THE PSYCHOMETRIC PROPERTIES OF THE DEVELOPMENTAL READING ASSESSMENT

By

Rachel Stewart

A thesis submitted to the Graduate Program in Education
in conformity with the requirements for the
Degree of Master of Education

Queen’s University
Kingston, Ontario, Canada
August, 2016

Copyright © Rachel Stewart, 2016
ABSTRACT

The purpose of this study was to examine the technical adequacy of the Developmental Reading Assessment (Beaver & Carter, 2004). Internal consistency analysis, factor analysis, and linear regression analyses were used to test whether the DRA is a statistically reliable measuring of reading comprehension for Grades 7 and 8 students. Correlational analyses, decision consistency analyses, and a focus group of experienced Intermediate (Grades 7 and 8) teachers examined whether there is evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension. Results indicated that, as currently scored, internal consistency is low and skewness of distribution is high. Factor analyses did not replicate those cited by the DRA developers to prove construct validity. Two-way contingency analyses determined that decision consistency did not vary greatly between the DRA, EQAO, scores and report card marks. Views expressed during the focus group echoed many of the challenges to validity found in the statistical analysis. The teachers found that the DRA was somewhat useful, as there were limited alternative reading assessments available for the classroom, but did not endorse it strongly. The study found little evidence that the DRA provides valid interpretations regarding Intermediate students’ reading skills. Indicated changes to the structure and administration procedures of the DRA may ameliorate some of these issues.
Acknowledgements

I wish to acknowledge those individuals who have offered their support and expertise throughout this process. I would like to thank Dr. Don A. Klinger, my thesis supervisor. His unwavering support and understanding, along with his enthusiasm for statistics, have been instrumental in my growth as a researcher. Dr. Elizabeth Lee, who acted on my thesis committee, has been a source of inspiration.

I would also like to acknowledge the students and staff of the participating school. Without their cooperation, this study would not have been possible.

Finally, to my family, James, Owen, Aidan, and Iain. Your support and patience made this entire journey possible.
# TABLE OF CONTENTS

Abstract ..................................................................................................................................................i

Acknowledgements ............................................................................................................................ii

Table of Contents ...............................................................................................................................iii

List of Tables ........................................................................................................................................iv

Chapter One: Introduction ................................................................................................................1

  *Purpose........................................................................................................................................4
  *Significance.................................................................................................................................9

Chapter Two: Literature Review ......................................................................................................10

  *Current Trends in Assessment .....................................................................................................10
  *Components of Reading Comprehension ..................................................................................13
  *Adolescent Readers ...................................................................................................................14
  *Reading Measures ....................................................................................................................15
  *The Developmental Reading Assessment ................................................................................17

Chapter Three: Methodology ..........................................................................................................26

  *Quantitative Study ....................................................................................................................27
    *Participants .............................................................................................................................27
    *Instrumentation .......................................................................................................................28
    *Data Collection .......................................................................................................................29
    *Analysis ....................................................................................................................................30
  *Qualitative Study .......................................................................................................................32
    *Participants .............................................................................................................................32
    *Instrumentation .......................................................................................................................33
    *Data Collection .......................................................................................................................34
    *Analysis ....................................................................................................................................35

Chapter Four: Results .....................................................................................................................37

  *Psychometrics Within the DRA .................................................................................................38
    *Descriptive Statistics and Correlation Analyses ..................................................................38
    *Internal Consistency Analysis ...............................................................................................39
    *Factor Analysis .......................................................................................................................41
LIST OF TABLES

Table 1: Correlations Between DRA Items Including DRA Text Level and Survey ........39
Table 2: DRA Section Composition by Subtask ...........................................................41
Table 3: Factor Loadings of All DRA Components .......................................................42
Table 4: Factor Analysis of Oral and Written Items from the DRA ...............................44
Table 5: Regression Coefficients of DRA Components ................................................46
Table 6: Descriptive Statistics for the DRA, EQAO Results and Report Card Marks ........48
Table 7: Correlations Between DRA Results, Report Card Marks and EQAO Reading Scores .................................................................................................................................49
Table 8: At Grade Level Decision Consistency For EQAO Results and Report Card Marks ........................................................................................................................................50
Table 9: At Grade Level Decision Consistency for EQAO Results and DRA Scores ........51
Table 10: At Grade Level Decision Consistency for DRA Scores and Report Card Marks..51
CHAPTER 1
INTRODUCTION

The goal of any education system is to provide quality instruction to students, enabling them to reach their potential and to succeed. Changes in societal expectations have created an emphasis on accountability for teachers, principals, and school boards. Here in Ontario, the Ministry of Education responded to these changes with the creation of the Education Quality and Accountability Office (EQAO). EQAO’s mandate is to administer and report the results of province-wide, national (e.g., Pan Canadian Assessment Program), and international (e.g., Programme for International Student Assessment) tests to school board officials, parents, and the public. This agency “acts as a catalyst for increasing the success of Ontario students by measuring their achievement in reading, writing and mathematics ... resulting data provide a gauge of quality and accountability in the Ontario education system” (Calman & Crawford, 2013, p. 2). Ontario is not unique in its use of large-scale assessments (LSAs). Other provinces in Canada use such testing programs (Klinger, DeLuca, & Miller, 2008), as do the vast majority of American states. The No Child Left Behind legislation (2002) and the more recent Race to the Top initiative (2009) have had profound effects on the student testing programs in the United States.

As a result of such ongoing and increasing LSA use, researchers have renewed their examination of the development and impact of LSAs. Perhaps, not surprisingly, researchers previously found that LSAs have limitations in their ability to monitor achievement: “Most large-scale tests still serve only institutional purposes, are administered to big groups in single sittings on a few dates per year, make little use of new technology, and are premised on a
psychological model that probably owes more to the behaviorism of the first half of the century than to the cognitive science of the current half” (Bennett, 1997 as cited in Simon, Ercican, & Rousseau, 2013, p. 3). While LSAs continue to serve as a tool to monitor student achievement (e.g., Klinger et al., 2008), most jurisdictions recognize it is certainly not the primary source of information regarding student achievement.

Given the increasing accountability expectations, and looking for a more immediate, external, and objective indication of student achievement, some school boards have mandated the use of externally developed teacher administrated assessments. The purposes of these assessments are to inform programming, identify struggling learners for remediation, and to provide a benchmark for comparison to spring assessments. Given its prominence in the learning process, literacy has been the most common focus of these assessments. The intention for the use of these literacy assessments is to provide teachers with a relatively easy to use diagnostic measure that will help inform subsequent teaching. In Growing Success (2010), the term diagnostic assessment is defined as “assessment that is used to identify a student's needs and abilities and the student's readiness to acquire the knowledge and skills outlined in the curriculum expectations” (p.146). This informed teaching will then lead to improved results on the external LSAs used within the accountability frameworks. Not surprisingly, test developers have found a strong market for these diagnostic measures and have aggressively worked to promote their use.

