GENDER AND ABORIGINAL DIFFERENCES IN ELEMENTARY SCHOOL STUDENTS' CBM READING, WRITING, AND DIBELS SCORES

by

Shelley Wiltshire

B.H.E., University of British Columbia, 1984

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

in

EDUCATIONAL COUNSELLING

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

July 2004

Your file Votre référence
ISBN: 0-494-04700-3
Our file Notre référence
ISBN: 0-494-04700-3

ABSTRACT

This study uses Curriculum Based Measurement data of students' reading and writing fluency and Dynamic Indicators of Basic Early Literacy Skills data to investigate the relationship between scores on these achievement measures, the gender of the students, and the aboriginal status of the students. The sample consists of 2272 elementary students randomly selected for the Prince George School District norming project. The measurements were collected by teachers and other school district staff in each elementary school during October, January, and April of the 2002/2003 school year. Scores were analyzed using a 2 X 2 analysis of variance (gender by aboriginal status). Gender, aboriginal status and the dependent variables of reading and written expression scores were analyzed for each of Grade 1 through 7. Gender, aboriginal status and the dependent variables of pre-reading and early reading skills scores were analyzed for Kindergarten and Grade 1. Repeated measures for October, January, and April were compared for trends in reading and written expression fluency and pre-literacy skills over the school year. Although male students' mean scores in reading, writing, and in early literacy skills were lower than female students' mean scores at every grade level and every testing period, the only consistent statistically significant gender effect was found in written expression fluency and only for Grade 2 to 7. A consistent statistically significant aboriginal status effect was found only for reading expression fluency from Grade 1 through 7 and for early literacy skills for Kindergarten and Grade 1. Aboriginal students' mean scores in early literacy skills and in reading and writing fluency were lower than non-aboriginal students' mean scores at every grade level and testing period except the grade five January testing for all variables.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to thank Dr. Peter MacMillan, Associate Professor at UNBC, for his endless support and teaching throughout the thesis process. Without his help this thesis would definitely not have been possible. I would also like to thank Catherine McGregor, Lecturer at UNBC and PhD candidate, for her support and assistance with my defense and the final document.

I would like to thank Dr. Robert Tait, Dean of Graduate Studies, and the other members of my committee, Dr. Carl Anserello, School Services Administrator for School District #57, and Paul Michel, Adjunct Professor in First Nations Studies at UNBC and Coordinator for the First Nations Centre, for their support and encouragement. I would also like to thank the Aboriginal Education Board of School District #57 for their support of this research.

I would like to dedicate this thesis to my husband Ed and my daughters Savannah and Sian. Without their constant patience, love, support, and encouragement I would never have been able to complete this thesis or the masters program.

CHAPTER ONE: INTRODUCTION

The issue of literacy and the factors that influence the success or failure of students is an increasingly examined and discussed topic. The relative importance of gender and levels of achievement is being discussed and debated at both a school and university level. There are a number of academic indicators that point to differences between boys and girls with respect to literacy. When examining the Foundation Skills Assessment results or provincial examination results girls are outscoring boys in numerous areas including literacy. Hedekar's (1997) study using Curriculum Based Measurement (CBM) also found definite gender differences in literacy.

Another issue for educators is the matter of literacy among aboriginal students in British Columbia. There has been a long history of achievement differences between aboriginal and non-aboriginal students. Again indicators such as the Foundation Skills Assessment and provincial examination results highlight the need to examine these differences so that these issues can be addressed.

The Prince George School District in British Columbia has identified a high number of students lacking in early literacy skills, particularly males and aboriginals, and have made improving student literacy, particularly in these two groups, a priority (School District No. 57, Prince George, 2003). Also of concern to the Prince George School District is the fact that the Foundation Skills Assessment results indicate the gender gap favouring females is larger in the Prince George School District than it is at the provincial level. The two test instruments the school district is using to assess these early literacy skills are Curriculum Based Measurement (CBM) and Dynamic Indicators of Basic Early Literacy Skills (DIBELS). In the Prince George School District CBM and DIBELS are

being used to measure the curriculum being taught and growth in student learning against previously established district norms. Based on the Foundation Skills Assessment results, provincial government exam results, Hedekar's (1997) previous results and the fact that the Prince George School District has identified the area of gender and aboriginal differences in literacy as a concern, the importance of my research study is to examine the CBM reading and writing scores and the DIBELS scores for approximately 2200 students in order to analyze the effects of gender and aboriginal status on the acquisition of early literacy skills in the Prince George School District.

Curriculum Based Measurement (CBM) is a series of short, informal achievement tests that are standardized yet based on curriculum being used in the classroom (Scott & Weishaar, 2003). The CBM measures of literacy used in this study include Words Read Correctly (WRC), Words Spelled Correctly (WSC), and Total Words Written (TWW). Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a standardized, individually administered set of tests that measure pre-reading and early reading skills (University of Oregon (a), n.d.). The DIBELS measures used in this study include Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Oral Reading Fluency (ORF).

Description of School District

The Prince George School District (SD #57) has been using CBM as an assessment tool since 1996. School District #57 is located in the central interior of British Columbia and covers an area of almost 52,000 square kilometres. The communities covered by this district include Prince George, Mackenzie, McBride, Valemount, and Hixon as well as many small settlements in between. Because of the

vast area covered by the school district the schools are located in a variety of settings including inner city, suburban, and rural. In 2003 there were approximately 16,400 students in the district, of these approximately 18% are aboriginal. There are 37 elementary schools in the district.

Research Questions

1. Is there a gender difference in reading or writing fluency of elementary school students based on CBM/DIBELS measures? Is this gender difference consistent throughout the grades?
2. Is there a difference in reading or writing fluency for aboriginal elementary school students versus non-aboriginal elementary school students based on CBM/DIBELS measures? Is this effect consistent across all grade levels?
3. Is there an interaction between gender and aboriginal status for elementary school students when examining gender and aboriginal status differences in reading or writing fluency, based on CBM/DIBELS measures?

Hypotheses

The following are a number of statistical hypotheses that were generated by the research questions and tested during this study.

1. Within a given grade level the mean reading fluency (as measured by the variable Words Read Correctly) of male students equals that of female students.

   a) $H_0: \mu(r)_{gm} - \mu(r)_{gf} = 0$

   $H_1: \mu(r)_{gm} - \mu(r)_{gf} \neq 0$

where r refers to reading fluency as measured by Words Read Correctly, g refers to Grades 1 through 7, and m and f refers to male and female respectively.

Writing fluency is measured by two highly correlated variables Words Spelled Correctly (WRC) and Total Words Written (TWW).

b) H_0: $\mu(w)_{gm} - \mu(w)_{gf} = 0$

H_1: $\mu(w)_{gm} - \mu(w)_{gf} \neq 0$

where g, m, and f are defined as above and where w refers first to a test with the variable WSC and then with the variable TWW.

2. To investigate the second research question the means of the reading and writing fluency variables were compared for aboriginal elementary students and non-aboriginal elementary students.

a) H_0: $\mu(r)_{gab} - \mu(r)_{gnab} = 0$

H_1: $\mu(r)_{gab} - \mu(r)_{gnab} \neq 0$

where ab refers to aboriginal and nab refers to non-aboriginal and the other symbols are defined as previously stated.

b) H_0: $\mu(w)_{gab} - \mu(w)_{gnab} = 0$

H_1: $\mu(w)_{gab} - \mu(w)_{gnab} \neq 0$

where w refers first to a test with the variable WSC and then with the variable TWW. Other symbols are defined as previously stated.

3. Finally to investigate if there is any interaction between gender and aboriginal status the means for reading and writing fluency for both gender groups and aboriginal, non-aboriginal groups were compared.

α) H_0: $\mu(r)_{gab \times gen} - \mu(r)_{gab} - \mu(r)_{ggen} + \mu(r)_g = 0$

H_1: $\mu(r)_{gab \times gen} - \mu(r)_{gab} - \mu(r)_{ggen} + \mu(r)_g \neq 0$

where gen refers to gender. Other symbols are defined as previously stated.

b) H_0: $\mu(w)_{gab \times gen} - \mu(w)_{gab} - \mu(w)_{ggen} + \mu(w)_g = 0$

H_1: $\mu(w)_{gab \times gen} - \mu(w)_{gab} - \mu(w)_{ggen} + \mu(w)_g \neq 0$

The symbols are defined as previously stated.

CHAPTER TWO: LITERATURE REVIEW

This chapter consists of four sections in which I will discuss literature relevant to this study. In the first section I will investigate literacy as measured by Curriculum Based Measurement (CBM). In the second section I will discuss literacy as measured by Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The third section will be where I review gender studies relating to reading and writing and in the last section I will examine aboriginal studies and issues relating to reading and writing.

Literacy as Measured by CBM

Curriculum Based Measurement initially developed in the area of special education. It was developed with the intention of testing a special education intervention model that would formatively evaluate teacher instruction in order to improve their effectiveness (Deno, 2003). In the 1980's there was a need to come up with an alternative measurement system to commercial standardized achievement tests and teacher observations. This alternative would provide a data base to evaluate students' overall proficiency in basic skills and to assist teachers in their instructional planning with the end goal of improving student achievement (Fuchs & Fuchs, 1991). CBM has emerged as a set of procedures used by teachers to evaluate student progress and instructional effectiveness (Deno, 1985). Although CBM was initially developed and tested for reliability and validity for testing reading skills, it is also used to reliably and validly test written expression and spelling skills (Deno, 1985).

Reasons for Using CBM

There are a variety of reasons for using CBM as an alternative measurement system the first of which is the validity and reliability of CBM measures. Due to the

standardized nature of CBM, a large number of reliability and validity studies have been conducted (Deno, 1992). Deno (1985) also reported that all CBM measures are highly correlated with performance on the standardized, norm-referenced tests with a particularly close relationship between reading aloud from text and comprehension scores.

A second reason for using CBM as an alternative measure is the improved level of communication of information that can be provided by using CBM. The graphical images that can be produced using data collected by CBM procedures are clear and simple to interpret making it easy for teachers, parents, and students to see individual levels of performance and rates of change or growth in achievement over time. These levels can then be referenced to the student's individual goals, to the instructional program and to peers in the class, the school or the district (Fuchs & Fuchs, 1991).

A third reason for using CBM is the flexibility. Although CBM procedures are standardized, teachers have the freedom to identify the curriculum materials to be used in the testing as well as the level within that curriculum that they want to be mastered by the end of the year (Fuchs & Fuchs, 1991). This allows for individual needs and interests of the teacher or school to be met.

With the current situation in education of cutbacks in funding and increased curricular demands, cost and time effectiveness is the fourth reason to use CBM. The fact that additional testing materials do not need to be purchased to use CBM is a cost saving. With commercial standardized tests is the hidden expense of the procedure to yield a norm-referenced score which will give little information about the individual student's performance in the local curriculum (Deno, 1985). The time saving for

administering CBM is crucial as well. Due to the multiple sampling approach of CBM, performance samples are generally 1 to 3 minutes long whereas the time to administer standardized achievement tests is generally an hour or more (Deno, 2003). Another consideration for time and cost-effectiveness is the amount of time and money required to train teachers or others to administer the CBM test samples. According to Deno (2003) it is easy for professionals, paraprofessionals and parents to learn to use CBM and still obtain reliable data.

The final reason for using CBM is the fact that research has shown that when CBM is used to monitor the effectiveness of an instructional program and formulate improvements the quality of instruction as well as student achievement goes up (Fuchs & Fuchs, 1991). The reasons why a school or district might choose to use CBM is twofold in that it not only provides an assessment tool but it can assist in improving the level of both instruction and student achievement. Despite the benefits of using CBM I could not find any information in the literature to indicate that CBM is being widely used in school districts in British Columbia or Canada.

Limitations of CBM

While CBM may initially be viewed as an answer to achievement measurement concerns, there are some problematic issues that need to be identified. It was previously mentioned there is a strong correlation between CBM measures for reading and reading comprehension scores, Deno (1985) also cautions that reading aloud from text may be detached from comprehension as in the case of "word callers", students who read fluently but do not understand what they read. A study by Hamilton and Shinn (2003) investigates the question of whether or not "word callers" read fluently but lack

comprehension by comparing the oral reading and comprehension skills of teacher-identified "word callers" with that of peers who were identified by the teacher as fluent readers with good comprehension skills. The 66 students involved in the study were all in Grade 3 and were administered four reading tests: the Curriculum-Based Measurement of Reading (R-CBM), the Curriculum-Based Measurement-Maze (CBM-Maze), a comprehensive oral question answering test (CQT), and the Passage Comprehension subtest of the Woodcock Reading Mastery Test (WRMT-PC). The results of the study indicate that students in the "word caller" group not only comprehended significantly ($p < .001$) less well than their peers who read and comprehend well but the "word callers" also had significantly ($p < .001$) lower oral reading fluency scores (Hamilton & Shinn, 2003).

Another finding of Hamilton and Shinn's (2003) study was that teachers over predicted the reading fluency scores of both groups of student which brings into question the accuracy of teachers' judgements regarding students' reading fluency skills. This study seems to indicate teachers' judgements about the reading fluency of whom they identify as "word callers" may not be accurate which gives strength to the argument that CBM measures of reading fluency are valid measures of reading comprehension.

Another problem arises in the area of training. Deno (1985) states teachers must be carefully trained and extremely efficient in using CBM if it is to remain a time-effective approach to measurement of achievement. In another paper Deno (2003) also states time as being the most important barrier to teachers in implementing the measurement procedures.

Lastly, the question of the most effective use of CBM needs to be addressed. As

far as formative evaluation of individual students is concerned CBM is most effective in settings where special education teachers have the time and skills to chart the progress of individual students and then adjust the student's program in response to the data these charts provide (Deno, 2003). With the inclusion of students with disabilities into regular classrooms and increases in class sizes it is unlikely that CBM will be as effective at improving student achievement in these settings as compared to more individualized settings. However, CBM can still be used as an effective assessment tool to measure the progress of students and the curriculum being taught in the classroom.

Reliability and Validity of CBM as a Measure of Literacy

There are many aspects and modes of literacy, however for the scope of this study literacy will be defined as reading and writing fluency. The reliability and validity of CBM as a measure of literacy, specifically reading, writing, and spelling has been widely researched and will also be addressed here. The criterion validity of performance on some of the CBM tasks, specifically cloze procedures (supplying words deleted from text), word meanings and reading aloud, are examined by Deno (1985) with respect to commercial standardized norm-referenced tests. The results indicate that all CBM measures except for word meanings are highly correlated (.70 to .95) with standardized norm-referenced tests such as the Literal and Inferential subtests of the Stanford Achievement Test and the Woodcock Reading Mastery Test (Deno, 1985). Similar results of validity were found for the written expression and spelling measures of CBM.

In a review by Good and Jefferson (1998) criterion-related validity coefficients were examined for the CBM measures of oral reading fluency passages and correct writing sequences with story starters. The tests with which the CBM measures were

validated were published, norm-referenced or criterion-referenced tests, or tests from a published basal reader series. The results indicate that the median validity coefficients for the CBM reading measure for Grades 2 to 6 range from .62 to .73, which is within the acceptable range of concurrent, criterion-related validity coefficients of .60 to .80 (Good & Jefferson, 1998). The results for the CBM writing measure are not quite as impressive with the median validity coefficients for Grades 2 to 11 ranging from .48 to .68 (Good & Jefferson, 1998). This provides less support for the construct validity of the CBM writing measure.

In another study in 1992 Shinn, Good, Knutson, Tilly, and Collins (as cited in Good & Jefferson, 1998) used multiple reading measures to test the construct validity of these measures with respect to reading comprehension. This study involves Grade 3 students and Grade 5 students. For the Grade 3 students the construct examined is reading competence and the CBM reading probes tested indicate construct validity coefficients of .88 to .90 (Good & Jefferson, 1998). For the Grade 5 students the constructs examined are decoding and comprehension and the CBM reading probes tested indicate construct validity coefficients of .74 to .90 (Good & Jefferson, 1998).

In a study conducted in School District 57 (Prince George), Fewster and MacMillan (2002) found that school-based information, such as teacher-awarded grades, adds to the validity of CBM. Their study examined the validity of elementary school CBM scores to predict grades in future courses that are reading and writing intensive and to predict program placements. Their results indicate CBM measures of words read correctly and words spelled correctly are significant predictors of future grades particularly for words read correctly and at the Grade 8 level. The same validity is not

indicated for WSC as a measure of overall writing competency.

A study by Gansle, Noell, VanDerHeyden, Naquin, and Slider (2003) looks at the need for a variety of other new writing measures beyond TWW or correct word sequences as an indication of students' written skill levels. Their study includes third and fourth grade students from one school who completed two 3-minute writing probes on two consecutive days (Gansle et al., 2003). Students were also ranked in terms of their writing skills by their classroom teacher plus standarized criterion test scores were analyzed, specifically the Iowa Test of Basic Skills (ITBS) for the third grade students and the Louisiana Educational Assessment Program (LEAP) for the fourth grade students. The CBM measure of TWW is one of a number of predictor variables including parts of speech, long words, words spelled correctly, total punctuation marks, correct punctuation marks, correct capitalization, complete sentences, words in complete sentences, words in correct sequence, sentence fragments, simple sentences, computer-scored variables. These predictor variables are measured to determine the best predictor of the three criterion variable scores. The largest correlations in this study between predictor variables and the criterion variable of ITBS are for the variables of correct punctuation marks and words in correct sequence which had correlation coefficients ranging from .35 to .44 (Gansle et al., 2003). For correlations between the predictor variables and the criterion variable of LEAP the results are highest for number of verbs, .33, and the computer-scored variable of vocabulary complexity, .24. The largest correlations between the predictor variables and the criterion variable of classroom teacher rankings are for the variables of words in correct sequence, .37, and correct punctuation marks, .35. These results indicate that TWW is not the best predictor of

written skills as measured by the criterion variables of teacher rankings, the Iowa Test of Basic Skills, and the Louisiana Educational Assessment Program.

