arts: A Digest of current research*, (pp.3-33). Urbana, IL: A Research Bulletin of The National Conference on Research in English, National Council of Teachers of English.

Sokal, L. (2010). Prevalence of gendered views of reading in Thailand and Canada. *Alberta Journal of Educational Research, 56*(1), 44–56.

Sokal, L., & Katz, L. (2008). Effects of technology and male teachers on boys' reading. *The Australian Journal of Education, 52*(1), 80–93.

Sokal, L., Katz, H., Adkins, M., Gladu, A., Jackson-Davis, K., & Kussin, B. (2005). Boys will be "boys": Variability in boys' experiences of literacy. *Alberta Journal of Educational Research, 51*(3), 216–230.

Sokal, L., Katz, H., Adkins, M., Grills, T., Stewart, C., Priddle, G. . . . Chochinov-Harder, L. (2005). Factors affecting inner-city boys' reading: Are male teachers the answer? *Canadian Journal of Urban Research, 14*(1), 107–130.

Sokal, L., Katz, H., Chaszewski, L., & Wojcik, C., (2007). Good-bye, Mr. Chips: Male teacher shortages and boys' reading achievement. *Sex Roles: A Journal of Research*, 56(9/10), 651-659.

Sokal, L., Monette, J., McBey, S., & Wojcik, C. (2006). *Factors affecting boys' reading gains*. Paper presented at the annual meetings of the American Educational Research Association, San Francisco, CA.

Sokal, L., Thiem, C., Crampton, A., & Katz, H. (2009). Differential effects of male and female reading tutors based on boys' gendered views of reading. *Canadian Journal of Education, 32*(2), 245–270.

Spache, G. (1972). *Diagnostic reading scales*. Monterey, CA: McGraw-Hill.

Standish, E. (1960). Group tests of reading readiness. *Educational Research, 2*(2), 155–160.

Stauffer, R. (1971). Thorndike's "reading as reasoning": A perspective. *Research Quarterly, 6*(4), 443–448.

Steiner, E., Steinen, R., & Newman, I. (1981). A current investigation of sex differences in reading achievement and the sex-typing of reading. Paper presented at the Annual Meeting of the *American Educational Research Association*, Los Angeles, CA.

Stone, C. (1922). *Silent and oral reading: A practical handbook of methods based on the most recent scientific investigations*. Boston, MA: Houghton Mifflin.

Stotsky, S., Denny, G., & Tschepikow, N. (2010). Single-sex classes in two Arkansas elementary schools: 2008–2009. *Education Working Paper Archive*. Fayetteville, AR: University of Arkansas. Retrieved from http://isu468research.wikispaces.com/file/view/Single+-+Sex+classes+-+Stotsky.pdf

Stroud, J. B., & Linquist, E. F. (1942). Sex differences in achievement in the elementary and secondary schools. *Journal of Educational Psychology*, 33, 657–667.

Sutherland, N. (1976). *Children in English-Canadian society: Framing the twentieth-century consensus*. Toronto, ON: University of Toronto Press.

The Education Alliance. (2007). *Gender differences in reading achievement: Policy implications and best practices*. Pittsburgh, PA: Claude Worthington Benedum Foundation.

Thompson, H. (1903). The mental traits of sex: An experimental investigation of the normal mind in men and women [Electronic version]. In C. Green (Ed.), *Classics in the History of Psychology* (Chapter 7). Toronto, ON: York University.

Thorndike, E. [Ed.]. (1903). Heredity, correlation and sex differences in school abilities. *Columbia University Contributions to Philosophy, Psychology and Education*, *6*(2), 1–60.

Thorndike, E. (1904). *An introduction to the theory of mental and social measurements.* New York, NY: Science Press.

Thorndike, E. (1914). *Educational psychology* (2nd ed.). New York, NY: Teachers College, Columbia University**.**

Thorndike, E. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *The Journal of Educational Psychology*, *8*(6), 323–332.

Thorndike, R. (1973). *Reading comprehension education in fifteen countries: International studies in education III*. New York, NY: Halsted Press.

Tostberg, R. (1971). Biographical essay on Arthur I. Gates. In R. J. Havighurst (Ed.), *Leaders in American education: The 70th yearbook of the National*

Society for the Study of Education, Part II (pp. 222–230). Chicago, IL: University of Chicago Press.

Traxler, A. E. (1935). Sex differences in rate of reading in the high school. *Journal of Applied Psychology*, 19, 351–352.

Traxler, E., & Townsend, A. (1946). *Another five years of research in reading: Summary and bibliography* [Educational Records Bulletin No. 46]. New York, NY: Educational Records Bureau.

UNESCO Institute for Education. (1962). *Educational achievements of thirteen-year-olds in twelve countries: results of an international research project 1959–61*. Hamburg, DE: Unesco Institute for Education.

Venezky, R. (1996). The development of literacy in the industrialized nations of the West. In R. Barr, M. Kamil, P. Mosenthal, & D. Pearson (Eds.), *Handbook of reading research volume II* (pp. 3–26). Mahwah, NJ: Lawrence Erlbaum Associates.

Vrooman, M. (2009). *An examination of the effects of single-gender classes on reading and mathematics achievement test scores of middle school students* (Unpublished doctoral dissertation). Norman, OK: University of Oklahoma.

Wargacki, J. (2008). *High-stakes testing and the gender gap: A study of fourth grade reading performance on the Ohio achievement test* (Unpublished master's thesis). Bowling Green, OH: Green State University.

West, C., & Zimmerman, D. (1987). Doing gender. *Gender and Society*, *1*(2), 125–151.

West, P. (1999). Boys' underachievement in school: Some persistent problems and some current research. *Issues in Educational Research, 9*(1), 33–54.

Wheat, H. (1923). *The teaching of reading: A textbook of principles and methods*. Boston, MA: Ginn and Company.

Whipple, G. (Ed.). (1925). The development of a meaningful, vocabulary, and of independence in word recognition. In G. Whipple (Ed.), *The twenty-fourth yearbook of the National Society for the Study of Education. Part I. Report of the National Committee on Reading*. Bloomington, IL: Public School Publishing Company.

White, B. (2007). Are girls better readers than boys? Which boys? Which girls? *Canadian Journal of Education, 30*(2), 554–581.

Wilhelm, J. (2009). *Let's get boys reading and writing: An essential guide to raising boys' achievement.* Don Mills, ON: Oxford University Press.

Willms, J. (2003). *Student engagement at school: A sense of belonging and participation: Results from PISA 2000*. Paris, FR: OECD publications.

Willms, J. (2004). *Variation in literacy skills among Canadian provinces: Findings from the OECD PISA.* Report commissioned by Statistics Canada. Catalogue number: 81-595-MIE20040012 { http://www.statcan.ca/cgi-bin/downpub/listpub.cgi?catno=81-595-MIE2004012 }

Woolley, H. (1910). A review of the recent literature on the psychology of sex. *Psychological Bulletin, 7,* 335–342.

Wozencraft, M. (1963). Sex comparisons of certain abilities. *The Journal of Educational Research, 57*(1), 21–27.

Wyatt, N. (1966). Sex differences in reading achievement. *Elementary English*, 43, 596–600.

Yarborough, B., & Johnson, R. (1980). A six-year study of sex differences in intellectual functioning, reading/language arts achievement, and affective development. *The Journal of Psychology*, 106, 55–61.