The province of Ontario included the requirement for teachers to administer and use such diagnostic assessments in Policy/Program Memorandum 155 (2013). At my school, this policy was initially interpreted to mean all intermediate teachers administer either the Developmental
Reading Assessment (DRA, Beaver & Carter, 2003) or Comprehension Attitude Strategies Interests Grade 4 - 8 (CASI) by a certain date in the autumn, and the results were sent to the board office to be collated with all other schools. Subsequent interpretations have emphasized teachers’ professional judgement in the selection and timing of these tests. There is a continued and extensive use of these diagnostic literacy (mostly reading) assessments throughout the province. Nevertheless, there is no mandated standard reading comprehension test for this preliminary diagnostic literacy assessment, leaving school boards to develop their own, or to purchase such tests from a publisher. These tests vary in terms of cost, amount of time required to administer, reading levels covered, format, target population, and so on. For valid decisions to be made regarding student programming, school boards must ensure that they are using technically adequate assessments. Unfortunately, not all reading comprehension assessments are technically adequate. Test validity, “the extent to which inferences made on the basis of numerical scores are appropriate, meaningful, and useful” (McMillan & Schumacher, 2010, p. 173), and test reliability, “the extent to which the results are similar over different forms of the same instrument or occasions of data collection” (McMillan & Schumacher, 2010, p. 179), may not have been determined.

One such popular external measure is the DRA (2003, Canadian edition, 2004). Its advertised purposes are to (a) monitor student growth, (b) help with diagnosis of students’ needs and plan for instruction, (c) prepare students for classroom and testing expectations, and (d) keep stakeholders informed about student achievement (Beaver & Carter, 2004). Given the popularity of the DRA, and other similar externally developed assessments, it is critical to examine the extent to which these assessments can deliver on their promised measures. There are reasons to
doubt the quality of the DRA. During its development, the DRA was almost exclusively tested with primary students who read at grade level (DRA2 K - 8 Technical Manual, Beaver & Carter, 2009). The DRA comprises three sections: the student survey, oral reading fluency, and written comprehension, but the DRA Technical Manual reports validation and reliability analysis of just the latter two sections.

**Purpose**

The reading comprehension measure commonly used in my school board is the Developmental Reading Assessment (DRA2, Beaver & Carter, 2004). The DRA claims to:

1. monitor student growth on a variety of crucial skills and strategies that successful readers utilize; 2. help teachers diagnose student needs and plan for timely instruction; 3. prepares students to be successful at meeting today’s classroom and testing expectations, and; 4. support teachers and school districts in keeping parents and other stakeholders informed about the level of student achievement. (Beaver & Carter, 2004, p. 4)

Almost none of the field-testing of this assessment, necessary to collect data for analysis of test reliability and validity, was conducted with intermediate students (DRA2 K-8 Technical Manual, Beaver & Carter 2009). As a Grade 8 teacher, with many years of experience teaching language, I questioned the use of the DRA. This test takes a lot of time to administer and mark, and its administration procedures create a challenge for classroom management. I wanted to be confident in the scores and reading levels that the DRA produces. During informal staffroom discussions of the DRA, my intermediate grade colleagues expressed similar concerns. Hence
the purpose of my research was to examine the technical adequacy of the DRA with intermediate students. My research was guided by the following research questions:

1) Is the DRA a statistically reliable measure of reading comprehension for Grade 7 and 8 students?

2) Is there evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension?

My research examined DRA results from intermediate students, Grades 7 and 8, with a wide range of reading abilities, including children with gifted, learning disabled, and mild intellectual delayed identifications. Through statistical analysis of these results, the DRA’s ability to be a reliable test that provides valid interpretations of reading for this population of students was examined. Analyses of the components of the DRA provided evidence to the extent to which the subtests of the DRA are more strongly correlated with overall reading ability. Further, my analyses attempted to determine if the student survey represents an irrelevant or incidental prerequisite. The predictive validity of the DRA was also examined by correlating DRA scores with previously administered EQAO assessments and subsequent report card marks. By comparing these results, it was determined which measures were better aligned with the others. Data collected from a focus group of educators provided evidence of teachers’ application of the DRA in intermediate classrooms, and whether or not, in their professional opinion, they believe the DRA provides for a valid assessment of student reading.

(context)

According to the Ontario Ministry of Education (2010), “assessment is the process of gathering information that accurately reflects how well a student is achieving the curriculum
expectations in a subject or course. The primary purpose of assessment is to improve student learning” (p. 28). In school boards across Ontario, teachers are mandated to administer and report on reading comprehension achievement in the fall and spring. Decisions regarding reading groups and referral for extra support are often made from these early fall assessments. Timely intervention is also vital for struggling readers. Volante (2010), as cited by the Ontario Ministry of Education states that, “Fairness in assessment and evaluation is grounded in the belief that all students should be able to demonstrate their learning regardless of their socio-economic status, ethnicity, geographic location, learning style, and/or need for special services” (p. 8). As such, teachers must use reading comprehension assessments that not only “look right,” but that are reliable and free from bias. Face validity, or “looking right” is not a sufficient form of validity. Technical validity is based on interpretations made from the assessment. Test developers must identify assumptions made during test development, and justify the inferences made by the test, collecting evidence to support these assumptions. While school board staff are directed by the Ontario Ministry of Education “to ensure that the tools (diagnostic assessments) support the collection of valid and reliable evidence” (Ontario Ministry of Education, 2013, p. 3), finding a technically valid or reliable test can be challenging.

The DRA provides a technical manual containing a description of the procedures undertaken to determine the DRA’s reliability and the extent to which it measures what it is supposed to measure (Beaver & Carter, 2004). The Technical Manual of the DRA (2009) indicates that it was field tested almost exclusively with primary and junior grade students who read at grade level. Invernizzi, Landrum, Howell, and Warley (2005) warn teachers, “not to use
assessments designed for specific groups of students on other, more diverse groups of students” (p. 613).

Adolescence, the time between puberty and approximately 18 years, is a time of big changes as children mature into young adults. Physical and psychological differences, especially changes in motivation and engagement, make adolescents different from primary and junior age children. Intermediate students, those in Grades 7 and 8, are adolescents and come to school with a perspective and experience very different from when they first started school (Irvin et al, 2007).