The reliability of CBM measures is inherent in the very nature of the frequent collection of data to assess growth in the skills being measured. Traditional achievement tests that are norm-referenced or grade-equivalent scored do not reliably reveal an individual student's growth in reading proficiency (Deno, 1985). With CBM assessment it is possible to repeat data collection frequently with the same sample of students and with a larger number of students than would be possible with other more traditional assessment tools. In addition it was found that reading aloud from text was reliable in discriminating which students were in special education programs and which ones were not (Deno, 1985). Simple data such as words read correctly can reliably be used to monitor growth in reading. The study by Fewster and MacMillan (2002) also shows CBM reliably predicts program placements especially for honours programs.

In a previous study done by Hedekar (1997) in the Prince George school district reliability and validity coefficients were reported for the CBM measures of WRC, WSC, and TWW. For Hedekar's (1997) study the Pearson correlation coefficients indicate a high correlation between WSC and TWW, $.91 < r < .99$, and a low to medium correlation between WRC and TWW, $.31 < r < .48$. The reliability across the 6 month testing period for Hedekar's (1997) study also shows stability with coefficients for WRC ranging from .77 to .86 and for TWW coefficients ranging from .48 to .62. The inter-rater reliability for Hedekar's (1997) study was also examined and was found to be very reliable with correlations of .97 to .99 between the scores given by different raters on the same tests.

In the norming project for the current study in the Prince George school district

the Pearson correlation coefficients also indicate a high correlation between WSC and TWW, $.94 < r < .99$, and a low to medium correlation between WRC and TWW, $.27 < r < .49$ (Fewster, Fortier, Foulds, MacMillan, Struthers, & Walraven, 2003). The reliability across the 6 month testing period for the norming project also shows stability with coefficients for WRC ranging from .81 to .86 and for TWW coefficients ranging from .58 to .65 (Fewster et. al., 2003).

Literacy as Measured by DIBELS

Measuring literacy at the Kindergarten and Grade 1 level is a difficult task. The challenge is to find measures that will assess students' literacy through reading and writing skills when students have not yet acquired these skills. DIBELS is a logical measurement system due to the fact that it tests early literacy skills in the grades where pre-reading, pre-writing, early reading and early writing skills are initially taught. Testing at the Kindergarten and Grade 1 level using DIBELS measures "provides a reliable and valid indicator of children's progress toward the acquisition of early literacy skills" (Elliot, Lee, & Tollefson, 2001, p. 35).

What is DIBELS?

The DIBELS assessments are a standardized set of short, individually administered measures that assess three of the essential early literacy domains: phonological awareness; alphabetic principle; and fluency with connected text (University of Oregon (a), n.d.). The original DIBELS measures are a set of 10 that were initially designed as downward extensions of the CBM reading probes (Elliot et al., 2001). The DIBELS measures used in my research include: Letter Naming Fluency (LNF), an indicator of risk for difficulty in achieving early literacy benchmark goals;

Initial Sound Fluency (ISF) and Phonemic Segmentation Fluency (PSF), used to assess phonological awareness; Nonsense Word Fluency (NWF), used to assess alphabetic principle; and Oral Reading Fluency (ORF), used to assess fluency with connected text (University of Oregon (b), n.d.). The measures are intended to be used together in order to be empirically valid and reliable, as supported by Good, Kaminski, Smith, Simmons, Kameenui, and Wallin (2003) who outlined that at the Kindergarten level instructions on phonemic awareness, especially blending and segmentation, needs to be explicitly integrated with sounds of letters to ensure reading development later on.

Uses of DIBELS

DIBELS is a standardized assessment system to test pre-cursor skills for early literacy. The DIBELS assessment is administered three times a year and can be used with students from Kindergarten through to Grade 3. It provides a series of benchmarks for each measure at each grade level.

The resulting data that are produced from DIBELS measures has innumerable uses. The data can be used to assess the quality of instruction and supplemental programs, school outcomes, professional development, curriculum and supplemental materials adequacy and appropriateness, and additional intervention which are all elements of an effective beginning reading program (Good et al., 2003). Another positive of using DIBELS as an assessment tool is the benefits that children may gain from being exposed to these skills (Elliot et al., 2001).

Reliability and Validity of DIBELS as a Measure of Literacy

The question of whether or not DIBELS measures emerging literacy skills needs to be addressed. The study by Elliot et al. (2001) addresses this question by correlating

average DIBELS scores and a variety of achievement-related criterion measures such as the Woodcock-Johnson Psycho-Educational Achievement Battery-Revised (WJ-R) Broad Reading and Skills clusters, the Test of Phonological Awareness (TOPA), the Teacher Rating Questionnaire (TQR), the Developing Skills Checklist (DSC), and the Kaufman Brief Intelligence Test (K-BIT). The results generally support previous research on DIBELS that pre-literacy abilities in Kindergarten are associated with later reading fluency (Elliot et al., 2001). The use of DIBELS over the previously mentioned standardized tests is preferred because as well as proving technical adequacy, the DIBELS measures are more practical because they are more easily administered and repeated, more easily adapted to curriculum, more easily scored, and can be used with minimal training and materials (Elliot et al., 2001).

Having demonstrated the effectiveness of DIBELS as an appropriate measure of pre-literacy skills, also of significant importance to this study is the compatibility of DIBELS with CBM measures. The DIBELS measures were originally developed as extensions to the CBM measures and so a discussion of the correlation between the two systems of assessment is essential (Elliot et al., 2001). The test for LNF asks students to name as many letters as they can in one minute from a random presentation of upper- and lower-case letters. The LNF measure is a standardized measure of risk used to assess risk of not achieving early literacy benchmark goals in Kindergarten and has a predictive validity of .71 with the Grade 1 CBM Oral Reading Fluency (ORF) measure (Good, Wallin, Simmons, Kame'enui, and Kaminski (2002). The ISF measure tests students' ability to identify and produce the beginning sound of an orally and pictorially presented word. The predictive validity of ISF with the CBM Oral Reading Fluency (ORF)

measure taken in the spring of Grade 1 is .45 (Good et al., 2002). The PSF measure is used from the winter of Kindergarten through to the middle of Grade 1 and assesses the students' ability to fluently segment three- and four-phoneme words into their individual phonemes. The PSF assessed in the spring of Kindergarten has a predictive validity of .62 with the spring of Grade 1 CBM ORF (Good et al., 2002). The NWF measure uses a list of nonsense words that the student has to either read or reproduce the letter sounds of each word in one minute. The predictive validity of NWF in January of Grade 1 with the CBM ORF in May of Grade 1 is .82 and with the CBM ORF in May of Grade 2 is .66 (Good et al., 2002). The DIBELS Oral Reading Fluency (DORF) assessment is a set of passages used to assess oral reading fluency from Grade 1 through to Grade 3 and has a median concurrent validity of .95 with the Test of Reading Fluency (TORF) which is a version of the CBM ORF (Good et al., 2002).

The norming project for the current study in the Prince George school district examined correlations between the four variables (ISF, LNF, PSF, and NWF) tested at the Kindergarten level (Fewster et. al., 2003). The Pearson correlation coefficients for ISF with the other three variables range from .426 to .593, the Pearson correlation coefficients for LNF with the other three variables range from .342 to .708, the Pearson correlation coefficients for PSF with the other three variables range from .342 to .593, and the Pearson correlation coefficients for NWF with the other three variables range from .446 to .708 (Fewster et. al., 2003). The reliability across the 3 month testing period for the norming project also shows stability with coefficients for three of the Kindergarten test variables being .687 for PSF, .695 for ISF, and .741 for NWF. The reliability across the 6 month testing period for the norming project shows stability with

the coefficient for the fourth Kindergarten test variable LNF being .649.

The norming project also examined correlations between the seven variables (PSF, NWF, LNF, ORF, WRC, TWW, and WSC) at the Grade 1 level. The Pearson correlation coefficients for PSF with the other six variables range from .231 to .559, for NWF with the other six variables the coefficients range from .428 to .821, for LNF the coefficients range from .422 to .738, for ORF the coefficients range from .231 to .925, for WRC the coefficients range from .309 to .925, for TWW the coefficients range from .413 to .944, and for WSC the coefficients range from .385 to .944 (Fewster et. al., 2003). The Grade 1 results for the norming project also indicate reliability across the testing period for the variables that were tested more than once, specifically PSF, NWF, and ORF. The Pearson correlation coefficients for these three variables over the testing period are .706 for PSF, .645 for NWF, and .903 for ORF (Fewster et. al., 2003).

In summary the measures used in the Prince George norming project are good, reliable measures of early literacy skills in the Prince George school district. These measures comprise the data being analyzed in this study.

Modification of DIBELS Measures

One study by Elliot et al. (2001) looked at modifying the DIBELS measures and investigating their technical adequacy for identifying Kindergarten children at risk for reading failure. The measures that are modified in their study are PSF and ISF. The measures for PSF and ISF are changed to Phoneme Segmentation Ability (PSA) and Initial Sound Ability (ISA), respectively, to differentiate these modified measures from the original DIBELS measures because the modified measures stress the measurement of accuracy instead of the measurement of fluency. The experimental measure of Sound

Naming Fluency is also included in the study done by Elliot et al. (2001) because letter-sound connections have been instructed and measured extensively with children in Kindergarten. The results of the study indicate that initial support of SNF is positive but additional work is needed on instrumentation, improved training and administration of the PSA and ISA measures (Elliot et al., 2001).

Gender Studies Relating to Literacy

The current state of a gender gap in literacy, with respect to reading and writing skills is without question. Numerous examples of females outperforming males can be found in assessment results across Canada. In a survey of gender differences by Gambell and Hunter (2000), provincial exam results for Quebec, British Columbia and Saskatchewan indicated females outperform males in all literacy based courses such as English, French, Communications and Literature. While this data is interesting, for the purposes of this study a closer examination of assessments for younger students is more appropriate.

Foundation Skills Assessment Results

The Foundation Skills Assessment (FSA) is a vital source of information regarding basic literacy skills at the Grade 4, 7 and 10 levels. Every year in British Columbia over 140,000 students participate in the FSA that assesses reading comprehension, writing and numeracy in order to provide external information about performance levels in these basic skill areas and to evaluate how well these basic skills are being taught (British Columbia Ministry of Education, 2001). The British Columbia Ministry of Education cautions that the FSA results are just a snapshot of students' basic academic skills in relation to provincial standards and should be considered in

conjunction with numerous other forms of information collected by schools and districts (British Columbia Ministry of Education, 2001).

For the purpose of this study only reading and writing assessment at the Grade 4 and 7 levels will be discussed. The reading comprehension portion of the FSA assessment consists of multiple-choice and written-response questions and the writing component consists of one longer, extended writing task and one shorter, focused writing task (British Columbia Ministry of Education, 2001). These questions are developed from the prescribed provincial learning outcomes that outline expectations of what students in British Columbia should know and be able to do. The results are reported by stating that students are at one of three levels: "exceeds expectations" which means the student has fully met or is beyond the expectations of the grade level on this test; "meets expectations" which means the student meets the widely held expectations of the grade level on this test; and "not yet within expectations" which means the student does not yet have the skills to meet expectations of the grade level on this test (British Columbia Ministry of Education, 2001). The measurement for these results is in percentages of student who fall into the various expectation categories. Statistical measurement has been used so results from year to year can be compared and comparisons of results between district and provincial levels can also be made. For discussion purposes in this study, percentages of students who "exceeds expectations" and "meets expectations" will be combined.

The FSA results at the provincial level over the last 4 years for Grade 4 reading comprehension indicate the gender gap in favour of females has remained steady at about a 6 % difference for percentage of students meeting or exceeding expectations (British

Columbia Ministry of Education, 2003a). The results are presented in Table 1. The provincial Grade 4 results for writing also indicate a gender gap in favour of females and it has ranged from a 5% to 8% difference for percentage of students meeting or exceeding expectations. When examining the same provincial results for students at the Grade 7 level gaps similar to Grade 4 are present, ranging from a 5% to 7% difference in favour of females for reading comprehension. The Grade 7 results for writing indicate the gap is over twice as large as that at Grade 4 with a 13% to 18% difference in favour of females for percentage of students meeting or exceeding expectations.

Table 1

Provincial FSA Trends of Percentage of Students Meeting or Exceeding Grade Expectations by Gender

Grade Level & Year	Reading Female	Reading Male	Writing Female	Writing Male
Grade 4				
2000	83	77	95	88
2001	81	75	95	87
2002	83	77	96	91
2003	80	75	97	91
Grade 7				
2000	84	78	88	74
2001	78	73	90	72
2002	79	74	91	78
2003	80	73	87	72

Another aspect worth discussing is the trend in the FSA data over the last 4 years for both female and male results at the provincial level (see Table 1). The reading comprehension results for Grade 4 females in the province over the last 4 years has shown a slight downward trend going from 83% meeting or exceeding expectations in 2000 to 80% in 2003. A similar trend is evident for Grade 4 males in reading

comprehension with a slight downward trend going from 77% meeting or exceeding expectations in 2000 to 75% in 2003. The trend in the provincial Grade 4 writing results is slightly upward for both males and females, going from 95% in 2000 to 97% in 2003 for females, and from 88% in 2000 to 91% in 2003 for males. The Grade 4 female and male trends follow the overall provincial trends. The trends for the provincial Grade 7 reading comprehension results also show a slight downward trend for both genders, going from 84% in 2000 to 80% in 2003 for females, and from 78% in 2000 to 73% in 2003 for males. The provincial Grade 7 writing results for both females and males indicate a small peak between 3% to 4% in 2002 from 2000, but then decline again in 2003. The Grade 7 female and male trends follow the overall provincial trends.

For this study it is also relevant to review the FSA results for the Prince George school district which are recorded only for 2001 to 2003. The Prince George results over the past 3 years for Grade 4 reading comprehension indicate a gender gap favouring females by a 7% to 8% difference for percentage of students meeting or exceeding expectations, until 2003 where the gap is a 1% difference in favour of males (British Columbia Ministry of Education, 2003b). Refer to Table 2 for percentages. This gender gap in reading for Prince George is 2% to 3% higher than the provincial gap and the males catching up and passing the females in the 2003 results for Prince George does not reflect the provincial results. The Prince George Grade 4 FSA results for 2003 match up with the Grade 4 participants in the current study of gender differences in Prince George.

The Prince George Grade 4 writing results indicate a gender gap favouring females by a 7% to 11% difference over the past 3 years for percentage of students meeting or exceeding expectations. Again the gender gap in Prince George is 2% to 3%

higher than the provincial gap for Grade 4 writing over the past 3 years. The Grade 7 reading comprehension results for Prince George indicate a gender gap favouring females by a difference of 6% to 13% over the past 3 years for percentage of students meeting or exceeding expectations. The gap in Prince George at the Grade 7 level for reading is also larger than the provincial gap by 1% to 6%. The Grade 7 writing results for Prince George indicate an even larger gender gap than for reading, with a difference favouring females by 21% to 25% over the past 3 years for percentage of students meeting or exceeding expectations. Again the Prince George gap is higher, by 7% to 8%, than the provincial gap for writing at the Grade 7 level. The Grade 7 Prince George FSA results for 2003 match up with the Grade 7 participants in the current gender difference study for Prince George.

With the exception of the grade 4 reading results in 2003 the Foundation Skills Assessment results in Prince George indicate a larger gender gap favouring females in literacy than for the overall provincial results. It is important to investigate these results because the Prince George School District has identified gender differences in literacy as an area of concern and made it a priority to address these gender differences in literacy in their district.

Table 2

Prince George FSA Trends of Percentage of Students Meeting or Exceeding Grade Expectations by Gender

Grade Level & Year	Reading Female	Reading Male	Writing Female	Writing Male
Grade 4				
2000	n/a	n/a	n/a	n/a
2001	77	70	93	82
2002	77	69	97	90
2003	72	73	94	86
Grade 7				
2000	n/a	n/a	n/a	n/a
2001	72	66	83	62
2002	78	73	88	74
2003	76	63	82	57

It is particularly worthwhile to examine the Prince George results for any trends over the past 3 years due to the fact that the district has previously identified a concern for improving literacy levels especially among male students (see Table 2). The Grade 4 Prince George results for reading comprehension indicate two different trends for females and males over the past 3 years. The percentage of females meeting or exceeding expectations for reading has declined from 77% in 2001 to 72% in 2003, while the percentage of males meeting or exceeding expectations for reading has risen from 70% in 2001 to 73% in 2003. The trend in reading for Grade 4 females in Prince George follows the provincial trend but the trend in reading for Grade 4 males in Prince George is opposite to the provincial trend. This trend in reading in the Prince George School District could indicate that the district is beginning to address the gender gap in literacy levels for males.

The Grade 4 writing results indicate an altogether different trend from the reading

results. For both females and males the writing results peak in 2002, 97% of females and 90% of males meeting or exceeding expectations, and then the results decline in 2003 to 94% of females and 86% of males meeting or exceeding expectations. This Prince George Grade 4 trend for writing only partially follows the provincial trend, which does not experience a decline in 2003.

The Grade 7 results in Prince George for reading comprehension for females and males also indicate two different trends. The percentage of females meeting or exceeding expectations has risen over the past 3 years from 72% in 2001 to 76% in 2003 with a peak of 78% in 2002. This mirrors the provincial trend for Grade 7 females in reading comprehension. The results in reading for males during this time period also experienced a peak in 2002, of 73%, but overall from 2001 to 2003 the trend has indicated a decline from 66% to 63% of students meeting or exceeding expectations. The provincial trend for Grade 7 males in reading comprehension remained stable during this time frame. The Prince George Grade 7 results for writing for both genders indicate a rise from 2001 to 2002, 83% up to 88% for females and 62% up to 74% for males, and then a decline in 2003, down to 82% for females and 57% males. The Grade 7 writing results for Prince George follow a similar trend in the provincial results but to a larger extent. The Prince George district results that peak in 2002, Grade 4 writing results for both genders and all the Grade 7 results for both genders, indicate an anomaly.

Studies of Specific Gender Differences

Literacy is comprised of many component skills so to say there are gender gaps in literacy is a very broad statement that needs to be more distinctly defined. The volume of studies and literature regarding gender differences in literacy will help with this task.