I investigated DRA use with intermediate students in the school at which I teach Grade 8. The school is an exclusively intermediate school that houses four separate programs: (a) a program for students with learning difficulties; (b) an academic program for students with aptitude or gifted identifications; (c) a regular program for students from neighbouring elementary schools; and (d) an arts program for students with aptitude in the arts. The only internal consistency reliability study conducted by the DRA developers was done in 2006 and used fewer than 200 intermediate students, of which 10 students were from Grade 8 (DRA2 K-8 Technical Manual, Beaver & Carter, 2009).

Unfortunately, no information regarding sample demographic characteristics are provided in the DRA Technical Manual. My research included 83 intermediate students, split fairly evenly between Grade 7 and Grade 8. The sample involved students with a wide range of reading abilities, including students with identified language impairments and those identified as gifted. Unlike previous research, the size of this sample was sufficient to conduct a factor analysis of DRA results with intermediate students.
A significant line of inquiry in Ontario has examined the relationship between EQAO scores and report card marks. Large-scale assessments, such as those administered by the EQAO, offer a snapshot of student achievement. They generally use only one format and modality. Report card marks are generated from assessment data obtained through a much larger variety of formats and over a much greater time frame. These differences can have an effect on the conclusions made regarding student achievement. Ross and Gray (2008) found that for Grade 3 and Grade 6 students, report card marks were consistently higher than EQAO results, with the opposite being true for Grade 9 students. They examined the alignment of these scores and possible explanations for the results. Correlation coefficients between EQAO and report card marks ranged from .32 to .59 (Ross & Gray, 2008).

More recently, Ross and Kostuch (2011) confirmed that in the elementary panel, teachers assigned higher marks than EQAO test results. They examined the stability of this relationship across gender, Public versus Catholic school districts, and English versus French languages. These relationships were deemed robust, with agreement between measures higher for Grade 6 students than for Grade 3 students. They concluded that report cards, “provided information about students’ future achievement that was accurate and delivered up to 2 years prior to the administration of external assessments” (p. 159). My study investigated the relationship between EQAO and report cards in the opposite direction. Do EQAO results provide information about students’ future achievement up to two years hence, as measured by report cards and DRA results?
Significance

Reading comprehension assessments are mandated in many jurisdictions. The DRA is commonly used across many grade levels and reading abilities. Given the large amount of time required to administer the DRA, it is vital that teachers have confidence in its assessment of students’ reading comprehension. Unlike the DRA’s technical manual, my research reported on the DRA’s effectiveness in the intermediate grades. Additionally, students with Special Education identifications were included in the study. If specific subtasks of the DRA are found to correlate more strongly than other subtasks, teachers could use only these subtasks, thereby reducing the class time needed to administer the DRA. Comparing EQAO results to report card marks continues the line of inquiry of researchers such as Ross and Kostuch (2011) and Ross and Gray (2008). Including DRA scores, and using Grades 7 and 8 students, will add to the body of knowledge regarding alignment of grades and assessments.

Organization of the Study

This research study is presented in five chapters. Chapter 1 includes the background of the study, statement of the research questions, purpose of the study, and significance of the study. Chapter 2 presents a review of the literature, which includes reading comprehension and reading measures. Chapter 3 describes the methodology used for this study. It includes the selection of participants, instrumentation, data collection, and data analysis. Chapter 4 presents the results of the quantitative data analysis of student results and the qualitative analysis of focus group data. Chapter 5 provides a summary of the study, discussion of the findings, implications for teaching practice, recommendations for further research, and conclusions.
CHAPTER 2
LITERATURE REVIEW

This chapter presents the rationale for conducting research on the statistical reliability of the DRA as a measure of reading comprehension for Grades 7 and 8 students; and whether there is evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension. The following review of literature represents the literature pertinent to my research study. Specifically, Chapter 2 is organized into four sections: a) assessment, b) reading comprehension, c) adolescents, and d) current reading measures.

Assessment

Assessment and evaluation are often used interchangeably to indicate information gathering about student achievement, but there is a distinction between them. While assessment is the process of obtaining information about students, evaluation is “the process of making a value judgement about the worth of a student’s product or performance” (Brookhart & Nitko, 2008, p. 5). The Ontario Ministry of Education defines assessment as the “process of gathering, from a variety of sources, information that accurately reflects how well a student is achieving the curriculum expectations in a subject or course” (Growing Success, 2010, p. 143). The drive to demonstrate school effectiveness and to standardize assessment, evaluation, and reporting across Ontario prompted the publication of the Growing Success policy document. Central to the document is the division of assessment for, of, and as learning.

Assessment for learning can be both diagnostic and formative (Growing Success, 2010). It is expected to provide students with feedback and coaching. It is also used to guide
programming decisions by teachers, and to allow students to monitor their own progress.

Formative assessment involves both formal and informal measures, with the purpose of improving student motivation and learning. (McMillan et al., 2011). Assessment as learning occurs with teacher and peer input, for students to track their own progress. Assessment of learning is more formal and occurs at the end of a unit or course (McMillan et al., 2011). It is a summation of student learning that is often used to make important decisions about academic futures. There is an assessment and evaluation process in which assessment is used to obtain data regarding students’ learning and achievement and evaluation then uses the data to make a judgement regarding student achievement relative to expectations and standards. In Ontario, student work is now assessed in relation to criterion-referenced standards, as opposed to norm-referenced standards. The Ministry of Education points out that norm-referenced assessments and evaluation were not always fair, and that, “The goal of using a criterion-based approach is to make the assessment and evaluation of student achievement as fair, reliable, and transparent as possible” (Ontario Ministry of Education, 2010, p. 19). The Language Curriculum for Grades 1 - 8 (2006) lists 11 strategies teachers must use, “to ensure that assessment and evaluation are valid and reliable” (p. 15). These strategies include concepts of variety of forms, appropriateness for learning activity and teacher purpose, fairness to all students, accommodations based on Special Needs, student feedback, goal setting, and access to sample work. The Ontario Ministry of Education has set four levels of achievement to be used for evaluation and reporting: Level 1 work falls much below the provincial standard; Level 2 approaches the standard; Level 3 meets the standard; and Level 4 surpasses the provincial standard. These evaluation levels are converted into letter grades on the provincial report cards for Grades 1 through 6. From Grade 7
upwards, these levels are converted into percentages. Appendix 1 displays the Achievement Chart from the 2006 Ontario Ministry of Education Language Document. It represents categories of knowledge and skills, which remain constant from Grades 1 - 8. The chart is used by teachers to evaluate achievement across all subject areas. The overall expectations are broad themes which vary little from grade to grade and subject to subject. They include: Knowledge of content; Use of planning skills; Use of conventions; and Transfer of knowledge and skills. There are numerous grade and subject specific expectations, and teachers “will use their professional judgement to determine which specific expectations should be used to evaluate achievement” (Ontario Ministry of Education, 2006, p. 16).