In Gambell and Hunter's (2000) survey of gender differences in Canada, a cross-Canada assessment of approximately 36000 students aged 13 and 16 years was completed as part of the School Achievement Indicators Programme (SAIP) Reading and Writing Assessment in English. One half of the sample completed a reading assessment with a follow up questionnaire detailing characteristics regarding demographics, education, curriculum, home, self-evaluation, and reading practices. The other half of the sample completed a writing assessment followed up by a questionnaire regarding characteristics about the students, curriculum, home, self-evaluation, and writing practices. Several gender gaps became evident in reading and writing preferences, practices and attitudes (Gambell & Hunter, 2000). Some of these items of difference include a greater percentage of females who: spend time reading for enjoyment; use reading strategies; rate themselves as confident readers; report liking to write; edit their writing; write down ideas as they think about the assignment; and use the dictionary when writing. Where gender gaps favour males there are: patterns of greater amounts of time spent on watching television; and using the computer to complete assignments. Another gender gap is evident in the genre preferred by readers. Females have much broader, more eclectic tastes in reading and were more aware of social issues than males. Some of these preferences, practices, and attitudes were found to predict reading and writing performances. Specifically, enjoyment of reading, self-confidence with respect to reading, and use of context as a reading strategy predicted 20% to 29% of the variation in reading test scores. Gambell and Hunter (2000) found that the results from the writing questionnaire did not have as much predictive power, only 10% to 20% of the variation in writing test scores could be predicted by editing practices, grammar handbook use, and

self-confidence as a writer. Gambell and Hunter's (2000) study also lends some credence to the gender gap with respect to identification with genre and character-personification which could lead to assessment design bias on tests such as the SAIP. More research is needed to understand how the gender differences come about.

In a study by Pomplun, Sundbye, and Kelley (1999) the Kansas Reading Assessment was used as a vehicle to examine the gender gap in performances on differing item formats, specifically constructed-response items. A total of 400 exam booklets were processed for female and male students at the Grade 7 and 10 levels. For the study done by Pomplun et al. (1999) students who had taken the regular assessment, a narrative passage accompanied by 8 to 12 objective items, were then asked to take the parallel assessment which consisted of an expository passage accompanied by eight constructed-response questions. The variables measured had the following rater reliabilities: .66 for handwriting, .76 for mechanics errors, .91 for number of correct answers, .97 for total number of words written, .99 for number of T-units written (a main clause plus any dependent structure), .86 for total number of reproductions, .51 for total number of transformations, and .89 for total number of unrelated clauses produced by the student (p. 59). The results indicate that gender differences favouring females were found in number of correct answers, reproductions, mechanics errors, handwriting, number of words written, T-unit length, and unrelated clauses which may explain why females perform better than males on constructed-response items.

Another area of literacy to be examined for gender differences is the area of spelling ability. In a study by Allred (1990) 3000 students from Grade 1 through 6 (approximately 250 of each gender at each grade level) were tested using the

Comprehensive Tests of Basic Skills (CTBS) to assess proof-reading skills and a written spelling test (WST) using the same words from the CTBS. Data were collected in two ways, a count of females' and males' performances for each word on each test and analyses of variance on the average differences across both tests by gender for each grade (Allred, 1990). The results indicate females in Grade 1 through 6 significantly outscored males on both the CTBS and the WST with all p values $< .001$. Gender differences in spelling relate to gender differences in reading achievement and Allred (1990) suggests that cultural expectations, specifically cross-cultural expectations placed on girls and boys with respect to sex-roles, play a large role in gender differences in reading but it is not the only cause.

In a prior study done by Hedekar (1997) in the Prince George school district a gender difference favouring females was found in all the analyses for WSC and TWW for grades one through seven. A gender difference favouring females was also found in 14 of the 19 analyses for WRC for Grades 1 through 7 in the same study. The effect sizes, Cohen's d, for all analyses in Hedekar's (1997) study range from .15 to .78.

Aboriginal Studies Relating to Literacy

The term aboriginal was chosen to be used in this study because it is the term used by the British Columbia Ministry of Education and it refers to anyone of aboriginal ancestry which includes Status Indian, Non-Status Indians, Inuit, and Metis (British Columbia Ministry of Education, 2002). In British Columbia, students in the education system identify themselves as aboriginal on a voluntary, self-identifying basis in the September of each year (British Columbia Ministry of Education, 2002).

The education system in British Columbia, and for that matter Canada, in both the

public and private sectors has a long and tragic history of failure with aboriginal peoples. This general failure continues today when graduation rates of aboriginal students in British Columbia are considered. Even though graduation rates have been increasing, only 46% of aboriginal students completed high school in 2003 as compared to 79% for the entire province (British Columbia Ministry of Education, 2004, p. 1). Following the progress of a group of Grade 8 cohorts, who started in the system in 1995, at Grade 9 about 5% of the aboriginal students, as compared to about 1% of non-aboriginal students, had left the system. Between Grade 11 and 12 the percentage of aboriginal students lost increases to about 30% as compared to about 6% for non-aboriginal students (British Columbia Ministry of Education, 2002). At the end of the cohort period in 2000 of those aboriginal students remaining only a little over 40% received their Dogwood graduation certificates as compared to a little over 70% for non-aboriginal students. This document shows that not only is there a large gap in graduation rates between aboriginal and non-aboriginal students, but there is also a large gap in drop out rates at a fairly early age. This is another indication of the failure of the education system with respect to aboriginal students.

Foundation Skills Assessment Results

In addition to the gap in graduation and drop out rates there is vast documentation of the gap in achievement between aboriginal and non-aboriginal students. Some areas of achievement that have been documented in British Columbia are in the area of literacy and numeracy under the auspices of the Foundation Skills Assessment (FSA) that is administered to grades 4, 7, and 10 students each year. Due to the scope of this study being Kindergarten to Grade 7 students, only literacy results for grade 4 and 7 students

will be discussed. The two aspects of literacy that are measured by the FSA are reading comprehension and writing. As previously mentioned the British Columbia Ministry of Education cautions that the FSA results are just a snapshot of students' basic academic skills in relation to provincial standards and should be considered in conjunction with numerous other forms of information collected by schools and districts (British Columbia Ministry of Education, 2001).

The FSA results at the provincial level over the last 4 years for Grade 4 reading comprehension indicate that the proportion of aboriginal students meeting or exceeding expectations is 21% to 24% less than for the province as a whole (British Columbia Ministry of Education, 2003a). The results are presented in Table 3. The gap between aboriginal and provincial FSA results for writing at the Grade 4 level over the last 4 years is smaller with differences ranging from 9% to 13% (British Columbia Ministry of Education, 2003a). When examining the same provincial results for students at the Grade 7 level similar gaps are present, ranging from 23% to 25% for reading comprehension, and 18% to 21% for writing.

Table 3

Provincial FSA Trends of Percentages of Students Meeting or Exceeding Grade Expectations by Aboriginal Status

Grade Level & Year	Reading Aboriginal	Reading All	Writing Aboriginal	Writing All
Grade 4				
2000	56	79	78	91
2001	55	78	77	91
2002	56	80	84	94
2003	56	77	85	94
Grade 7				
2000	57	81	60	81
2001	51	76	61	81
2002	52	76	66	84
2003	53	77	61	79

Another issue worth mentioning is the trend in the FSA data over the last 4 years for both the provincial and aboriginal results (see Table 3). The Grade 4 reading comprehension data for the province indicates an insignificant increase in 2002 but then decreases again in 2003, while the aboriginal results replicate the increase in 2002 but remain steady for 2003. The Grade 4 writing data for the province indicates a slight increase for 2002, but the aboriginal results for this measure indicate a larger increase of 7% in 2002, over twice the size of the increase for the province as a whole. The trends for the Grade 7 measures for reading comprehension for both the provincial and aboriginal results indicate a similarly significant decrease in 2001 and then both begin to increase slightly in 2003. The trend for the Grade 7 writing measures for both the provincial and aboriginal results indicate an increase in 2002 and then both decrease by 5% in 2003. Overall, when comparing the 2000 to 2003 results of the reading comprehension and writing measures for both grades, the trends for the aboriginal and the

provincial data are very similar with the exception of the aboriginal writing result in 2002 which had an increase two times that of the provincial increase.

For this study it is relevant to review the FSA results for the Prince George school district as well (see Table 4). The Prince George results over the last 3 years for the Grade 4 reading comprehension measure indicate that the proportion of aboriginal students meeting or exceeding expectations is 14% to 17% less than for the district as a whole (British Columbia Ministry of Education, 2003b). The gap between aboriginal and district FSA results for writing at the Grade 4 level for the last 3 years is slightly smaller than for reading with the exception in 2002 where the gap is only a 5% difference. The gap, between aboriginal and district results for Grade 7 students, ranges from 17% to 21% for the reading measures for the 3 year period, but for the writing measure the gap ranges from 4% to 20%.

Table 4

Prince George FSA Trends of Percentages of Students Meeting or Exceeding Grade Expectations by Aboriginal Status

Grade Level & Year	Reading Aboriginal	Reading All	Writing Aboriginal	Writing All
Grade 4				
2000	n/a	76	n/a	n/a
2001	60	74	77	88
2002	57	73	89	94
2003	55	72	78	90
Grade 7				
2000	n/a	79	n/a	n/a
2001	52	69	53	73
2002	54	75	63	81
2003	52	69	65	69

When examining the Prince George district data from 2001 to 2003 some trends

are indicated (see Table 4). For the Grade 4 reading comprehension measures there is a slight downward trend for both the aboriginal and district results from 2001 to 2003. For the writing measure at Grade 4 there is a bit of an anomaly in 2002 for aboriginal results which increase significantly in that year alone. The district results for the Grade 4 writing measure also increase but not as significantly. At the Grade 7 level for the reading comprehension measure the aboriginal and district results have similar trends of a slight increase in 2002 and then in 2003 the results return to the 2001 level. For the Grade 7 writing measure the aboriginal results increase significantly in 2002 and continue with a slight increase for the next year. The Grade 7 writing measure results for the district show a similar significant increase in 2002 but then the next year drop back to the 2001 level. In summary the Prince George district FSA results for Grade 4 appear to have a slight downward trend in reading comprehension and a bit of an anomaly in 2002 for writing. The Grade 7 results have a somewhat level trend for reading and like the Grade 4 results indicate an anomaly for writing in 2002.

To complete the review of FSA results for reading and writing it is necessary to compare the aboriginal gap at the district level to the aboriginal gap at the provincial level. The aboriginal gap at Grade 4 for reading comprehension is 7% less at the district level than that for the provincial level. The aboriginal gap at Grade 4 for writing is similar at both the district and provincial levels. For Grade 7 the aboriginal gap for reading comprehension is again smaller at the district level, by about 4% to 5% in this case. The Grade 7 aboriginal gap for writing is again similar at both district and provincial levels.

When comparing provincial trends to district trends for aboriginal FSA results,

from 2000 to 2003, there are no similarities. At the Grade 4 level for reading the provincial trend is stable whereas the district trend shows an overall decline of about 5%. For writing at the Grade 4 level the provincial trend indicates an overall increase of 7%, the district trend indicates an anomaly in 2002 where the results increased by 12% and then dropped again by 11% in 2003. The trend in Grade 7 reading results for aboriginal students at the provincial level indicates a decline from 2000 to 2003 whereas the district results remain stable. The Grade 7 writing results provincially for aboriginal students indicates a small peak in 2002 whereas the district results indicate a steady rise over the same time period.

Studies of Specific Aboriginal Differences

In reviewing other literature regarding aboriginality and literacy the differences between aboriginal and non-aboriginal are not always quantifiable performance scores. There are many different types, modes and uses of literacy. Curwen Doige (2001) points out that aboriginal literacy has been neither respected nor explicated throughout our history nor has it been accepted as part of the definition of being aboriginal. She goes on to say that reading and writing are the most narrow definition of literacy and that the language and symbols of aboriginal literacy communicate history, culture, knowledge, tradition, and systems of education and understanding: in other words literacy is vitally connected to who we are. Gaikezehongai (2003) also addresses the important contributions to aboriginal literacy made by aboriginal prophecies, history and traditional teachings being passed down. A similar point is made by Dunn (2001) with respect to the Australian aboriginal people when she talks about implementing a culturally responsive pedagogy that includes things such as knowledge of Australian aboriginal

social history, culturally appropriate literacy education, recognizing and addressing group and individual learning preferences, and accepting a child's primary discourse as legitimate.

Differences between aboriginal and non-aboriginal students with respect to attitudes about literacy are addressed by Ward, Shook, and Marrion (1993) in their research regarding attitudes about writing in a cross-cultural setting. The study carried out by Ward et. al. (1993) in Lytton, British Columbia surveyed students in Grade 1 and two about what they thought the purpose of writing was, their personal writing preferences, and their self-concept as writers. The results indicate that aboriginal students were not able to list as many forms of writing as the non-aboriginal students, a higher proportion of aboriginal than non-aboriginal students enjoyed writing stories, and a slightly higher percentage of aboriginal than non-aboriginal children saw themselves as good writers.

This study in the Prince George school district examines aboriginal differences in reading and writing fluency as well as the previously mentioned gender differences. The earlier study done in Prince George by Hedekar in 1997 does not examine aboriginal differences in literacy due to the political direction given at that time; the Aboriginal Education Board did not want a separate study undertaken on aboriginal students (P. D. MacMillan, personal communication, June 3, 2004). As well, in Hedekar's 1997 study, relative age differences were examined with respect to reading and writing fluency but due to a lack of significant differences the variable of relative age was not included in this study.

CHAPTER THREE: METHODS

This chapter contains three sections. The first section describes the participants who were tested and how they were selected for the CBM/DIBELS norming project and this study. The second section explains the test instruments used for the CBM/DIBELS norming project and this study. The third section is a description of the procedures followed for my research.

Participants

This study uses the CBM/DIBELS norming data, which is an intact data set collected by teachers and district staff in School District #57 (Prince George) during the 2002-2003 school year. Therefore, this researcher did not select the participants or collect the data. The district (SD #57) deemed no signed consent forms for student participation were required because the data consisted of measures routinely collected by the school district. See Foulds (2002) or Fewster and MacMillan (2002) for earlier instances of these procedures. Participants were selected using stratified random sampling of the elementary school population from Kindergarten to Grade 7. Participants in the study comprise approximately 20% of the total elementary student population. Each school has provided approximately 20% of its total school population.

In the Technical Report of the CBM Norming Project, Fewster et. al. (2003) indicate there were a total of 2272 students used in the norming sample from Kindergarten to Grade 7. The breakdown for each grade is as follows: 245 Kindergarten students, 248 Grade 1 students, 265 Grade 2 students, 281 Grade 3 students, 308 Grade 4 students, 277 Grade 5 students, 313 Grade 6 students, and 335 Grade 7 students. Students participating in the norming project were tested three times throughout the

school year, once each in October, January and April. Data for all three norming periods was cleaned and entered into SPSS 9 (Fewster et. al., 2003). Therefore, no further cleaning of the data was required by this researcher.

Instruments

The Kindergarten and Grade 1 participants for the CBM/DIBELS norming project were given a different series of tests from their older counterparts. Both Kindergarten and Grade 1 students were tested on Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Letter Naming Fluency (LNF). Only Kindergarten students were tested on Initial Sound Fluency (ISF). Only Grade 1 students were tested on Oral Reading Fluency (ORF) and only in the January and April testing periods. The Kindergarten scores for PSF and NWF were recorded only for the January and April periods whereas these scores for the Grade 1 participants were scored for all three testing periods. Scores on LNF were recorded for Grade 1 students in October only, but were recorded for all three periods for the Kindergarten students. See Table 5 for a complete schedule of the testing times for the Kindergarten and Grade 1 DIBELS measures. See Table 6 for a complete description of the DIBELS variables.

Table 5

Schedule of Testing Periods for DIBELS Measures

DIBELS Measures	Fall (October)	Winter (January)	Spring (April)
Kindergarten			
ISF	X	X	
LNF	X	X	X
PSF		X	X
NWF		X	X
ORF			
Grade 1			
ISF			
LNF	X		
PSF	X	X	X
NWF	X	X	X
ORF			X

Participants from Grade 2 through Grade 7 were tested on Total Words Written (TWW), Words Spelled Correctly (WSC), and Words Read Correctly (WRC). Grade 1 students were also tested on TWW, WSC, and WRC but only for the April testing period. See Table 6 for variable descriptions. See Fewster et. al. (2003) for further details about any aspect of the norming project.

Table 6

Description of DIBELS and CBM Variables

Variable	Description
ISF	Number of correctly identified and produced initial sounds (of an orally presented word) in 1 minute
LNF	Number of letters (upper and lower case) correctly named in 1 minute
PSF	Number of correct phonemes (in 3- and 4-phoneme words) produced in 1 minute
NWF	Number of correct letter-sounds produced or read from nonsense words in 1 minute
ORF	Number of words read correctly on a 1 minute to read passage
WRC	Number of words read correctly on a 1 minute to read passage
WSC	Number of words spelled correctly in a 3 minute written response to a verbal cue
TWW	Total number of words written in a 3 minute written response to a verbal cue (highly correlated with WSC)

An analysis performed by Fewster et. al. (2003) in the Technical Report of the Curriculum Based Measurement Norming Project provides evidence that none of the probes used in the testing showed any significant difference in difficulty level from the others (p. 29). Therefore, for this study, the reading and writing probes used at each grade level will be considered equivalent. Also from Fewster et. al.'s (2003) analysis is evidence that there is a high correlation between Total Words Written (TWW) and Words Spelled Correctly (WSC), $.94 < r < .99$, and a low to moderate correlation between TWW and Words Read Correctly (WRC), $.27 < r < .49$. Correlations across the 6 month norming period for both TWW and WRC show consistency and good stability with

coefficients ranging from .59 to .65 and from .81 to .86 respectively (Fewster et. al., 2003).

For the DIBELS data in Fewster et. al.'s (2003) analysis at the Kindergarten level there is a low to moderate correlation among the four variables tested (ISF, LNF, PSF, and NWF), $.342 < r < .708$. Correlations across the 3 month norming period for PSF, ISF, and NWF and the 6 month norming period for LNF show consistency and good stability with coefficients of .687, .695, .741, and .649 respectively. The results for the Grade 1 DIBELS and CBM data indicates a low to high correlation among the seven variables tested (PSF, NWF, LNF, ORF, WRC, TWW, and WSC), $.231 < r < .944$. At the Grade 1 level for the three DIBELS variables that were tested more than once (i.e.: PSF, NWF, and ORF), the correlations across the 6 month (for PSF and NWF) and 3 month (for ORF) norming period show consistency and good stability with coefficients of .706, .645, and .903 respectively.

Procedures

The data that have been collected for School District #57 (Prince George) for the CBM/DIBELS Norming Study 2002/2003 will be used to investigate gender and aboriginal differences in Kindergarten to Grade 7 students with respect to their CBM reading, writing and DIBELS scores. The data were collected by the school district during the 2002-2003 school year, after which John Cook prepared a technical report for the school district under the supervision of Dr. Peter MacMillan of the University of Northern British Columbia. Due to the fact that this study is using an intact data set ethics approval was obtained from the University of Northern British Columbia prior to proposal approval. Relevant documentation is located in the Appendix.

The DIBELS data for Kindergarten and Grade 1 will be analyzed with a series of 2 X 2 gender-by-aboriginal status *ANOVA* using the SPSS statistical program to determine if there are any effects attributable to gender or aboriginal status and also for the variables of Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Oral Reading Fluency (ORF). A total of 18 *ANOVA*'s were performed in order to examine the five variables for Kindergarten and Grade 1 over the three testing periods. Descriptive statistics will be reported by grade and then by all the differences. Data will be examined across test periods and grades for consistency and then a Bonferroni correction will be applied, α / n (test) $= \alpha\beta$ (e.g. .05/3 = .016). No multivariate statistical testing will be applied.

The CBM data sample of students in Grade 1 through 7 will also be analyzed with a series of 2 X 2 gender-by-aboriginal status *ANOVA* using the SPSS statistical program. A determination will be made as to whether or not there are any effects attributable to gender or aboriginal status for the variables of Words Read Correctly (WRC), Words Spelled Correctly (WSC) and Total Words Written (TWW). A total of 57 *ANOVA*'s were performed in order to examine the three variables (WRC, WSC and TWW) for each grade level for the three different testing periods. Descriptive statistics will be reported by grade and then by all the differences. Data will be examined across test periods and grades for consistency and then a Bonferroni correction will be applied, α / n (test) $= \alpha\beta$ (e.g. .05/3 = .016). No multivariate statistical testing will be applied.

CHAPTER FOUR: RESULTS

The results of the data analysis will be discussed in three parts. Part one will discuss the results of the analysis of the DIBELS data. These data have been analyzed for differences in early literacy skills in Kindergarten and Grade 1 for gender, aboriginal status and the interaction between these two independent variables. The second part will discuss the results of the analysis of the CBM data. These data have been analyzed for differences in reading and writing fluency from Grade 2 to 7 for gender, aboriginal status and the interaction between these two variables. Part three will discuss effect sizes and trends for the analysis of both the DIBELS and CBM data.

Results of the DIBELS Data Analysis

The early literacy skills measured in this study include Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Oral Reading Fluency (ORF). Kindergarten and Grade 1 students were tested using these DIBELS variables during the recommended testing periods (see Table 5 in Chapter 3).

The sample sizes varied slightly from testing period to testing period and from grade to grade. The largest sample size was 252 for Grade 1 at the January testing of Phoneme Segmentation Fluency. The smallest sample size was 180 for Kindergarten at the January testing of Nonsense Word Fluency. The most common sample size was in the 240's. See Table 7 for sample sizes for all DIBELS results.

Table 7

Descriptive Statistics for ISF, LNF, PSF, NWF, and ORF

Grade and testing period	*N*	*M*	*SD*	Skew	*SE* of Skew	Kurtosis	*SE* of Kurtosis
ISF							
Kindergarten							
October	245	11.16	5.57	1.34	.16	2.24	.31
January	242	14.09	10.42	1.00	.16	1.05	.31
LNF							
Kindergarten							
October	245	10.04	11.41	2.07	.16	7.05	.31
January	243	20.06	14.94	.91	.16	1.50	.31
April	241	29.85	15.78	.32	.16	.00	.31
Grade 1							
October	248	33.17	17.04	.25	.16	- .58	.31
PSF							
Kindergarten							
January	242	14.31	15.06	1.26	.16	1.37	.31
April	240	20.65	16.41	.47	.16	- .84	.31
Grade 1							
October	248	24.50	19.05	.55	.16	- .67	.31
January	252	35.90	18.83	.06	.15	- .43	.31
April	231	41.07	16.44	- .42	.16	- .21	.32
NWF							
Kindergarten							
January	180	7.01	9.08	1.63	.18	2.38	.36
April	239	14.89	13.87	1.93	.16	7.26	.31
Grade 1							
October	249	19.77	17.06	1.91	.15	7.70	.31
January	251	37.41	21.48	.65	.15	.77	.31
April	233	53.59	30.40	.89	.16	.54	.32
ORF							
Grade 1							
January	250	19.73	20.79	1.93	.15	4.17	.31
April	232	39.24	28.29	1.03	.16	.73	.32

For each variable tested the mean score increased over the testing periods for each grade and across grades. The standard deviation also increased from testing period to testing period for each variable with the exception of the PSF testing for Grade 1, which shows a decrease in the standard deviation over the testing periods. The other statistic to note is the increase of the standard deviation for ISF from the October testing to the January testing at the Kindergarten level. This January standard deviation is almost twice that of the October standard deviation.

For a number of testing results the magnitude of skewness was six times the standard error. These cases include: the ISF testing for Kindergarten in October and January; the LNF testing for Grade 1 in October (see Figure 1); the PSF testing for Kindergarten in January; the NWF testing for Kindergarten in January and April and for Grade 1 in October; and the ORF testing for Grade 1 in both January and April. The skew in these cases would indicate that some students have acquired the skill being tested but most have not. One testing period is negatively skewed (the Grade 1 April testing of PSF). This raises little concern due to the assumption that for equal and unequal *n*'s, skewed populations have very little effect on the level of significance or power (Glass & Hopkins, 1996). In addition, the fact that a directional or one-tailed test is not being performed means the skew is of no consequence.

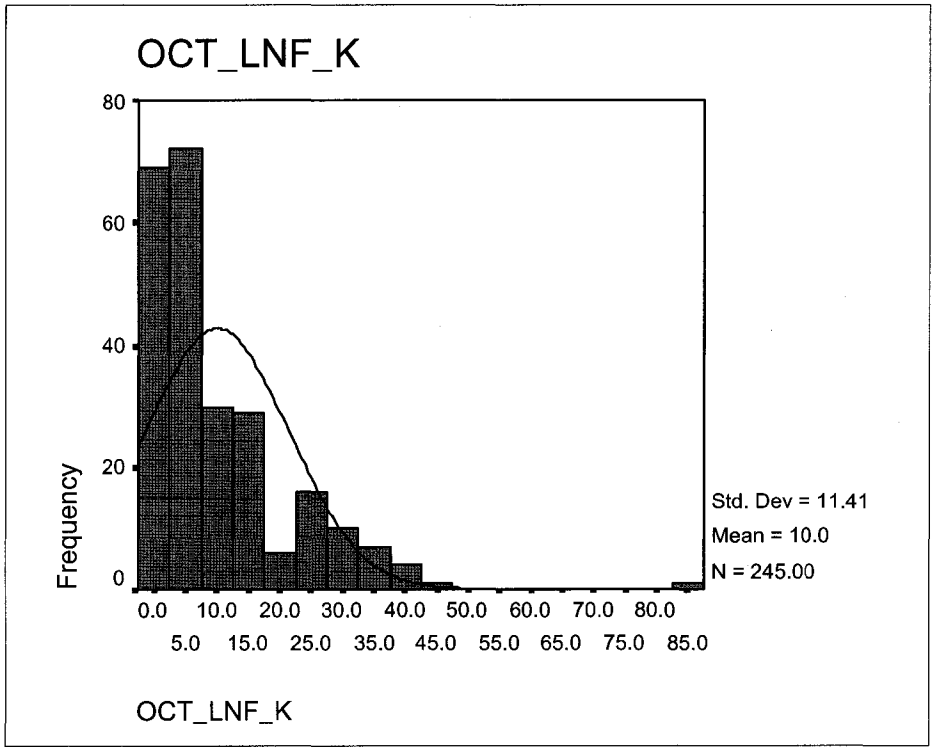

Figure 1. Histogram of scores and their frequency for the October testing of LNF at the Kindergarten level, showing a positively skewed, leptokurtic distribution.

A number of the testing results are leptokurtic with a kurtosis of six times the standard error. These cases include: the ISF testing for Kindergarten in October; the LNF testing for Kindergarten in October (see Figure 1); the NWF testing for Kindergarten in both January and April and for Grade 1 in October; and the ORF testing for Grade 1 in January. A number of the testing results are also platykurtic (see Table 7). The kurtosis effects are slight with the actual α being less than the nominal α in leptokurtic populations and the actual α exceeding the nominal α in platykurtic populations (Glass & Hopkins, 1996).

At the Kindergarten level a 2 X 2 between groups ANOVA (gender by aboriginal status) was run wherever data existed for the three testing periods (October, January, and

April). The four variables analyzed for Kindergarten are: initial sound fluency (ISF), letter naming fluency (LNF), phoneme segmentation fluency (PSF), and nonsense word fluency (NWF). At the Grade 1 level a 2 X 2 between groups *ANOVA* (gender by aboriginal status) was run for the three testing periods (October, January, and April) where data existed. The four variables analyzed for Grade 1 are: letter naming fluency (LNF), phoneme segmentation fluency (PSF), nonsense word fluency (NWF), and oral reading fluency (ORF). A total of 18 analyses of variance were calculated for Kindergarten and Grade 1 students. Values of F and p are reported in Table 8. The degrees of freedom between (J - 1) is always equal to 1 when there are two genders or two categories of aboriginal status. The degrees of freedom within (N - J) are always $N-2$ for the main effect and $N-JK$ for the interaction so for all analyses of variance these will not be shown in the respective tables. Summaries of the DIBELS analyses of variance are found in Table 8 for gender, aboriginal status and the interaction of gender and aboriginal status (G X Ab). Analysis of variance that are significant at $p < .05$ are marked with a single asterisk, analysis of variance that are significant at $p < .01$ are marked with a double asterisk.

To examine the assumption of homogeneity of variances the Levine's test was run for all analyses of variance and for all cases there was no violation of this assumption (all $p > .05$). Of the 18 analyses of variance performed, every calculation indicated there were no significant interactions between gender and aboriginal status for early literacy skills (all $p > .10$). Therefore all main effects can be interpreted without reference to any interaction. The results are found in the G X Ab rows of Table 8.

There are five cases where gender differences are evident for early literacy skills. All three testing periods for phoneme segmentation fluency (PSF) at the Grade 1 level had a significant gender difference: $F(1, 248) = 5.107, p < .05$; $F(1, 252) = 10.343, p < .05$; $F(1, 231) = 10.198, p < .05$. The other two cases where significant gender differences occurred were at the Grade 1 level for nonsense word fluency (NWF) in October and for oral reading fluency (ORF) in April: $F(1, 249) = 6.713, p < .05$; $F(1, 232) = 4.334, p < .05$. All other early literacy skills analyses did not indicate a significant gender difference. With only 5 of the 18 *ANOVA* results showing a significant gender difference, there is not consistent evidence of a gender difference across Kindergarten and Grade 1 for early literacy skills. If a modest Bonferroni correction for the two or three testing periods (e.g.: $\alpha / 2 = .025$, $\alpha / 3 = .016$) in a year is applied there would only be 3 of the 18 *ANOVA* results showing a significant gender difference.

A significant difference ($p < .05$) between aboriginal students and non-aboriginal students was detected in 15 of the 18 analyses of variance for early literacy skills. The three cases where the results were non-significant all occurred at the Kindergarten level. The non-significant results occurred in the January and April testing of nonsense word fluency (NWF) and in the April testing of phoneme segmentation fluency (PSF): $F(1, 180) = 2.824, p > .05$; $F(1, 239) = 1.779, p > .05$; $F(1, 240) = 2.677, p > .05$. With 15 of the 18 *ANOVA* results showing a significant difference there is consistent evidence of an aboriginal status/non-aboriginal status difference for early literacy skills across Kindergarten and Grade 1.

Table 8

Analysis of Variance for Gender and Aboriginal Diff in ISF, LNF, PSF, NWF, ORF

	October		January		April	
Source	*F*	*p*	*F*	*p*	*F*	*p*
ISF Kindergarten						
Gender	0.447	.504	0.348	.556		
Aboriginal	5.278	.022*	5.654	.018*		
G X Ab	0.028	.867	0.195	.659		
LNF Kindergarten						
Gender	2.339	.127	3.512	.062	3.119	.079
Aboriginal	7.599	.006**	4.151	.043*	4.480	.035*
G X Ab	0.213	.645	0.049	.826	0.054	.816
LNF Grade 1						
Gender	2.768	.097				
Aboriginal	9.965	.002**				
G X Ab	0.089	.765				
PSF Kindergarten						
Gender			0.374	.542	0.002	.963
Aboriginal			8.827	.003**	2.677	.103
G X Ab			1.256	.264	2.152	.144
PSF Grade 1						
Gender	5.107	.025*	10.343	.001**	10.198	.002**
Aboriginal	11.788	.001**	14.087	.000**	6.276	.013*
G X Ab	1.609	.206	0.172	.678	0.312	.577
NWF Kindergarten						
Gender			1.675	.197	0.366	.546
Aboriginal			2.824	.095	1.779	.184
G X Ab			2.161	.143	1.075	.301
NWF Grade 1						
Gender	6.713	.010*	2.050	.154	1.744	.188
Aboriginal	10.146	.002**	10.616	.001**	12.685	.000**
G X Ab	0.080	.778	0.320	.572	0.393	.531
ORF Grade 1						
Gender			1.241	.266	4.334	.038*
Aboriginal			10.874	.001**	15.529	.000**
G X Ab			0.019	.890	0.060	.807

Note: * $p < .05$, ** $p < .01$; $p < .0005$ is recorded as .000

Results of the CBM Data Analysis

Grade 2 to 7 students were tested for reading and writing literacy using CBM measures of Words Read Correctly (WRC), Words Spelled Correctly (WSC), and Total Words Written (TWW). Grade 1 students were also tested using all three CBM measures but were only tested in the April testing period. The sample sizes for these variables varied from testing period to testing period and grade to grade. The largest sample size was 335 (Grade 7) and the smallest sample size was 247 (Grade 1). The average sample size was 284.

For each of the three variables tested the mean score increased over the testing periods for each grade. The standard deviation remained relatively constant for the WRC results but for the WSC and TWW results the standard deviation doubled from the October testing in Grade 2 to the Grade 6 testing in October.

The majority of the testing results for WRC, WSC, and TWW are normally distributed with a skew of two times the standard error or less (see Figure 2 for an example of this); see Table 9, 10, and 11 for complete results. For a small number of test results the magnitude of skewness is six times the standard error and occurs at the Grade 1 level for WRC in April and at the Grade 2 level: once for the October testing of WRC; and again for the October and January testing of WSC. The highly positive skew for WSC at Grade 2 indicates that some students performed well at this skill but most students were not performing well at this skill. There are two testing periods that are very slightly negatively skewed. These include the Grade 6 January and April testing of WRC. This small number of skewed results raises little concern due to the assumption that for equal and unequal *n*'s, skewed populations have very little effect on the level of

significance or power (Glass & Hopkins, 1996). The fact that a directional or one-tailed test is not being performed means the skew is of no consequence.

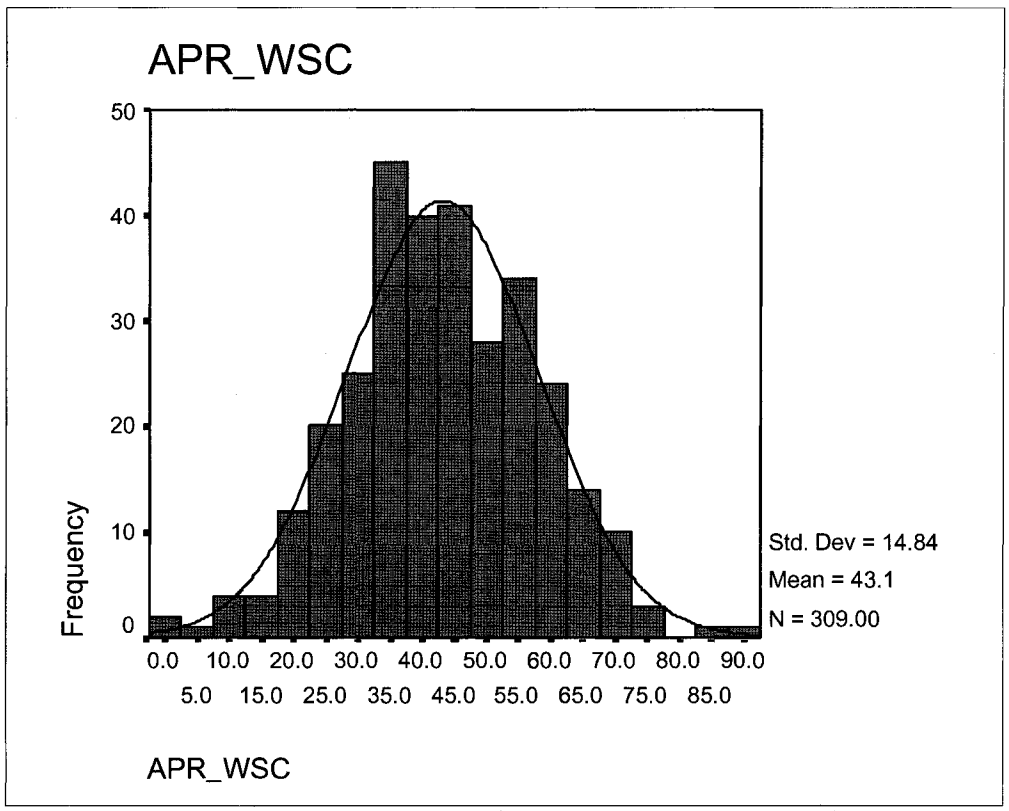

Figure 2. Histogram of scores and their frequency for the April testing of WSC at the Grade 4 level, showing a normal (non-skewed), and mesokurtic distribution.