Education in Canada is administered on a provincial and territorial level. The Royal Commission on Learning (1994) observed that although Ontario students participated in Large Scale Assessments (LSA), there was little publicly available information about the data, as compared to other jurisdictions. The Commission recommended:

That the Ministry develop detailed, multi-year plans for large-scale assessments (program reviews, examination monitoring), which establish the data to be collected and the way implementation will be monitored, and report the results publicly, and provide for the interpretation and use of results to educators and to the public (The Royal Commission on Learning, 1994, Chapter 11 # 54)

Consequently, the Office of Learning Assessment and Accountability, now called the Education Quality and Accountability Office (EQAO) was founded.

EQAO’s mandate is to provide information for accountability to students, parents, educators, and the public. Considerable effort has been made by EQAO to develop assessment
measures that provide technically sound data on student achievement in language and mathematics. EQAO language and mathematics assessments are administered in all publicly funded Ontario elementary schools in Grades 3 and 6, Grade 9 students write mathematics assessments, and Grade 10 students write the Ontario Student Literacy Test (OSSLT). A passing mark is required on the OSSLT for graduation.

EQAO recognizes that there are several sources for error in their measures. A comprehensive administrative procedure (with serious consequences for noncompliance) accompanies each package of EQAO tests. This action is to mitigate differences in teachers’ administration and therefore ensure fairness. Inter-rater reliability issues surrounding the marking of the tests is addressed by extensive training of scorers and use of exemplars and rubrics. Appendix 2 provides examples from the 2012 EQAO Grade 6 test for Reading.

An extensive investigation of the technical properties of the EQAO measures has been published in the EQAO Research Series, available publicly at www.eqao.com. Included are a psychometric analysis of the generalizability of the Grade 3 and Grade 9 measures (EQAO Research Series #3, 1999), and a comparison of student performance on the Grade 10 OSSLT and the PISA assessment through the lens of Item Response Theory.

**Reading Comprehension**

Reading comprehension is a complex process and not directly measurable. Shuy et al. (2006) and Appelman (2010) assert that reading comprehension is made up of various component skills, strategies, and background information. In their article examining the *Simple View of Reading* (SVR), Kirby and Savage (2008) expand on this by proposing that reading comprehension is the product of listening comprehension and decoding. Decoding is also a
complex process involving components such as phonological, orthographic and semantic awareness, word meaning, and morphology (Kirby & Savage, 2008). SVR is an explanation of why students with good decoding skills can comprehend poorly, and why students with good comprehension can have difficulty decoding. The Ontario Curriculum for Language (2006) concurs that reading is a complex process. The Ministry document focuses on the strategies students must use before, during, and after reading, such as activating prior knowledge and drawing inferences. In Ontario, Grades 7 and 8 students with good reading comprehension skills are expected to read and demonstrate an understanding of a variety of texts, to recognize a variety of text forms, to read fluently, and to reflect on their reading ability. Appleman (2010) lists nine metacognitive strategies that expert readers understand. Poor readers lack these metacognition strategies, which include visualizing, monitoring, revising meaning, and evaluating (Appleman, 2010). Irvin et al. (2007) concur and include, “Plan for the demands of different kinds of text” as a strategy used by good readers (p. 59).

Adolescent Readers

Adolescents, approximately from ages 10 to 20, are different than primary and junior aged children (Lerner et al., 2005). Aside from the obvious physical maturation, adolescents’ cognitive, emotional, and social development differentiates them from younger children. Physically, puberty causes: rapid growth, resulting in considerable increases in height and weight; the development of primary and secondary sex characteristics; changes in the quantity and distribution of body fat and muscle; and changes to cardiovascular and muscular endurance and strength (Steinberg, 2008). Adolescents’ brains are also growing. They are developing: improved abstract thinking skills; metacognition skills; the ability to think about more than a
single issue or perspective; and the ability to see issues are relative, rather than black and white. (Steinberg, 2008). Adolescents often feel emotions intensely but their ability to self-regulate and for decision making lags behind (LiteracyGains, 2012). Social development and sense of identity are another hallmark of adolescence. Students are developing: a personal identity; a sense of independence; and strong relationships outside of the family (LiteracyGains, 2012).

These developmental changes have educational implications. Students may: (a) feel hungry or tired at different times of the school day; (b) question the relevance of their learning; (c) have an increased ability to make connections; (d) want more choice or say about their learning; (e) feel the need to be independent; (f) fear rejection, failure, or inadequacy; and (g) feel self-conscious and lack self-confidence (LiteracyGains, 2012). Motivation and engagement with learning becomes “critical” for adolescent readers (Irvin et al., 2007; Lerner et al., 2005.) “If they are not engaged, adolescents with strong literacy skills may choose not to read or write. The number of students who are not engaged with or motivated by school learning grows at every grade level, reaching epidemic proportions in high school” (Appleman, 2010, p. xiv). Adolescents need choice, autonomy, purpose to keep them involved in literacy tasks, including reading assessments (Irvin et al., 2007). Adolescents are not just older children, and the changes that come with puberty mean it is a time of adaptation.

**Reading Measures**

Reading comprehension measures have changed over time. Fuchs et al. (2001) pointed to the decrease in the use of oral fluency as a proxy by looking at tests from 1929 to 1999. If reading comprehension tests have changed so much, do they measure the same things, and can they be interchanged? Nation and Snowling’s (1997) work was among the first to examine this
issue. They compared two commonly used instruments from the United Kingdom, the *Neale Analysis of Reading Ability - Revised* (Neale, 1989) and the *Suffolk Reading Scale* (Hagley, 1987) by correlating student scores from these two tests, with scores on several other measures. They found that different tests weighted the various sub-skills of reading, such as word recognition, in different amounts. Following Nation and Snowling’s work was the much cited work by Cutting and Scarborough (2006) examining measures commonly used in North America: the *Wechsler Individual Achievement Test* (WIAT; Wechsler, 1992); the *Gates-MacGinitie Reading Test - Revised* (G-M; MacGinitie, MacGinitie, Maria, & Dreyer, 2000); and the *Grey Oral Reading Test - Third Edition* (GORT-3; Wiederholt & Bryant, 1992). Again, it was found that these reading comprehension tests did not seem to measure the same cognitive processes.

More recent work by Keenan, Bjerman, and Olson (2008) confirms Cutting and Scarborough’s (2006) findings. Additionally, they found that the same tests vary in sensitivity across different ages and across different skill levels. Keenan et al. (2008) compared the reading comprehension subtests of the GORT-3, the *Qualitative Reading Inventory* (QRI), the *Woodcock-Johnson Passage Comprehension subtest* (WJPC), and the *Peabody Individual Achievement Test* (PIAT). These four tests were chosen as they are widely used in North America and they represent a wide variety of test formats. Using a factor analysis, it was found that two factors emerged - listening comprehension and decoding. Reading scores on the four measures varied, depending on how heavily the test weighted the two factors. A subsequent regression analysis found that age contributed unique variance for all four tests, but especially for the PIAT and WJPC. Keenan et al. (2008) stated that the younger one is, or the weaker a reader one is, the
more the PIAT and WJPC is a measure of decoding ability rather than listening comprehension. For older and more advanced readers, the PIAT and WJPC are a test of listening comprehension skills.

Keenan and Betjemann (2006) previously found that the GORT was not an assessment of reading comprehension as much as it was a measure of the ability to reason from prior knowledge. They found that the GORT questions could be answered with better than 50% accuracy without reading the text passage at all. Children with little decoding skill could do as well on the GORT as those with excellent skills. The best performance indicator for the GORT was not how accurately the test subjects read, but how easily the question could be answered without reading the text passages. The authors of this study concluded by hoping that researchers and clinicians would critically evaluate the reading comprehension tests now available on the market.

The DRA

In its document, Diagnostic Assessment Toolkit (2013), my school board lists reading comprehension assessments available for use within the board. The DRA is the measure most often cited for use with intermediate students. Previously, many intermediate teachers used the Comprehension Attitude Strategies Interests (CASI) Grades 4 - 8 because of its ease of administration. Students read a text and then wrote answers to the provided questions. The CASI required less class time to administer and posed fewer classroom management challenges than the DRA. The DRA, with its oral reading fluency measure, is considered a more informative measure, and it has been found to have stronger technical qualities (Klinger, Glickman, Wade-Woolley, & Clegg, 2005a, 2005b). As with all assessment, the purpose of using
the DRA in my school board is to achieve “...the three core priorities of the Ministry of Education: high levels of student achievement, reduced gaps in student achievement, and increased public confidence in publicly funded education” (Ontario Ministry of Education, 2013, p. 1).

The DRA and DRA2 are reading assessments geared towards children in Kindergarten to Grade 3, and from Grades 4 to 8. The Canadian Edition of the DRA2 is almost identical in content to the original. It is comprised of three sections: a student Engagement survey; an Oral Reading Fluency component; and a written Comprehension Skills / Strategies assessment.

The first step in administering the DRA is the survey. This one page written task asks students to list what they have read recently, their perceived strengths as readers, reading goals, and plans to achieve these goals. The DRA developers (2004) indicate that this survey may be administered to the whole class at once or in smaller groups, and should take approximately 15 minutes to complete. This survey is divided into two components, Wide Reading, and Self-Assessment and Goal-Setting. These components, as well as all other components from the Oral Fluency and Comprehension sections, are scored on a scale from 1 - 4. Level 1 is called the Intervention level, Level 2 is Instructional, Level 3 is Independent, and Level 4 is labelled Advanced. Wide Reading is assessed by counting the number of different genres listed by the students and whether the teacher feels that these titles are below, at, or above grade level. The DRA Technical Manual (2009) provides a list at each text level (40, 50, 60, 70, and 80) of approximately 12 book titles which the developers feel correspond with Grades 4, 5, 6, 7, and 8. Of note is that chapter books and comics rate a Level 1 score, and that the suggested titles are almost exclusively fiction, not reflecting the, “wide variety of titles across 3 or more
genres” (Beaver & Carter, 2004, p. 61) required to score a Level 4. Self-Assessment and Goal Setting is scored on the number of reading strengths listed by the student, the specificity of the goals, and the complexity of the reading improvement plan.

The second step in the DRA is the individual Oral Fluency assessment. Using a text at the level determined by the teacher, the student reads aloud a designated passage while the teacher takes a running record. This individual conference takes approximately 10 minutes, during which the teacher evaluates the student’s: Expression, determined by emphasis of words and phrases; and Phrasing, determined by the phrase length. The oral reading is timed so that the teacher can calculate the student’s word reading rate which is then converted to a score from 1 - 4. The Accuracy rate is calculated by examining the miscues from the running record. The number of non-self-corrected miscues are tallied and converted to a percentage from which a score from 1 - 4 is assigned. Further analysis categorizes the types of miscues, whether they interfered with meaning, and strategies used by the student.

The third step in the DRA is the written assessment of Comprehension Skills and Strategies. Before reading the entire text and using what they remember from the oral assessment, students complete the Prediction question. This section is scored on the number and reasonableness of the predictions and questions provided by the students. The Summary section asks the students to summarize the text, including the important ideas and supporting details. The completeness and organization of the writing, as well as the students’ use of their own language is used to score the Summary. The Literal Comprehension question asks students to find directly stated information from the text. This score is based on the amount of information provided and how effectively the students respond to the question. Students are asked to infer
beyond what is stated in the text for the Interpretation question. Students scoring at Level 4 (Advanced) demonstrate, “insightful understanding”, and provide important supporting details. Those students scoring at Level 1 demonstrate, “little or no understanding” (Beaver & Carter, 2004, p. 83). The Reflection section asks students to “think deeply” about an important or significant part of the text. Level 4 responses are “insightful”, and, “successfully identify a significant message, event, or information” (Beaver & Carter, 2004, p. 87). The final written question asks students to explain the strategies they used to understand the text. The students’ Metacognitive Awareness score depends on the explicitness of linking their reading strategies to specific points in the text.

The scores from these three steps are tallied to produce (a) a Reading Engagement Score, (b) an Oral Reading Fluency score, and (c) a Comprehension Skills/ Strategies score. These scores are then added to the Text Level, 40 for Level 40, 50 for Level 50, and so on. This DRA Total Score is then compared to the DRA Stage Table to determine the reading Grade Level.

The Teacher Manual (2004) indicates that most students will need 35 - 45 minutes to complete the written section of the DRA, as well teachers will require 10 - 12 minutes, with practice, per student to calculate oral rates, analyse the running record, mark the survey and comprehension section, and calculate grade level. In my experience, many students require well over an hour to complete the written section and the marking takes closer to 30 minutes per pupil. In a class of 28 it takes days to get through the oral assessments, followed by many hours of marking.