The majority of testing results are also mesokurtic with a kurtosis of two times the standard error or less. A very small number of the testing results are leptokurtic with a kurtosis of six times the standard error. These three cases include: the Grade 2 October and April testing of WSC; and the Grade 3 October testing of TWW. A number of the testing results are also slightly platykurtic (see Tables 9, 10, and 11). The kurtosis effects are slight with the actual α being less than the nominal α in leptokurtic populations and

the actual α exceeding the nominal α in platykurtic populations (Glass & Hopkins, 1996).

Table 9

Descriptive Statistics for WRC

Grade and testing period	*n*	*M*	*SD*	Skew	*SE* of Skew	Kurtosis	*SE* of Kurtosis
Grade 1							
April	247	36.02	29.60	1.15	.16	.80	.31
Grade 2							
October	266	51.72	39.57	1.00	.15	.75	.30
January	264	67.65	39.80	.31	.15	-.93	.30
April	265	81.03	42.32	.33	.15	-.46	.30
Grade 3							
October	281	88.65	40.26	.29	.15	-.34	.29
January	282	101.72	41.62	.29	.15	-.01	.29
April	281	110.31	39.47	.18	.15	-.03	.29
Grade 4							
October	309	102.89	40.89	.10	.14	-.76	.28
January	309	114.07	40.13	.17	.14	-.37	.28
April	309	120.29	38.30	.07	.14	-.36	.28
Grade 5							
October	278	115.05	36.08	-.02	.15	-.37	.29
January	277	121.50	37.83	.08	.15	-.27	.29
April	276	130.57	38.55	-.09	.15	-.32	.29
Grade 6							
October	313	128.01	38.55	-.07	.14	-.32	.28
January	310	131.48	39.70	-.18	.14	-.18	.28
April	312	137.78	38.17	-.21	.14	.22	.28
Grade 7							
October	334	135.32	40.49	.29	.13	-.27	.27
January	335	139.16	40.66	.18	.13	-.22	.27
April	335	143.93	40.18	.14	.13	-.07	.27

Table 10

Descriptive Statistics for WSC

Grade and testing period	*n*	*M*	*SD*	Skew	*SE* of Skew	Kurtosis	*SE* of Kurtosis
Grade 1							
April	247	9.77	7.03	.92	.16	.50	.31
Grade 2							
October	264	12.72	8.05	1.21	.15	2.37	.30
January	267	17.98	9.47	1.01	.15	1.59	.30
April	265	22.63	10.83	.97	.15	1.70	.30
Grade 3							
October	281	23.00	10.94	.92	.15	1.52	.29
January	283	28.34	12.19	.39	.15	.01	.29
April	279	31.72	12.25	.32	.15	.34	.29
Grade 4							
October	307	32.07	12.54	.30	.14	.11	.28
January	307	36.42	13.22	.12	.14	-.05	.28
April	309	43.12	14.84	.02	.14	.11	.28
Grade 5							
October	278	40.84	14.21	.22	.15	-.20	.29
January	280	43.84	14.10	.03	.15	.07	.29
April	277	49.17	15.74	.23	.15	.79	.29
Grade 6							
October	313	51.01	16.54	.20	.14	.04	.28
January	312	53.36	16.17	.26	.14	-.06	.28
April	311	56.96	17.33	.15	.14	.68	.28
Grade 7							
October	335	59.40	16.43	.21	.13	-.12	.27
January	333	60.87	16.80	.36	.13	.37	.27
April	334	63.29	16.90	.41	.13	1.38	.27

Table 11

Descriptive Statistics for TWW

Grade and testing period	*n*	*M*	*SD*	Skew	*SE* of Skew	Kurtosis	*SE* of Kurtosis
Grade 1							
April	247	13.45	8.28	.71	.16	.13	.31
Grade 2							
October	264	16.80	8.84	.72	.15	1.04	.30
January	267	22.21	9.81	.66	.15	.88	.30
April	265	26.84	10.98	.73	.15	1.30	.30
Grade 3							
October	281	26.59	11.06	.96	.15	2.10	.29
January	283	31.90	12.16	.34	.15	.05	.29
April	279	35.01	12.39	.23	.15	.61	.29
Grade 4							
October	307	35.44	12.89	.36	.14	.22	.28
January	307	39.28	13.54	.10	.14	-.02	.28
April	309	46.03	15.00	.05	.14	.31	.28
Grade 5							
October	278	43.73	14.37	.23	.15	-.17	.29
January	280	46.53	14.26	-.03	.15	.25	.29
April	277	51.64	15.77	.20	.15	.90	.29
Grade 6							
October	313	53.75	16.38	.18	.14	.25	.28
January	312	55.71	15.87	.28	.14	.06	.28
April	311	59.14	17.19	.17	.14	.89	.28
Grade 7							
October	335	61.82	16.62	.24	.13	-.12	.27
January	333	63.20	16.99	.36	.13	.36	.27
April	334	65.40	16.77	.39	.13	1.40	.27

For the CBM data a 2x2 between groups *ANOVA* (gender by aboriginal status) was run for each testing period for each of the three variables: WRC, WSC, and TWW. A total of 57 analyses of variance Ire calculated. Values of F and p are reported in Tables 12, 13, and 14. The degrees of freedom between $(J-1)$ is always equal to 1 when there are two genders or two categories of aboriginal status. The degrees of freedom within $(N-J)$ are always $N-2$ for the main effect and $N-JK$ for the interaction so for all analyses of variance these will not be shown in the respective tables. Summaries of the CBM analyses of variance are found in Tables 12, 13, and 14 for gender, aboriginal status and the interaction of gender and aboriginal status (G X Ab). Analysis of variance that are significant at $p < .05$ are marked with a single asterisk, analysis of variance that are significant at $p < .01$ are marked with a double asterisk.

As with the DIBELS data, in order to examine the assumption of homogeneity of variances the Levine's test was run for all analyses of variance for the CBM data and for all cases there was no violation of this assumption (all $p > .05$).

Of the 57 analyses of variance performed, every calculation indicated there were no significant interactions between gender and aboriginal status for reading and writing fluency at the .05 probability level and only 3 of the 57 for which $p < .10$. The results are found in the G X Ab rows of Tables 12, 13, and 14 and the lack of interaction between gender and aboriginal status is well illustrated in Figure 3 by the parallel lines.

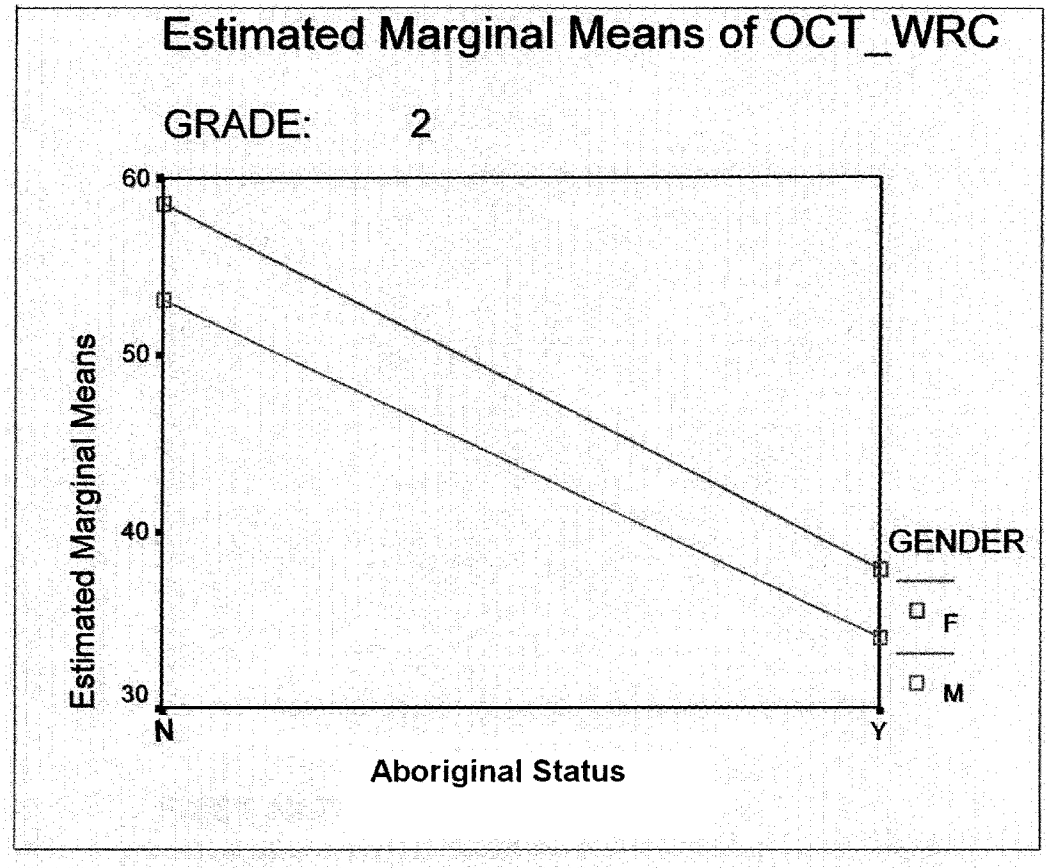

Figure 3. Line graph of estimated marginal means for male/female and aboriginal/non-aboriginal for the Grade 2 October testing of WRC, showing there is no interaction between the variables of gender and aboriginal status.

As illustrated in Figure 3, when the end points are subtracted the resulting gender gaps are approximately the same for both the non-aboriginal and aboriginal groups. When comparing the difference between the non-aboriginal and aboriginal end points the amount is approximately the same for both genders.

There are three cases where gender differences are evident for reading fluency (WRC). In 3 of the 19 *ANOVA*'s for WRC, significant gender differences were found. A significant gender difference was found for the April Grade 2 reading analysis: $F(1, 265) = 5.792, p < .05$ and the Grade 6 October and January reading analyses:

$F(1, 266) = 4.388, p < .05$; $F(1, 264) = 5.183, p < .05$. All other reading analyses did not indicate a significant gender difference. If a Bonferroni correction of $\alpha / 3 = .016$ is applied none of the 19 *ANOVA*'s indicate a significant difference.

In the case of writing fluency there is evidence of a significant gender difference. In 17 of the 19 *ANOVA*'s for WSC and in 18 of the 19 *ANOVA*'s for TWW significant gender differences were found ($p < .05$). The analyses which did not have a significant gender result were found in the Grade 2 October and Grade 3 October results for WSC: $F(1, 264) = 3.769, p > .05$; $F(1, 281) = 3.813, p > .05$. The other writing analysis that did not have a significant gender difference was the Grade 3 October result for TWW: $F(1, 264) = 1.346, p > .05$. The results for WSC and TWW are so similar because these two variables are very highly correlated. If a Bonferroni correction of $\alpha / 3 = .016$ is applied, 17 of the 19 *ANOVA*'s for WSC are still statistically significant.

With only 3 of the 19 *ANOVA* results showing a significant gender difference, there is not consistent evidence of a gender difference across all grades for reading fluency. In the case of writing fluency, with 17 of the 19 *ANOVA* results for WSC showing a significant gender difference, there is consistent evidence of a gender difference across all grades.

The *ANOVA* results for aboriginal differences for WRC, WSC, and TWW are found in Tables 12, 13, and 14 respectively. As previously mentioned a total of 57 analyses of variance were calculated. A significant difference between aboriginal students and non-aboriginal students was detected in 14 of the 19 ANOVA's for WRC. There were five cases where no significant differences in reading fluency between aboriginal and non-aboriginal students were found. These five cases are found at the

Grade 4 and 5 levels for various testing periods. These occurred in the January and April Grade 4 reading tests: $F(1, 309) = .673, p > .05$; $F(1, 309) = 3.718, p > .05$; and in the October, January, and April Grade 5 reading tests: $F(1, 278) = .000, p > .05$; $F(1, 277) = .060, p > .05$; $F(1, 276) = .003, p > .05$. With the lack of significant results showing an aboriginal, non-aboriginal difference at the Grade 4 and 5 level for reading fluency it is difficult to state that there is a difference across all grade levels but there is a significant difference in reading fluency between aboriginal and non-aboriginal students at the Grade 2, 3, 6, and 7 level. This also indicates a lack of an explainable trend.

A significant difference between aboriginal and non-aboriginal students for writing fluency is not as evident. In 7 of the 19 *ANOVA*'s calculated for WSC there was a significant difference detected. These differences occurred in the Grade 1 testing in April: $F(1, 247) = 9.632, p < .05$, the Grade 2 testing in January: $F(1, 267) = 12.26$, $p < .05$, the Grade 4 testing in October: $F(1, 307) = 9.49, p < .05$, the Grade 5 testing in April: $F(1, 277) = 6.403, p < .05$, the Grade 6 testing in January and April: $F(1, 312) = 3.975, p < .05$; $F(1, 311) = 4.331, p < .05$, and the Grade 7 testing in April: $F(1, 334) = 5.531, p < .05$. Due to the high correlation between WSC and TWW the results for TWW were very similar to those for WSC. The effect sizes for six out of seven of these significant differences were all small. With 12 of the 19 ANOVA results showing no significant difference between aboriginal and non-aboriginal students for WSC, there is not enough evidence showing a significant difference in writing fluency between aboriginal and non-aboriginal students across the grade levels which could be a sample size issue. Whether or not students are aboriginal does appear to impact on their

reading fluency scores at Grade 1, 2, 3, 6, and 7, but does not appear to have an impact on their writing fluency scores across all grade levels.

Table 12

Analysis of Variance for Gender and Aboriginal Differences in WRC

	October		January		April	
Source	*F*	*p*	*F*	*p*	*F*	*p*
Grade 1						
Gender					1.764	.185
Aboriginal					15.224	.000**
G X Ab					0.151	.698
Grade 2						
Gender	0.586	.445	2.238	.136	5.792	.017*
Aboriginal	10.773	.001**	16.384	.000**	20.180	.000**
G X Ab	0.015	.901	0.002	.965	0.000	.984
Grade 3						
Gender	0.991	.320	2.151	.144	1.595	.208
Aboriginal	10.001	.002**	9.086	.003**	9.023	.003**
G X Ab	0.022	.882	0.255	.614	0.000	.984
Grade 4						
Gender	0.170	.680	0.163	.687	0.048	.826
Aboriginal	4.956	.027*	0.673	.413	3.718	.055
G X Ab	0.048	.827	0.038	.846	0.177	.674
Grade 5						
Gender	0.014	.906	0.039	.843	0.144	.705
Aboriginal	0.000	.997	0.060	.807	0.003	.960
G X Ab	2.769	.097	3.546	.061	2.683	.103
Grade 6						
Gender	4.388	.037*	5.183	.024*	3.075	.080
Aboriginal	11.506	.001**	16.869	.000**	16.425	.000**
G X Ab	0.042	.837	0.211	.646	0.763	.383
Grade 7						
Gender	3.412	.066	3.108	.079	0.942	.333
Aboriginal	4.746	.030*	6.061	.014*	4.445	.036*
G X Ab	0.002	.963	0.353	.553	1.302	.255

Note: * $p < .05$, ** $p < .01$; $p < .0005$ is recorded as .000

Table 13

Analysis of Variance for Gender and Aboriginal Differences in WSC

	October		January		April	
Source	*F*	*p*	*F*	*p*	*F*	*p*
Grade 1						
Gender					5.168	.024*
Aboriginal					9.632	.002**
G X Ab					1.013	.315
Grade 2						
Gender	3.769	.053	9.983	.002**	8.080	.005**
Aboriginal	3.214	.074	12.260	.001**	2.482	.116
G X Ab	0.330	.566	0.093	.761	0.116	.734
Grade 3						
Gender	3.813	.052	9.579	.002**	17.654	.000**
Aboriginal	2.928	.088	2.823	.094	0.471	.493
G X Ab	1.127	.289	2.213	.138	1.082	.299
Grade 4						
Gender	20.716	.000**	15.025	.000**	7.870	.005**
Aboriginal	9.490	.002**	2.751	.098	3.188	.075
G X Ab	0.925	.337	0.420	.517	0.007	.932
Grade 5						
Gender	10.779	.001**	15.598	.000**	11.963	.001**
Aboriginal	1.248	.265	0.200	.655	6.403	.012*
G X Ab	1.781	.183	1.563	.212	1.686	.195
Grade 6						
Gender	18.686	.000**	13.428	.000**	12.629	.000**
Aboriginal	2.319	.129	3.975	.047*	4.331	.038*
G X Ab	0.041	.839	0.540	.463	0.408	.523
Grade 7						
Gender	19.773	.000**	20.017	.000**	27.485	.000**
Aboriginal	0.561	.454	1.017	.314	5.531	.019*
G X Ab	0.005	.944	0.100	.752	0.299	.585

Note: * $p < .05$, ** $p < .01$; $p < .0005$ is recorded as .000

Table 14

Analysis of Variance for Gender and Aboriginal Differences in TWW

	October		January		April	
Source	*F*	*p*	*F*	*p*	*F*	*p*
Grade 1						
Gender					3.898	.049*
Aboriginal					8.689	.004**
G X Ab					1.249	.265
Grade 2						
Gender	6.881	.009**	8.923	.003**	4.794	.029*
Aboriginal	2.567	.110	10.617	.001**	1.691	.195
G X Ab	0.005	.945	0.177	.674	0.189	.664
Grade 3						
Gender	1.346	.247	5.618	.018*	14.800	.000**
Aboriginal	1.939	.165	1.402	.237	0.171	.680
G X Ab	2.196	.140	3.619	.058	0.788	.375
Grade 4						
Gender	22.966	.000**	15.703	.000**	9.470	.002**
Aboriginal	8.798	.003**	3.116	.079	2.840	.093
G X Ab	1.613	.205	0.566	.452	0.217	.642
Grade 5						
Gender	10.412	.001**	17.012	.000**	10.717	.001**
Aboriginal	1.507	.221	0.338	.561	6.253	.013*
G X Ab	2.826	.094	3.314	.070	2.026	.156
Grade 6						
Gender	17.671	.000**	11.384	.001**	11.876	.001**
Aboriginal	1.575	.210	2.894	.090	3.303	.070
G X Ab	0.236	.628	0.584	.445	0.237	.627
Grade 7						
Gender	14.807	.000**	18.823	.000**	24.998	.000**
Aboriginal	0.286	.593	0.785	.376	5.175	.024*
G X Ab	0.052	.819	0.252	.616	0.519	.472

Note: * $p < .05$, ** $p < .01$; $p < .0005$ is recorded as .000

Effect Sizes and Trends

Effect Sizes for the DIBELS Data

With this many analyses of variance being run it is necessary to calculate Cohen's *d* for ISF, LNF, PSF, NWF, and ORF for the appropriate grade level(s) and testing periods. They are reported by gender first in Table 15 and then by aboriginal status in Table 16. Effect sizes where there was a significant difference of $p < .05$ are marked with a single asterisk.