In their article, Towards the Peaceful Coexistence of Test Developers, Policymakers, and Teachers in an Era of Accountability, Invernizzi et al., (2005) note the difficulty in developing
assessments that are not only technically sound, but also useful and feasible in a regular classroom. Eight standards for the evaluation of educational assessments are discussed in this article (a) validity, (b) reliability, (c) test development, (d) fairness in testing, (e) scales, norms, and score comparability, (f) standard administration, scoring and reporting, (g) testing individuals of diverse linguistic backgrounds, and (h) responsibilities of policy makers. In her 2006 review, Rathvon examined if the DRA met these standards. She determined that technical adequacy was lacking. Although face validity was claimed by analysis of teacher surveys, low return rates reduced generalizability. The DRA materials state that participation of a teacher committee gives it content validity. However there is no evidence of education, reading, or assessment experts involvement in the DRA’s development. There was no report of enquiry into item discrimination, item difficulty, or differential item functioning. This is unfortunate as item analysis can find unsuspected problems with specific items, (Hopkins, 1998) and thereby identify items that need to be eliminated. Also missing were comments on test format. As mentioned earlier, Cutting and Scarborough (2006) found that reading comprehension results varied according to test format.

The DRA’s Technical Manual (2009) cites various other analyses conducted to establish the validity of the DRA’s interpretations however lack of bias information was missing. The absence of a standardized method of text selection, coupled with no rationale or empirical data to support this omission, also challenged the DRA’s validity. Criterion-related validity data for the DRA was collected exclusively from primary grade students. As previously mentioned, Keenan et al. (2008) found that test results varied according to age.
The construct validity evidence reported by the DRA was again only gathered from primary grade students. Using Kindergarten to Grade 6 students, the DRA developers claimed concurrent validity by correlating DRA scores with the GORT. Correlations ranged from 0.60 to 0.75, but sample sizes were small, and no rationale for omitting Grade 7 and 8 students was given. As previously cited, Keenan and Betjemann (2006) found that the GORT questions could be answered with better than 50% accuracy without reading the passages.

According to the DRA developers, Beaver and Carter (2003), predictive validity was confirmed by asking teachers to rate their students on a 5-point scale and comparing scores with subsequent DRA results. Correlations for the 188 Grades 1 to 6 students were 0.63 for fluency and 0.60 for comprehension.

The DRA developers used a factor analysis to support claims of construct validity. Using the Spring 2006 field test, which included 0.5% Grade 8 students, the authors state that a 2-factor solution emerged that corresponded with the Fluency and Comprehension sections of the DRA. This pattern does emerge at the lowest reading levels, with two factors accounting for 61.8% of variance for DRA Reading Levels 4 - 12. Closer examination of the data showed that for the other levels: 14 - 16, 16 - 24, 28 - 36, and 40 - 80, three factors had Eigenvalues greater than one. While this third factor accounted for 10% of variance, the authors dismiss it as, “negligible in terms of explaining additionally variability” (Beaver & Carter, 2009, p. 42). Table 14 of the Technical Manual (2009) reports on the factor loadings of many of the DRA reading levels. These levels are not reported on individually, but collected together into differently sized interval groups. It is not indicated how the authors determined the intervals for reading levels reported in this analysis. For example, the Levels 14 - 16 interval covers three reading levels, while the
Levels 40 - 80 interval covers 41 levels. Low participant numbers at these higher Reading Levels could explain why the authors chose to group so many levels together. The Factor Loadings table given as evidence of the Fluency and Comprehension 2-factor solution is missing much critical data. Factor loading values are correlations between each test item and the factors determined by factor extraction, in this case Fluency and Comprehension. Items most highly correlated with Comprehension have high loadings on Factor 1, while those with high values for Factor 2 are more closely correlated with Fluency. It is possible for items to be split onto both factors, or to not load heavily on either factor. Table 14 of the DRA Technical Manual omits loadings for Factors 1 or 2 for Expression and Rate items for Levels 4 - 12. Monitoring, Self-Correction, and Problem-solving data is also missing for Levels 14 - 80. Additionally, the loadings for only one factor is published for: Rate, Levels 14 - 16 and 18 - 24; Accuracy, Levels 14 - 16 and 18 - 24; Previewing, Levels 14 - 16; Retell - Teacher Support, Levels 4 - 12, 14 - 16, 18 - 24; and Reflection, Levels 14 - 16, 18 - 24, 28 - 38. No explanation is given for this omission. The DRA Technical Manual claims construct validity by showing that the DRA’s content is “developmental in nature as evidenced by the large relationship between age and DRA2 performance” (p.45). Cited as evidence of construct validity, Table 15 in the DRA Technical Manual displays age, sample size, and raw Fluency and Comprehension scores. No analysis or explanation of these raw scores was reported. Also missing was a discussion of the age ranges used. In Ontario, students in JK are either four or five years old in the Spring, and those in SK are either five or six years old. These youngest children are excluded from this age/score analysis as the range is set at 6 to 13 years. Similarly, many Grade 8 students will turn 14 by the Spring, but are excluded. After a steep increase in median scores during the primary
grades, progress plateaus as the students become older. Ceiling effects could explain this phenomenon.

The second standard proposed by Invernizzi et al. (2005) is reliability. Several studies were cited in the DRA Technical Manual (2009) as evidence of inter-rater reliability. A study by Williams (1999) used only students from Kindergarten to Grade 3. Weber’s (2000) study also used this age group. A comparison of teacher-teacher and teacher-expert raters by Fisher (2003) videotaped five student performances on the DRA. Only the ratings of the three elementary school students, and not the two middle school students, were reported.

Invernizzi et al. (2005) warned teachers “not to use assessments designed for specific groups of students on other, more diverse groups of students” (p. 613). The DRA, 4 - 8 was field tested with a handful of intermediate students. The DRA developers calculated internal consistency reliability by comparing oral fluency on the DRA with comprehension scores. Cronbach’s Alpha scores ranged from 0.54 to 0.85, with the lower internal consistency values generally found with the oral fluency scores and in the higher reading levels. Passage Equivalency tests were conducted to determine if texts at each level were interchangeable. In general, texts were found to be equivalent, but Grade 8 students were significantly underrepresented. They made up .5% of the Spring 2006 sample, and 0% of the 2007 field test sample. Both Grades 7 and 8 students were excluded from the Spring 2008 Test-Retest Reliability and Concurrent Validity studies.

The DRA 4 - 8 is composed of three sections: the previously mentioned oral fluency and comprehension sections, and a student survey section. The survey’s purpose is to measure student engagement with reading. The survey is scored and used, in part, to ascertain a student’s
reading level. The Technical Manual contains no empirical data, for any grade level, of the survey’s reliability or the validity of decisions based upon it. There is no evidence that the survey was field tested, and there seems to have been no attempt to measure its technical adequacy. As Hopkins points out, “logical relevance is the principal criterion for test validity” (Hopkins, 1998, p. 254). Despite the DRA authors’ assertions that the second edition provided additional psychometric measures of the test, there is still little empirical evidence that the DRA, 4 - 8 is a valid or reliable measure of reading comprehension for intermediate students.

The DRA is a classroom assessment tool and as such it needs to be more than technically sound. McMillan et al. (2011) assert that high quality classroom assessments must meet eight criteria, among them: appropriateness of assessment method; validity; reliability; alignment; and practicality and efficiency.

**Summary**

It is evident that valid and reliable instruments to assess student reading comprehension are increasingly important to teachers to inform programming, to provide accountability for administrators, and to provide feedback for students and parents. By gaining a better understanding of: 1) assessment, especially as implemented in Ontario; 2) the complexity of the reading process; 3) changes associated with adolescence; 4) previously used reading measures; and 5) the psychometric properties of the DRA, all stakeholders can better determine if the DRA represents a statistically reliable measure of reading comprehension for Grades 7 and 8 students, and if there is evidence that the DRA provides valid interpretations regarding that reading comprehension and is a high quality classroom assessment.
CHAPTER 3
RESEARCH METHODS

This chapter presents the methods for conducting research on the reliability of the Developmental Reading Assessment (DRA) as a measure of reading comprehension for Grades 7 and 8 students, and examines if there is evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension. The primary goal of this study was to test the research questions: 1) Is the Developmental Reading Assessment (DRA) a statistically reliable measure of reading comprehension for Grades 7 and 8 students? and 2) Is there evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension? This study employed quantitative and qualitative methods of data collection and analysis. Mixed methods approaches can, in some contexts, generate better understanding of a topic. Greene, Kreider, and Mayer (2011) list five purposes for mixing methods: (a) triangulation enhances validity and credibility of inferences by corroborating of results; (b) complementarity uses the different perspectives of various methods to more comprehensively understand complex situations; (c) development, the results of one method are used to inform the implementation of another method; (d) initiation, divergent results from different methods provide impetus for fresh analysis of phenomena; and (e) expansion, the focus of study can be expanded by using different methods for different constructs.

The methods employed to address the research questions are presented separately in this chapter. This chapter is organized into two sections: (a) quantitative methods, and (b)
qualitative methods. Each section includes the following subsections: i) selection of participants, ii) instrumentation, iii) data collection, and iv) data analysis.

**Quantitative Study**

McMillan and Schumacher (2010) point to the appropriateness of quantitative research to emphasize objectivity in measuring and describing phenomena using numbers, statistics, structure, and control. The desire to examine the stability of relationships between the DRA, report card marks, and EQAO results necessitates a correlational approach.

**Participants**

The target population for this portion of the research was all Grade 7 and Grade 8 students at my school (approximately 375 students). These students were from four different programs: 1) program for academically advanced students; 2) a program for students identified with learning exceptionalities or struggling academically; 3) a program for students with artistic aptitude; and 4) program for the students from the neighbouring community. A few students in specially modified programs, who would not normally be administered the DRA, did not participate in this study. The goal of a large student sample aimed to reduce the sampling error and increase the statistical power of my analyses. Given the structure of the school, there were sampling challenges. For example, the artistic program and the struggling students programs have gender imbalances. The artistic program is predominantly female while the other program’s students are predominantly male. Except for the local neighbourhood program, students are self-selected, come from a much wider geographical area, and are subjected to a vetting process. These three “district” programs have admittance procedures that examine report cards, written assessments, oral presentations, Individualized Education Plans, group work, and artist performances. This
The study provided a unique opportunity to examine the stability of the relationships between the DRA, report card marks, and EQAO results across these subpopulations of students.

The first step in the quantitative data collection process included orally introducing the nature of the study to the Grades 7 and 8 classes at the school. All students were provided with two copies of a bright yellow Letter of Information for Parents and Consent Form. Frequent oral and visual reminders, as well as duplicate forms to replace lost ones, were provided in the hopes of increasing students’ participation. Given the protracted approval process of the school board, only one week was available as a turn-around time to return signed consent forms before the end of the school year. Out of 375 students, 83 returned signed consents, for a return rate of 22%. Data were available for 80 students. Of these 80 students, 7 came from the program for children with academic difficulties, 8 from the academically gifted program, 15 from the neighbourhood program, and 50 from the artistic program. I teach in the artistic program and it is possible that the relationship I have with these classes accounted for their much higher rate of return. Because there were so few participants from the other three programs, accounting for only 38% of the sample, it was decided to collapse all of the programs into one group. Had there been larger numbers from these groups, logical group comparisons could have been made.

**Instruments**

All DRA materials used in the study came from the Canadian Edition of the Developmental Reading Assessment Grades 4 - 8 (2003). They included assessment texts, black line masters of assessment rubrics and student booklets, and a Teacher Guide. There are several versions of the DRA: DRA (2003), DRA2 (2006) the second edition, and the DRA Canadian edition (2004). My school board assures teachers that they are all valid, and that the slight
variations between versions does not affect the accuracy of scoring. (Limestone District School Board, 2013a). Report card marks were collected from the Trillium Web Assessment program used by the school board. EQAO results from these students’ Grade 6 reading assessment were also retrieved from the school board’s database.

**Data Collection**

Due to a recent change in school board policy, not all students were given the DRA in the Fall of 2013. Teachers had been instructed to use their professional judgement in determining which instrument to use in assess reading comprehension. Most classroom teachers administered the DRA as usual. Students first completed a Student Reading Survey, which identified independent reading habits and self-efficacy. Students then read orally from a one-page assessment text, in a one-on-one conference with the teacher. As per DRA procedures, teachers kept a timed running record of errors, omissions, and self-corrections during this subtask. Students then read the entire text silently at their desks, and gave written answers to approximately six questions. Using the rubrics and exemplars provided in the DRA materials, teachers assigned marks for all three sections to determine students’ scores. These scores were added to the DRA text level to produce a DRA total score. This total score was used to assign a reading grade level from Grade 2 to Grade 8.

The regular classroom teachers administered the DRA test to reflect current practices and the marking rubrics were collected from them. To preserve confidentiality, each student was given an identification number to identify data during the statistical analysis. All materials were kept in a locked cabinet at my school. Reading marks from report cards were retrieved from the Trillium Web Assessment software used by the school board. Using the provincial formula
(Ontario Ministry of Education and Training, 2010), numeric grades used on the report cards were transformed into four levels. The target of both the report cards and EQAO was to meet the provincial standard (at least Level 3). Results were coded using the previously mentioned I.D. numbers, and were stored with the DRA results in the locked cabinet. Any identifying data was only be seen by the classroom teachers and myself.