For early literacy skills the statistically significant difference effect size for gender ranges from small (.37 to .46) to medium (.56) with the median effect size being at the upper end of small (.45). There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect small effects based on sample size when the effect is below the upper end of small yet not so sensitive as to detect statistically significant but trivial effects.

The only test variables that indicate statistically significant effect sizes for gender are all at the Grade 1 level and include PSF, NWF, and ORF. The small number of statistically significant differences is due to a consistent lack of differences in performance on the early literacy skills test variables. The lack of consistent statistically significant differences in the sample indicate non-significant results, suggesting for the other Grade 1 results (LNF) and all the Kindergarten results no difference is detected. Therefore I do not believe there are gender differences in early literacy skills in the population.

Table 15

Effect Sizes for ISF, LNF, PSF, NWF, and ORF by Gender

	Female			Male			
Grade and testing period	*SD*	*n*	*M*	*N*	*M*	Effect Size	
ISF							
Kindergarten							
October	8.57	121	11.60	124	10.74	.10	Trivial
January	10.42	120	14.88	122	13.31	.15	Trivial
LNF							
Kindergarten							
October	11.41	121	11.87	124	8.25	.32	Small
January	14.94	121	22.64	122	17.51	.34	Small
April	15.78	116	32.54	125	27.34	.33	Small
Grade 1							
October	17.04	117	35.78	131	30.84	.29	Small
PSF							
Kindergarten							
January	15.06	120	16.05	122	12.60	.23	Small
April	16.41	116	22.03	124	19.53	.15	Trivial
Grade 1							
October	19.05	117	29.05	131	20.44	.45	Small*
January	18.83	118	40.50	134	31.86	.46	Small*
April	16.44	108	45.94	123	36.79	.56	Medium*
NWF							
Kindergarten							
January	9.08	88	9.07	92	5.03	.45	Small
April	13.87	116	16.52	123	13.35	.23	Small
Grade 1							
October	17.06	118	23.16	131	16.72	.38	Small*
January	21.48	118	39.54	133	35.53	.19	Trivial
April	30.40	110	56.35	123	51.13	.17	Trivial
ORF							
Grade 1							
January	20.79	118	21.91	132	17.78	.20	Trivial
April	28.29	110	44.73	122	34.30	.37	Small*

Note: * denotes cases where there was a significant gender difference ($p < .05$)

Also for early literacy skills the statistically significant difference effect size for aboriginal status ranges from small (.35 to .48) to medium (.50 to .63) with the median effect size being at the lower end of medium (.50). There are no statistically significant effects of trivial size. The analysis is sensitive enough to detect medium effects based on sample size when the effect is below the lower end of medium yet not so sensitive as to detect statistically significant but trivial effects.

The only test variable that does not indicate statistically significant effect sizes for aboriginal status is NWF at the Kindergarten level. The presence of statistically significant differences is due to consistent differences in performance on the early literacy skills test variables. Statistically significant differences in the sample indicate significant results, therefore I believe there are aboriginal differences in early literacy skills in the population.

The statistically significant difference effect for aboriginal status is slightly greater than the statistically significant difference effect for gender when comparing the median statistically significant effect sizes for the two groups. For gender the median effect size is .45 and for aboriginal status the median effect size is .50. This would merely be a sample size artifact as aboriginal groups are approximately 40 to 55 whereas gender groups are approximately 90 to 130.

Table 16

Effect Sizes for ISF, LNF, PSF, NWF, and ORF by Aboriginal Status

		Non-Aboriginal		Aboriginal			
Grade and testing period	*SD*	*n*	*M*	*n*	*M*	Effect	Size
ISF							
Kindergarten							
October	8.57	205	11.72	40	8.31	.40	Small*
January	10.42	202	14.79	40	10.55	.41	Small*
LNF							
Kindergarten							
October	11.41	205	10.91	40	5.55	.47	Small*
January	14.94	203	20.91	40	15.75	.35	Small*
April	15.78	202	30.77	39	25.08	.36	Small*
Grade 1							
October	17.04	193	34.99	55	26.78	.48	Small*
PSF							
Kindergarten							
January	15.06	202	15.56	40	8.00	.50	Medium*
April	16.41	201	21.39	39	16.85	.28	Small
Grade 1							
October	19.05	193	26.63	55	17.02	.50	Medium*
January	18.83	198	38.26	54	27.26	.58	Medium*
April	16.44	182	42.48	49	35.84	.40	Small*
NWF							
Kindergarten							
January	9.08	151	7.48	29	4.55	.32	Small
April	13.87	201	15.39	38	12.21	.23	Small
Grade 1							
October	17.06	194	21.64	55	13.18	.50	Medium*
January	21.48	197	39.79	54	28.74	.51	Medium*
April	30.40	184	57.32	49	39.61	.58	Medium*
ORF							
Grade 1							
January	20.79	197	21.97	53	11.38	.51	Medium*
April	28.29	183	42.99	49	25.24	.63	Medium*

Note: * denotes cases where there was a significant difference ($p < .05$) between aboriginal and non-aboriginal students

Effect Sizes for the CBM Data

As previously mentioned, with the number of analyses of variance being run it is necessary to calculate Cohen's *d* for the WRC, WSC, and TWW analyses for each grade level and each testing period. These effect sizes are reported by gender first in Tables 17, 18, and 19 and then by aboriginal status in Tables 20, 21, and 22 respectively. Effect sizes where there was a significant difference of $p < .05$ are marked with a single asterisk.

For WRC the statistically significant difference effect size for gender is small (.32 to .37) with the median effect size being at the mid-range of small (.35). There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect small effects based on sample size when the effect is below the mid-range of small but not so sensitive as to declare trivial effects to be statistically significant. The lack of statistically significant differences is due to a consistent lack of differences in performance on the WRC test variable. The lack of statistically significant differences in the sample indicates non-significant results, therefore I do not believe there are gender differences in reading in the population.

For WSC the statistically significant difference effect size for gender ranges from small (.38 to .44) to medium (.50 to .67) with the median effect size being at the lower end of medium (.55). There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect medium effects based on sample size when the effect is below the lower end of medium but not so sensitive as to declare trivial effects to be statistically significant. The presence of statistically significant differences is due to consistent differences in performance on the WSC test variable. Statistically significant

differences in the sample indicate significant results, therefore I believe there are gender differences in writing in the population.

For TWW the statistically significant difference effect size for gender also ranges from small (.33 to .49) to medium (.52 to .63) with the median effect size being at the lower end of medium (.52). There are no statistically significant effects that are trivial in size. The analysis is also sensitive enough to detect medium effects based on sample size when the effect is below the lower end of medium but not so sensitive as to declare trivial effects to be statistically significant. The presence of statistically significant differences is due to consistent differences in performance on the TWW test variable. As with WSC the statistically significant differences in the sample lead me to believe there are gender differences in writing in the population.

Table 17

Effect Sizes for WRC by Gender

Grade and testing period	Female *SD*	Female *n*	Female *M*	Male *n*	Male *M*	Effect Size	
Grade 1							
April	29.60	115	40.03	132	32.53	.25	Small
Grade 2							
October	39.57	120	54.77	146	49.21	.14	Trivial
January	39.80	115	73.21	149	63.36	.25	Small
April	42.32	118	89.70	147	74.07	.37	Small*
Grade 3							
October	40.26	126	92.33	155	85.66	.17	Trivial
January	41.62	122	106.57	160	98.02	.21	Small
April	39.47	124	115.23	157	106.42	.22	Small
Grade 4							
October	40.89	154	104.79	155	101.01	.09	Trivial
January	40.13	156	115.83	153	112.27	.09	Trivial
April	38.30	156	122.03	153	118.53	.09	Trivial
Grade 5							
October	36.08	140	118.82	138	111.23	.21	Small
January	37.83	139	126.54	138	116.43	.27	Small
April	38.55	141	135.34	135	125.58	.25	Small
Grade 6							
October	38.55	151	134.38	162	122.07	.32	Small*
January	39.70	152	138.65	158	124.58	.35	Small*
April	38.17	149	144.11	163	131.99	.32	Small
Grade 7							
October	40.49	164	140.64	170	130.19	.26	Small
January	40.66	166	145.28	169	133.15	.30	Small
April	40.18	171	148.64	164	139.03	.24	Small

Note: * denotes cases where there was a significant gender difference ($p < .05$)

Table 18

Effect Sizes for WSC by Gender

Grade and testing period	Female *SD*	Female *n*	Female *M*	Male *n*	Male *M*	Effect Size	
Grade 1							
April	7.03	116	11.45	131	8.27	.45	Small*
Grade 2							
October	8.05	119	14.31	145	11.42	.36	Small
January	9.47	118	20.65	149	15.87	.51	Medium*
April	10.83	118	25.08	147	20.67	.41	Small*
Grade 3							
October	10.94	126	25.73	155	20.78	.45	Small
January	12.19	122	33.01	161	24.80	.67	Medium*
April	12.25	124	35.68	155	28.55	.58	Medium*
Grade 4							
October	12.54	153	35.75	154	28.43	.58	Medium*
January	13.22	155	39.90	152	32.88	.53	Medium*
April	14.84	156	46.38	153	39.79	.44	Small*
Grade 5							
October	14.21	139	43.54	139	38.14	.38	Small*
January	14.10	140	47.39	140	40.29	.50	Medium*
April	15.74	140	52.23	137	46.05	.39	Small*
Grade 6							
October	16.54	151	55.96	162	46.40	.58	Medium*
January	16.17	152	57.97	160	48.98	.56	Medium*
April	17.33	149	61.84	162	52.48	.54	Medium*
Grade 7							
October	16.43	165	64.67	170	54.28	.63	Medium*
January	16.80	166	65.99	167	55.78	.61	Medium*
April	16.90	170	68.86	164	57.51	.67	Medium*

Note: * denotes cases where there was a significant gender difference ($p < .05$)

Table 19

Effect Sizes for TWW by Gender

Grade and testing period	Female *SD*	Female *n*	Female *M*	Male *n*	Male *M*	Effect Size	
Grade 1							
April	8.28	116	15.28	131	11.83	.42	Small*
Grade 2							
October	8.84	119	18.75	145	15.20	.40	Small*
January	9.81	118	24.91	149	20.08	.49	Small*
April	10.98	118	29.17	147	24.97	.38	Small*
Grade 3							
October	11.06	126	28.86	155	24.75	.37	Small
January	12.16	122	36.11	161	28.71	.61	Medium*
April	12.39	124	38.76	155	32.02	.54	Medium*
Grade 4							
October	12.89	153	39.24	154	31.67	.59	Medium*
January	13.54	155	42.86	152	35.63	.53	Medium*
April	15.00	156	49.32	153	42.67	.44	Small*
Grade 5							
October	14.37	139	46.12	139	41.35	.33	Small*
January	14.26	140	49.84	140	43.21	.47	Small*
April	15.77	140	54.36	137	48.85	.35	Small*
Grade 6							
October	16.38	151	58.34	162	49.48	.54	Medium*
January	15.87	152	59.95	160	51.68	.52	Medium*
April	17.19	149	63.76	162	54.90	.52	Medium*
Grade 7							
October	16.62	165	66.62	170	57.15	.57	Medium*
January	16.99	166	68.10	167	58.34	.57	Medium*
April	16.77	170	70.56	164	60.05	.63	Medium*

Note: * denotes cases where there was a significant gender difference ($p < .05$)

For differences in aboriginal versus non-aboriginal status the statistically significant difference effect size for WRC ranges from small (.29 to .48) to medium (.51 to .68) with the median effect size being at the lower end of medium (.51). There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect medium effects based on sample size when the effect is below the lower end of medium but not so sensitive as to declare trivial effects to be statistically significant. The presence of statistically significant differences is due to consistent differences in performance on the WRC test variable. Statistically significant differences in the sample indicate significant results therefore I believe there are aboriginal differences in reading in the population.

For aboriginal status differences the statistically significant difference effect size for WSC ranges from small (.24 to .48) to medium (.53) with the median effect size being at the mid-range of small (.31). The statistically significant effect sizes go from medium to small in a progression from Grade 2 to Grade 7. There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect small effects based on sample size when the effect is below the mid-range of small but not so sensitive as to declare trivial effects to be statistically significant. The lack of significant differences is due to a consistent lack of difference in performance on the WSC test variable. The lack of statistically significant differences in the sample indicate non-significant results, therefore I do not believe there are aboriginal differences in writing in the population.

For TWW the statistically significant difference effect size is small (.23 to .49) for aboriginal status differences. The median effect size is at the mid-range of small (.39).

There are no statistically significant effects that are trivial in size. The analysis is sensitive enough to detect small effects based on sample size when the effect is below the mid-range of small but not so sensitive as to declare trivial effects to be statistically significant. The lack of significant differences is due to a consistent lack of difference in performance on the TWW test variable. The lack of statistically significant differences in the sample indicate non-significant results, therefore I do not believe there are aboriginal differences in writing in the population.

Table 20

Effect Sizes for WRC by Aboriginal Status

Grade and testing period	SD	Non-Aboriginal n	M	Aboriginal n	M	Effect Size	
Grade 1							
April	29.60	197	39.63	50	21.78	.60	Medium*
Grade 2							
October	39.57	214	55.63	52	35.62	.51	Medium*
January	39.80	212	72.54	52	47.71	.62	Medium*
April	42.32	214	86.56	51	57.86	.68	Medium*
Grade 3							
October	40.26	238	91.97	43	70.26	.54	Medium*
January	41.62	239	105.14	43	82.74	.54	Medium*
April	39.47	240	113.25	41	93.05	.51	Medium*
Grade 4							
October	40.89	261	105.11	48	90.83	.35	Small*
January	40.13	263	114.87	46	109.52	.13	Trivial
April	38.30	263	122.05	46	110.24	.31	Small
Grade 5							
October	36.08	235	115.12	43	114.67	.01	Trivial
January	37.83	237	121.40	40	122.10	-.02	Trivial
April	38.55	232	130.66	44	130.05	.02	Trivial
Grade 6							
October	38.55	251	131.67	62	113.18	.48	Small*
January	39.70	246	136.09	64	113.78	.56	Medium*
April	38.17	250	141.97	62	120.85	.55	Medium*
Grade 7							
October	40.49	279	137.27	55	125.44	.29	Small*
January	40.66	281	141.40	54	127.54	.34	Small*
April	40.18	280	146.00	55	133.44	.31	Small*

Note: * denotes cases where there was a significant difference ($p < .05$) between aboriginal and non-aboriginal students

Table 21

Effect Sizes for WSC by Aboriginal Status

Grade and testing period	*SD*	Non-Aboriginal *n*	*M*	Aboriginal *n*	*M*	Effect Size	
Grade 1							
April	7.03	197	10.44	50	7.10	.48	Small*
Grade 2							
October	8.05	213	13.15	51	10.94	.28	Small
January	9.47	214	18.97	53	13.98	.53	Medium*
April	10.83	214	23.17	51	20.39	.26	Small
Grade 3							
October	10.94	238	23.46	43	20.44	.28	Small
January	12.19	240	28.82	43	25.65	.26	Small
April	12.25	240	32.00	39	29.95	.17	Trivial
Grade 4							
October	12.54	259	33.01	48	27.04	.48	Small*
January	13.22	261	36.97	46	33.33	.28	Small
April	14.84	264	43.80	45	39.13	.32	Small
Grade 5							
October	14.21	235	41.14	43	39.21	.14	Trivial
January	14.10	238	43.81	42	44.02	-.01	Trivial
April	15.74	233	49.97	44	44.93	.32	Small*
Grade 6							
October	16.54	251	51.77	62	47.94	.23	Small
January	16.17	248	54.27	64	49.81	.28	Small*
April	17.33	249	57.95	62	52.98	.29	Small*
Grade 7							
October	16.43	280	59.50	55	58.85	.04	Trivial
January	16.80	279	61.02	54	60.09	.06	Trivial
April	16.90	280	63.94	54	59.94	.24	Small*

Note: * denotes cases where there was a significant difference ($p < .05$) between aboriginal and non-aboriginal students

Table 22

Effect Sizes for TWW by Aboriginal Status

Grade and testing period	*SD*	Non-Aboriginal *n*	*M*	Aboriginal *n*	*M*	Effect Size	
Grade 1							
April	8.28	197	14.21	50	10.48	.45	Small*
Grade 2							
October	8.84	213	17.24	51	14.96	.26	Small
January	9.81	214	23.17	53	18.36	.49	Small*
April	10.98	214	27.26	51	25.08	.20	Small
Grade 3							
October	11.06	238	26.94	43	24.65	.21	Small
January	12.16	240	32.20	43	30.23	.16	Trivial
April	12.39	240	35.22	39	33.77	.12	Trivial
Grade 4							
October	12.89	259	36.37	48	30.44	.46	Small*
January	13.54	261	39.87	46	35.91	.29	Small
April	15.00	264	46.69	45	42.13	.30	Small
Grade 5							
October	14.37	235	44.07	43	41.88	.15	Trivial
January	14.26	238	46.52	42	46.57	-.00	Trivial
April	15.77	233	52.43	44	47.41	.32	Small*
Grade 6							
October	16.38	251	54.39	62	51.15	.20	Small
January	15.87	248	56.48	64	52.72	.24	Small
April	17.19	249	60.01	62	55.66	.25	Small
Grade 7							
October	16.62	280	61.87	55	61.55	.02	Trivial
January	16.99	279	63.31	54	62.67	.04	Trivial
April	16.77	280	66.03	54	62.19	.23	Small*

Note: * denotes cases where there was a significant difference ($p < .05$) between aboriginal and non-aboriginal students

When comparing the median effect sizes for the variables WRC, WSC, and TWW, the significant difference effect for gender is greater than the significant difference effect for aboriginal status for WSC and TWW only. For gender the WSC and TWW median effect sizes are .55 and .52 respectively and for aboriginal status the WSC and TWW median effect sizes are .31 and .39 respectively. For the variable WRC the significant difference effect for aboriginal status (median effect size of .51) is greater than the significant difference effect for gender (median effect size of .35). The non-significant results for aboriginal status could be due to the smaller sample size for this group.

DIBELS and CBM Data Trends

There are some trends that are evident in the data for both the DIBELS and CBM measures. One trend is that female participants outperform male participants in all measures for early literacy skills for Kindergarten and Grade 1, as well as for both the reading and writing measures for Grade 2 to seven students. This is based on a comparison of the mean scores for the variables tested in the DIBELS and CBM studies. Despite this overall trend of females outperforming males the only statistically significant gender differences detected are for writing for Grade 1 to 7 and for some of the early literacy skills for Grade 1.

There is a lack of a noticeable trend when examining the gap between female and male performance, males are not improving or regressing when comparing the mean scores from grade to grade and from testing period to testing period. For both genders from Kindergarten to Grade 7 there is an increase in mean scores from one testing period to another for each and every test variable.

When comparing effect sizes by gender for reading and writing fluency, PSF and LNF were the measures used for Kindergarten, and WRC and WSC were the measures used for Grade 1 to 7. For ease of graphical comparison, PSF for Kindergarten will be graphed along with WRC for Grade 1 to 7 and LNF for Kindergarten will be graphed along with WSC for Grade 1 to 7 (see Figure 4). There is an increasing trend in the effect size for both WRC and WSC from Kindergarten to Grade 7 with effect sizes for both measures roughly doubling between Kindergarten and Grade 7.

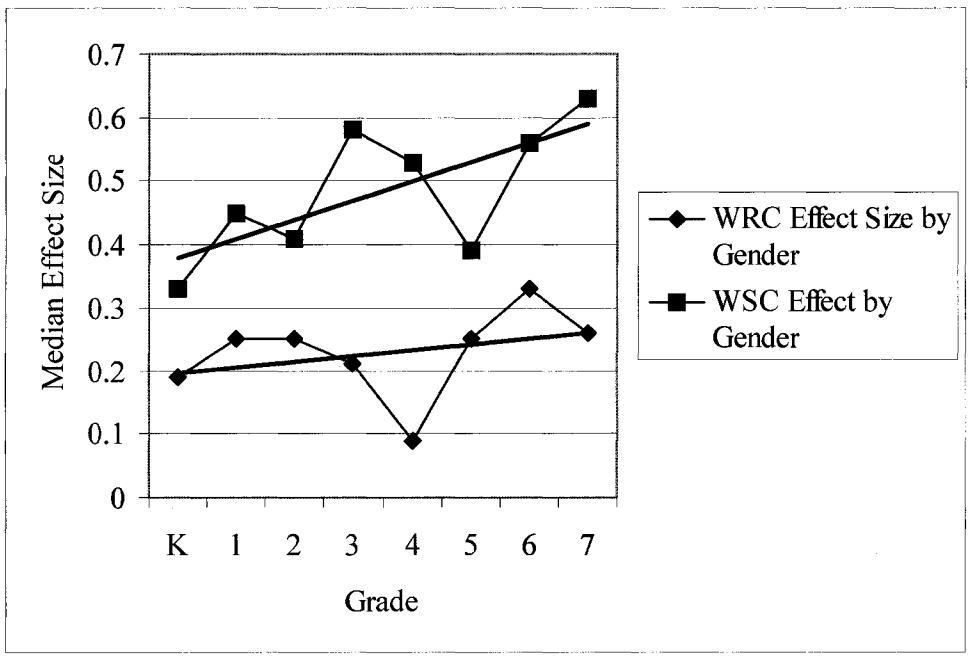

Figure 4. Line graph of median effect sizes for reading and writing by gender.

A trend for aboriginal status that is evident is that non-aboriginal students outperformed aboriginal students at every grade level and test variable except for the Grade 5 testing in January for all three test variables. This is based on a comparison of mean scores for all testing variables and periods for both the DIBELS and CBM data. PSF and LNF were the measures used for comparison at the Kindergarten and Grade 1

levels for reading and writing fluency. For both non-aboriginal and aboriginal students from Kindergarten to Grade 7 there is an increase in mean scores from one testing period to another for each and every test variable except at the Grade 7 level for the April testing of WSC. In this case the April mean test score for aboriginal students dropped from the January mean score.

When comparing effect sizes by aboriginal status for reading and writing fluency, PSF and LNF were the measures used for Kindergarten, and WRC and WSC were the measures used for Grade 1 to 7. For ease of graphical comparison, PSF for Kindergarten will be graphed along with WRC for Grade 1 to 7 and LNF for Kindergarten will be graphed along with WSC for Grade 1 to 7 (see Figure 5). Overall there is a decreasing trend in the effect sizes for both WRC and WSC from Grade 1 to Grade 7. This decrease indicates the difference between non-aboriginal students and aboriginal students is getting slightly smaller as the students reach the higher grade levels.

There is however a notable increase in the effect size by aboriginal status from Kindergarten to Grade 1. For WRC the increase from Kindergarten to Grade 1 is roughly doubled. This is an indication that the difference between non-aboriginal students and aboriginal students is getting larger at these two grade levels.

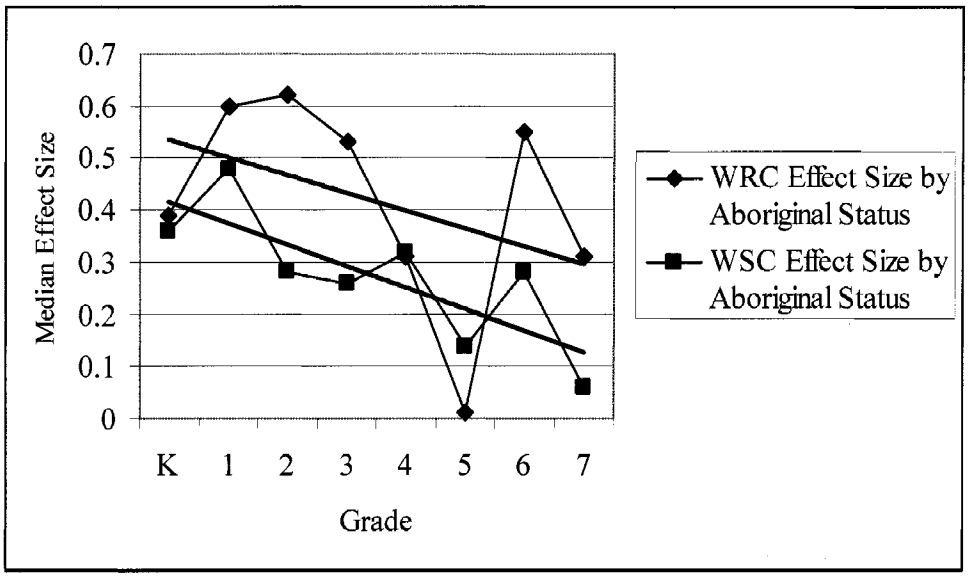

Figure 5. Line graph of median effect sizes for reading and writing by aboriginal status.

There is an anomaly at the Grade 5 level. Particularly for WRC, the effect size by aboriginal status for all three testing periods for this variable is smaller than the effect sizes for the other grade levels. The effect size for both WRC and WSC are trivial at the Grade 5 level. It is also at the Grade 5 level where aboriginal students outperformed non-aboriginal students in the January testing period for WRC, WSC and TWW.

When the median of the three means within grade scores for WRC for both gender and aboriginal status are plotted, there is a steady increase from Grade 1 right through to Grade 7. This increase is evident for all four groups: female; male; non-aboriginal; and aboriginal students (see Figure 6). Results for Kindergarten are not included in this graph because there is no accurate measure for Kindergarten for reading. There is a similar result seen for the median of the three means within grade scores for WSC for gender and aboriginal status (see Figure 7). The scores increase steadily from Grade 1 right through to Grade 7 for all four groups, female, male, non-aboriginal, and

aboriginal students. Results for Kindergarten are not included in this graph because there is no accurate measure for Kindergarten for words spelled or written correctly.

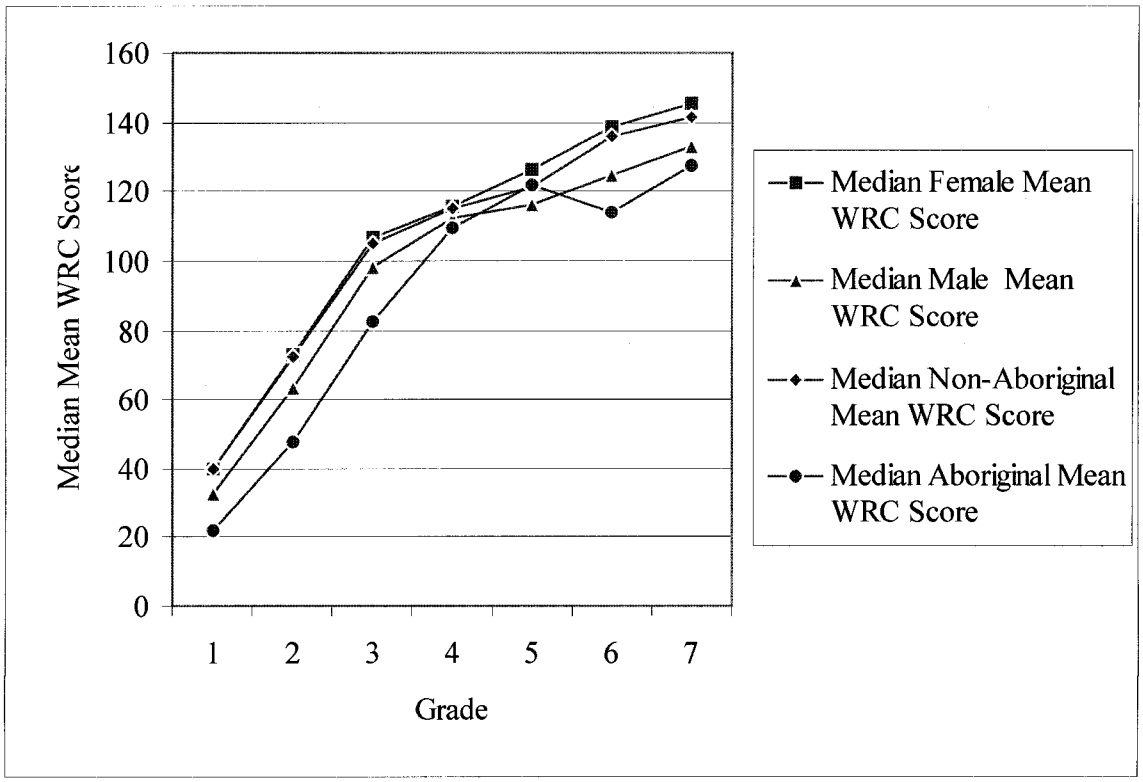

Figure 6. Line graph of median of the three means within grade scores for WRC for both gender and aboriginal status groups.

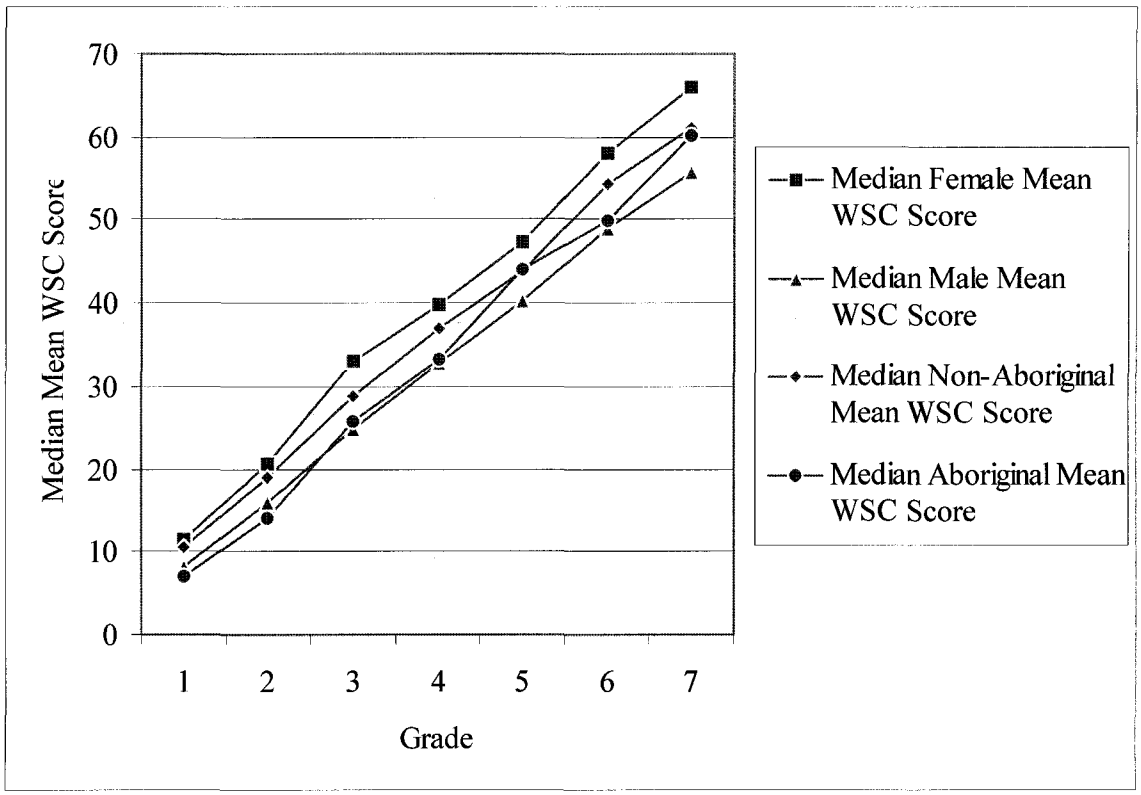

Figure 7. Line graph of median of the three means within grade scores for WSC for both gender and aboriginal status groups.

These two figures (6 and 7) show the gap between males and females is beginning to widen at the upper grades for both WRC and WSC while the gap between non-aboriginal and aboriginal students is beginning to narrow for both WRC and WSC. When comparing the median effect sizes by both gender and aboriginal status for WRC it is evident that there is a downward trend in effect sizes for the aboriginal results and an upward trend in the effect sizes for the gender results (see Figures 8). The effect sizes are more apparent when plotted on a line graph. The linear regressions and *R*-squared values for each trend line are as follows: $y = 0.012x + 0.1689$ and $R^2 = 0.1647$ for WRC effect size by gender; and $y = -0.025x + 0.515$ and $R^2 = 0.0838$ for WRC effect size by aboriginal status. The slope for the gender results is positive which again is evidence of

an upward trend in effect sizes for gender which means females are continuing to pull ahead of the males. The slope for aboriginal status is negative which is further evidence of a downward trend in effect sizes for aboriginal status which means that aboriginal students are getting closer to non-aboriginal students.

The effect sizes for WRC for gender are all in the range of just under 0.1 to just over 0.3, which is considered small for Cohen's limits. There is also a large dip in the effect size for WRC for aboriginal status at the Grade 5 level where Cohen's *d* drops to 0. This appears to be some sort of an anomaly.

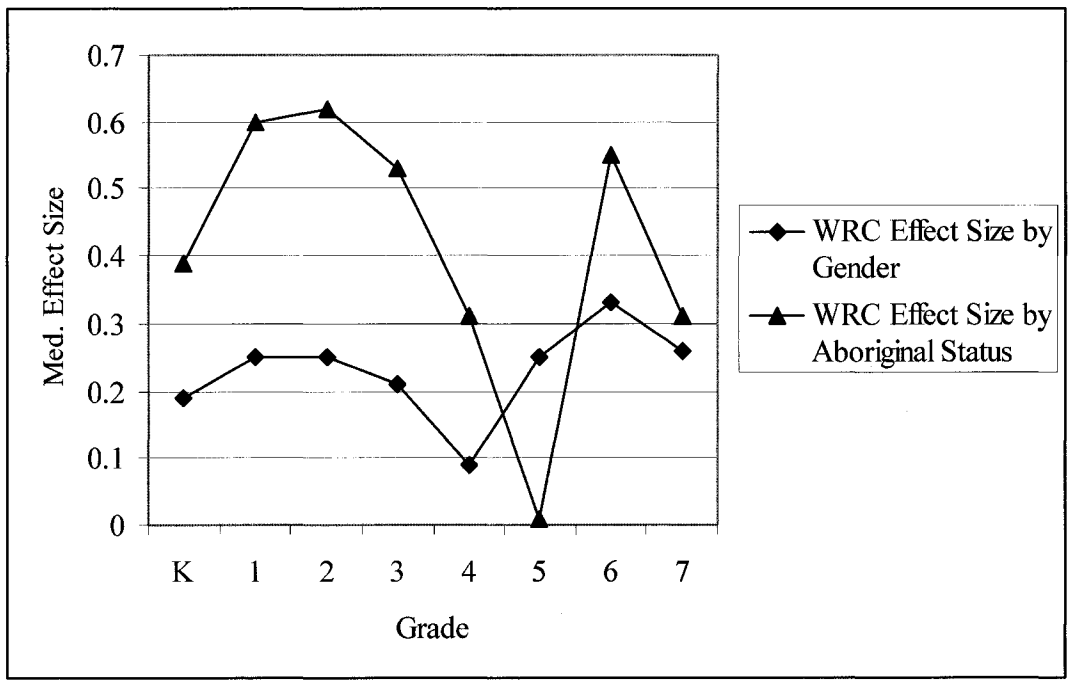

Figure 8. Line graph of median WRC effect sizes by gender and aboriginal status.

When comparing the median effect sizes by both gender and aboriginal status for WSC it is evident that there is a downward trend in effect sizes for the aboriginal results and an upward trend in the effect sizes for the gender results (see Figure 9). These effect

sizes become more apparent on the line graph. The linear regressions and *R*-squared values for each trend line are as follows: $y = 0.0302x + 0.3489$ and $R^2 = 0.4949$ for WSC effect size by gender; and $y = -0.0412x + 0.4579$ and $R^2 = 0.6146$ for WSC effect size by aboriginal status. The slope for the gender results is positive, which again is evidence of an upward trend in effect sizes for gender which confirms females are continuing to pull ahead of males. The slope for aboriginal status is negative which is further evidence of a downward trend in effect sizes for aboriginal status which confirms indications that aboriginal students are getting closer to non-aboriginal students.

The slopes for the WRC results for both the gender and aboriginal results are approximately half of the size of the slopes for WSC for both groups. The slope indicates a rate of change that for the WSC results is roughly twice as fast as for the WRC results. The R^2 values for WSC for gender and aboriginal status are very large, approximately 50% and 62% respectively, as compared to the WRC scores for gender and aboriginal status, approximately 17% and 8% respectively. This indicates that for WSC the data is more tightly centered around the slope line than the data is for WRC.

There are noticeable trends for both the gender and aboriginal effect sizes for WSC. For both gender and aboriginal status the WSC effect sizes are very similar at the Kindergarten and Grade 1 levels (small to medium). By Grade 2 the gender and aboriginal results begin to differ greatly. For gender the effect sizes for WSC begin to climb so that by Grade 6 the effect size is medium. For aboriginal status the effect sizes for WSC begin to fall and by Grade 5 the effect size is trivial. For both groups there is a dip in effect sizes at the Grade 5 level, another anomaly.

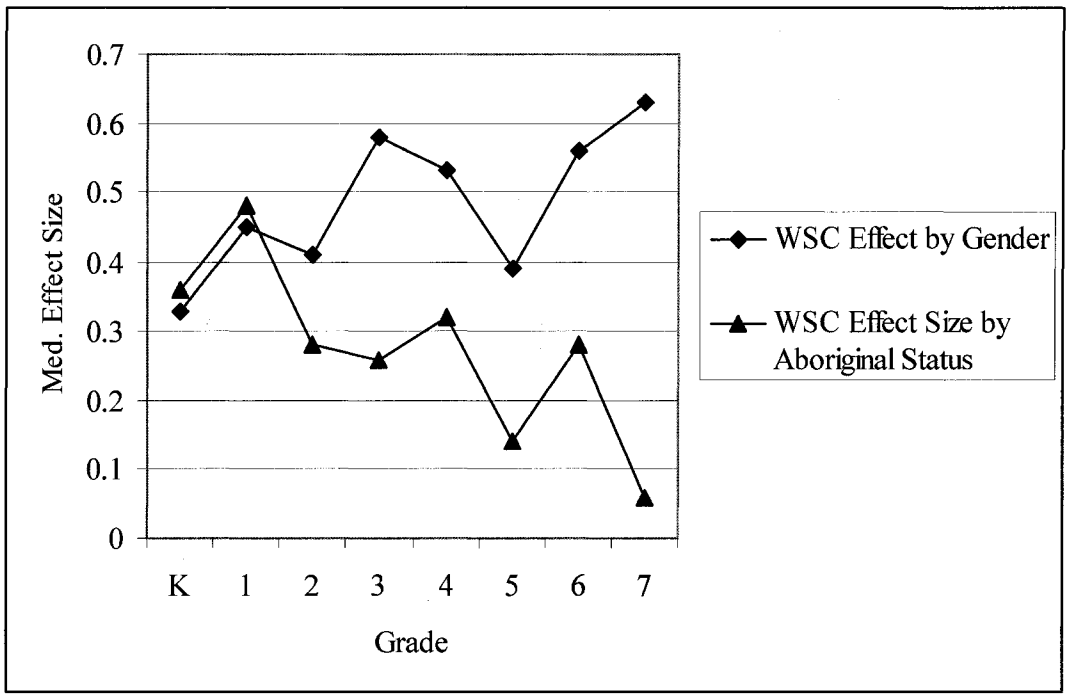

Figure 9. Line graph of median WSC effect sizes by gender and aboriginal status.

The gender gap favouring females is increasing for both WRC and WSC as is evident by the positive slope of the trend lines for the effect size results. This is not encouraging. However the gap favouring non-aboriginal students is decreasing for both WRC and WSC as is evident by the negative slope of the trend lines for the effect size results. This means that aboriginal students' performance is getting closer to non-aboriginal students' performance. In the upper elementary grades the effect size for the aboriginal results is not greater than medium and there is a steady decrease to becoming small. This is encouraging news.

CHAPTER FIVE: DISCUSSION AND CONCLUSIONS

This chapter consists of four sections. The first section will summarize the results of the study. The second section will discuss limitations of the study. The third and fourth sections will discuss implications for practice and implications for theory.

Summary and Conclusions

In the study, the question of whether gender and aboriginal status affect early literacy skills and reading and writing fluency as measured by DIBELS and CBM variables was analyzed. A total of 2420 students from Kindergarten to Grade 7 were tested during the fall, winter, and spring of the school year. The test results were analyzed using a 2x2 analysis of variance (gender by aboriginal status) to estimate the effects of each of the gender and aboriginal status groups and the interaction between gender and aboriginal status. The students' scores were analyzed as to whether there are differences in the DIBELS variables, ISF, LNF, PSF, NWF, and ORF, and in the CBM variables, WRC, WSC, and TWW.

The research found that although female students scored higher in early literacy skills and in reading and writing fluency at every grade level and every testing period, the only consistent statistically significant gender difference is in writing fluency, as measured by WSC and TWW from Grade 1 to 7. There were a few cases where statistically significant gender differences occurred at the Kindergarten to Grade 1 level in PSF, NWF, and ORF but this was not enough evidence to conclude a consistent, overall statistically significant gender difference for early literacy skills across these two grade levels.

The study found no statistically significant differences in early literacy skills when using the DIBELS measures and no statistically significant gender differences for reading fluency when using the CBM measure of WRC. However, the study did find a statistically significant gender difference for writing fluency when using the CBM measures of WSC and TWW. The CBM gender difference results do not completely correspond with an earlier study done by Hedekar (1997) in the same school district. Hedekar's study found a definite gender difference for both reading and writing fluency when using the CBM measures. The difference in results from Hedekar's study to the current study could indicate that the Prince George School District has begun to address the gender differences in reading and writing that Hedekar's study revealed and has been able to reduce the gender differences for reading fluency based on the CBM measures. Since Hedekar's study there is generally a noticeable improvement in the mean scores for both males and females for all the CBM variables with the difference for males being more consistent and of a greater magnitude.

The research also found that non-aboriginal students scored higher than aboriginal students in early literacy skills and in reading and writing fluency at every grade level and testing period except the Grade 5 January testing of WRC, WSC, and TWW. A consistent statistically significant difference for non-aboriginal and aboriginal students was found for early literacy skills using DIBELS measures and for reading fluency using the CBM measure of WRC. Some statistically significant differences for non-aboriginal and aboriginal students were found for writing fluency using the CBM measures of WSC and TWW but there was not enough evidence to conclude that a consistent, overall statistically significant difference for these two groups has occurred for writing fluency.

This study found statistically significant differences between non-aboriginal and aboriginal students in early literacy skills when using the DIBELS measures and in reading fluency when using the CBM measures. This study did not find statistically significant differences between non-aboriginal and aboriginal students in writing fluency when using the CBM measures.

There were no statistically significant interactions between gender and aboriginal status when early literacy skills, reading fluency, and writing fluency were measured. The fact that a student is both male and aboriginal does not put him at a greater risk for reading and writing difficulties than what would be expected from considering each factor separately.

Limitations of this Study

One limitation of this study is the meaning of the variable aboriginal status. The first problem is that if students are self-identifying, a question arises as to what criteria they are using. This could mean that some students are identifying themselves as aboriginal based on one set of criteria and other students could be identifying themselves as aboriginal based on a different set of criteria. The second problem that arises is that this study does not define what aspects of aboriginal status are contributing to the DIBELS and CBM measure results. For example, the student's socio-economic status, the education level of the parents, and rural versus urban living could be other aspects of aboriginal status that contribute to the results but were not measured or addressed in this study. Factors such as the lack of culturally relevant material and pedagogy, the current political climate, differing learning styles, and early literacy exposure prior to entering school could also be contributing to the differences found between non-aboriginal and

aboriginal students. Curwen Doige ((2001) and Dunn (2001) suggest that these variables may be important factors to consider in literacy development for aboriginal students, yet the effect of these variables in this study is completely unknown.

Another limitation to this study is the use of the DIBELS measure. Using this measure does not tell us the reading and writing fluency of students in Kindergarten and Grade 1 as does the CBM measure. The DIBELS measure only indicates a level of risk of students not being successful in acquiring reading and writing skills at grade appropriate times. The inherent problem of Kindergarten and Grade 1 students not being able to read and write makes it difficult to make the same comparisons between students in these two grades and students in Grade 2 to 7.

Implications for Further Research

Although the only statistically significant gender difference in this study occurred in writing fluency it is also important to consider the fact that girls still outscored boys for all measures when discussing further research. The question of why this gender gap occurs and the identification of factors contributing to this gap are topics for additional research. There is a need to investigate whether or not certain pedagogy and curriculum contributes to the gap between males and females and if so what are the reasons for these pedagogical and curricular differences; are they cultural, political, social, or economic. As part of this research we need to find new teaching strategies and curriculum to address this gender difference. The possible outcomes of implementing same sex classrooms would be an example of an area that needs to be researched further as well as the issue of bias in construction of test items.

Another issue that needs to be investigated is the gender gap at higher grades and how it compares with the gender gaps in this study. More specifically is this gap widening or narrowing as students get into the secondary grades and why is it narrowing or widening. Related to this issue is the question of why, when there were no statistically significant gender differences according to the DIBELS results for Kindergarten and Grade 1, do significant gender differences develop for writing for late grade 1 through to Grade 7. Further research is needed to investigate what is occurring that is contributing to this development.

With respect to aboriginal differences in performance, further research needs to be done to address what aspects of aboriginal status are contributing to these differences. Are the factors contributing to these differences social, cultural, political, or economic? Investigations into pedagogical strategies and curriculum that may be contributing to the gap between aboriginal and non-aboriginal students need to be done. For example research is needed into the area of learning styles of aboriginal versus non-aboriginal students and how and why the dominant forms of pedagogical strategies may be contributing to the gap between aboriginal and non-aboriginal students. In addition research needs to be done on which teaching strategies and curriculum would be successful in reducing the gap between aboriginal and non-aboriginal students.

Another issue that needs to be investigated is the gap between aboriginal and non-aboriginal students at the higher grade levels and how it compares with the aboriginal status gaps in this study. The question of whether or not the gap is widening or narrowing as students move into the secondary grades and why it is narrowing or widening needs to be researched. Due to the statistically significant differences in early

literacy skills at the Kindergarten to Grade 1 level another area of investigation would be to examine the possibility and benefit of implementing early intervention or early literacy programs for students prior to enrollment in Kindergarten.

Implications for Practice

One of the findings in this study is that although the only statistically significant gender difference was in writing fluency from Grade 1 to 7, females are still consistently outscoring males. This study partially replicates what has been concluded in other studies by Hedekar (1997) and by Gambell and Hunter (2000) as well as the FSA results. Educators will need to find ways to address these differences and to find interventions that will enable male students to equal the performance of female students. Some possible ways of doing this could include correcting any possible bias in classroom methodology, curriculum or test item construction. Educators also need to be cautioned about how they use the information from the CBM and DIBELS results. The intention is to use it for assessment and intervention not for labeling and funding purposes, although CBM is used for these purposes as well.

The other finding of this study is that statistically significant aboriginal differences were evident for early literacy skills and for reading fluency. In addition non-aboriginal students consistently outscore aboriginal students in early literacy skills as well as reading and writing fluency with the exception of the Grade 5 January testing of the CBM measures. As with gender differences, educators will need to find ways of addressing these differences and finding interventions that will enable aboriginal students to perform at the same level as their non-aboriginal counterparts. Some researchers or theorists suggest using more culturally relevant curriculum materials and pedagogical

strategies (Curwen Doige, 2001; Dunn 2001). It is also necessary for educators to use caution with this information so that there is not the temptation to fix the problem without knowing what the problem is.

REFERENCES

Allred, R. A. (1990). Gender differences in spelling achievement in grades 1 through 6. *Journal of Educational Research, 83,* 187-193.

British Columbia Ministry of Education. (2001). *Interpreting and communicating British Columbia Foundation Skills Assessment results* [Electronic version]. Victoria, BC: British Columbia Ministry of Educaton.

British Columbia Ministry of Education. (2002). *Performance plan 2001/2002 – technical notes.* Retrieved May 23, 2004, from http://www.bced.gov.bc.ca/abed/perf2002_notes.htm

British Columbia Ministry of Education. (2003a). *British Columbia Foundation Skills Assessment 2003 provincial results report* [Electronic version]. Victoria, BC: British Columbia Ministry of Education.

British Columbia Ministry of Education. (2003b). *British Columbia Foundation Skills Assessment 2003 district results report* [Electronic version]. Victoria, BC: British Columbia Ministry of Education.

British Columbia Ministry of Eduction. (2004). *School completion rates increase.* Retrieved May 23, 2004, from http://www2.news.gov.bc.ca/nrm_news_releases/2004OTP0004-000036-Attachment1.htm

Curwen Doige, L. A. (2001). Literacy in aboriginal education: an historical perspective. *Canadian Journal of Native Education, 25,* 117-128.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. (1992). The nature and development of curriculum-based measurement [Electronic version]. *Preventing School Failure, 36,* 5-10.

Deno, S. L. (2003). Developments in curriculum-based measurement [Electronic version]. *The Journal of Special Education, 37,* 184-192.

Dunn, M. (2001). Aboriginal literacy: Reading the tracks [Electronic version]. *Reading Teacher, 54,* 678-687.

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the dynamic indicators of basic early literacy skills – modified [Electronic version]. *School Psychology Review, 30,* 33-49.

Fewster, S., Fortier, P., Foulds, B. J., MacMillan, P. D., Struthers, L., & Walraven, G.(2003). *Curriculum-based measurement and dynamic indicators of basic early literacy skills in School District No. 57 guidebook.* Prince George, BC: School District 57 (Prince George).

Fewster, S., & MacMillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing [Electronic version]. *Remedial and Special Education, 23,* 149-156.

Foulds, B. A. (2002). *Gender and relative-age differences in math fluency using curriculum-based measurement.* Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia, Canada.

Fuchs, L. S., & Fuchs, D. (1991). Curriculum-based measurements [Electronic version]. *Preventing School Failure, 35,* 6-11.

Gaikezehongai, S. (2003). Aboriginal literacy: Raising standards, blazing trails [Electronic version]. *Canadian Journal of Native Education, 27,* 41-44.

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

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing [Electonic version]. *School Psychology Review, 31,* 477-497.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology.* Boston: Allyn and Bacon.

Good III, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York, NY: The Guilford Press.

Good III, R. H., Kaminski, R. A., Smith, S. B., Simmons, D. C., Kameenui, E., & Wallin J. (2003). Reviewing outcomes: using DIBELS to evaluate kindergarten curricula and interventions. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: systems for the observation of teaching and learning (*pp. 221-266). Baltimore, MD: Paul H. Brooks Publishing Co.

Good, R. H., Wallin, J. U., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). *System-wide percentile ranks for DIBELS benchmark assessment (Technical Report 9)* [Electronic version]. Eugene, OR: University of Oregon.

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgements of reading comprehension and oral reading skills. *School Psychology Review, 32,* 228-240.

Hedekar, L. (1997). *The effects of month of birth and gender on elementary reading and writing fluency scores using curriculum based measurement.* Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia, Canada.

Pomplun, M., Sundbye, N., & Kelley, J. (1999). Exploring gender differences in answers to constructed-response reading items [Electronic version]. *Scientific Studies of Reading, 3,* 55-67.

School District No. 57, Prince George (2003). *Focus on student learning: District plan for student success* [Electronic version]. Prince George, BC: School District 57 (Prince George).

Scott, V. G., & Weishaar, M. K. (2003). Curriculum-based measurement for reading Progress [Electronic version]. *Intervention in School & Clinic, 38,* 153-159.

University of Oregon (n.d.a.). *Official DIBELS homepage.* Retrieved February 18, 2004, from http://dibels.uoregon.edu/index.php

University of Oregon (n.d.b.). *Dynamic indicators of basic early literacy skills.* Retrieved March 13, 2004, from http://dibels.uoregon.edu/dibels_what.php

Ward, A., Shook, S. E., & Marrion, L. V. (1993). Concepts about writing: Native children in a cross-cultural setting. *Canadian Journal of Native Education, 20,* 96-108.

Appendix

Approval Forms