The use of the June 2013 reports and Fall 2013 DRA reading measures had several advantages. The DRA is normally administered in the Fall and is not a duplication of assessment. These DRA results and the June 2013 reports card marks were collected only a few months apart, lessening attrition and maturation threats. The Grade 6 EQAO reading results were also used for several reasons. The EQAO reading assessments are only administered in Grades 3, 6, and 10. The Grade 6 assessment was the closest one, chronologically, to my sample population. The teachers administering the DRA were not the same teachers these students had in Grade 6 during EQAO testing, avoiding confirmatory bias. As per school board request, the compiled list of student names was sent to the research office of the school board for the retrieval of Reading report card marks and EQAO reading level data. DRA reading level data was retrieved from classroom teachers’ files.

Analysis

Statistical analyses of the DRA data, report card marks, and EQAO results used SPSS Version 22 (2013) software. For the first research question: Is the DRA a statistically reliable measure of reading comprehension for Grade 7 and Grade 8 students, as measured by internal consistency, I conducted an internal consistency analysis of the DRA data. Reliability analysis
studies the properties of a test and its components, and provides information regarding the relationship between individual test items. Descriptive statistics and inter-item correlation data indicated score distribution and possible overlap of constructs measured in the DRA. Internal consistency (Cronbach’s alpha) indicated the level of internal consistency for the DRA, based on these average inter-item correlations.

The sample size of 80 students was potentially adequate to conduct a factor analysis of the DRA. Factor analysis examined whether more than one construct was needed to account for the observed pattern of results. Principal components analysis was used to make the initial decision regarding the number of factors underlying the DRA data. The Oblimin Rotation Method with Kaiser Normalization was used given the assumed overlap between reading constructs. The intention was to determine if the three factors, which map onto the three separate sections of the DRA, were observed and which factors accounted for the most variance on overall DRA scores. If distinct factors were found, correlational analysis of these factors determined the extent to which they were related. Linear regression analysis of these factors was also used to estimate which factors best predict overall scores on the DRA.

For the second research question, I tested the alignment of EQAO, DRA, and report card grades by comparing means and correlations. High correlational coefficients indicated that the constructs share common variance across the three assessments. Regression analysis was conducted with DRA results as the dependent variable, and report card marks and EQAO results as the independent variables. This provided estimates of their predictive power for determining subsequent DRA scores.
Qualitative Study

Qualitative research refers to face-to-face or observational techniques to collect data in natural settings (McMillan & Schumacher, 2010). Qualitative problems are emergent, open-ended, and describe phenomena instead of being static, closes-ended, and confirmatory. (McMillan & Schumacher, 2010). Qualitative data collection can take many forms, such as in-depth interviews, document and artifact collection, and field observations. A focus group is based on an interview and is made up of a group of selected participants brought together to discuss a relevant issue in an interactive manner. Focus groups produce primarily qualitative data and are used, “... for a variety of different purposes, and (academics) generally tend to rely on them to generate hypotheses and gain insight into participants’ views, perceptions, and attitudes on a given topic” (Litoselliti, 2003, p. 8). Focus groups provide a more natural environment for discussion than formal interviews, and tend to produce more realistic interactions. The emphasis during a focus group is not between the interviewer and interviewee, but on the exchanges between the participants (Litoselliti, 2003). This interaction, “... involves participants learning for each other and reconsidering or re-evaluating their own understanding and experiences” (Litoselliti, 2003, p. 19).

Participants

The target population of the focus group was all classroom teachers and Student Support Teachers (SST) at my school. This group was represented through a convenience sample of four teachers, one from the academically enriched program, two from the neighbourhood program, and one SST: who was responsible for students in all four programs. Ideally, a homeroom teacher from each of the school programs would have participated. There are only two teachers
in the learning difficulties program and neither was available. Of the four teachers in the artistic program, one was conducting the focus group. These four participants, two male and two female, were similar in terms of their education. Each had at least five years teaching experience with Grades 7 and 8 students, and they had a good working knowledge and first-hand experience using the DRA. Focus group participants tend to express personal views and disclose more to people they perceive as being similar to themselves (Greene, Kreider, & Mayer, 2011; Litoselliti, 2003; Morgan, 1997). Greene, Kreider, and Mayer (2011) state that focus groups participants should be “peer groups of professional teams” (p. 63) to make discussions as close to real-life. Litoselliti (2003) and Fern (2001) suggest that six to 10 is an optimal number of focus group participants, others report that groups of three of four people are now common in small-scale social research, especially if unique perspectives or specialized information are sought (Denscombe, 2007; Litoselliti, 2003; Morgan, 1997). Small groups are easier to organize, manage, and analyse (Litoselliti, 2003; Morgan, 1997) and are considered more appropriate for complex or controversial topics or for detailed accounts.

**Instrumentation**

The instrument of this section of the study was a semi-structured, open-ended focus group interview with the four participants. A focus group is a more time-efficient method of data collection than a series of individual interviews (Morgan, 1997) but also reveals, “... aspects of experiences and perspectives that would not be as accessible without group interaction” (Morgan, 1997, p. 20). Since focus groups are led by a moderator who directs discussion topics, they are more controlled than participant observation data collection methods,
but are less controlled than individual interviews. There is no evidence that this moderator
influence during focus groups is greater than during interviews (Morgan, 1997). I moderated the
focus group, which was conducted at my house. This site was chosen as an “off-site” from the
school to put participants at ease. Fern (2001) states that moderators should blend in with the
group by sharing life experiences. Like the focus group participants I have taught Intermediate
students for at least five years, teach at the target school, and have experience writing report
cards and administrating the DRA.

The focus group topics and questions were constructed to: a) correlate with this study’s
research questions, and b) compare Intermediate teacher perspectives to some of the claims
presented in the DRA Technical Manual (2009). The developers stated that the DRA, “... 
provides all the information required to individually pinpoint the reading level of a student ...”; 
that the administration of the test is “ easy” and “seamless” and that it provides, “ ... 
comprehensive teaching support and training options” (Beaver & Carter, 2009, p. 5).

Data Collection

Using the school specific messaging board provided by the school board, all teachers at
the target school were invited to participate in a focus group concerning the use of the DRA with
Intermediate students. Teachers were provided with a Letter of Information and an Informed 
Letter of Consent. Due to the aforementioned time constraints, this focus group was conducted
in July 2014, during the teachers’ summer vacation. Of the possible 15 teachers, 4 participated,
providing a participation rate of 27%.
The focus group was conducted at a scheduled and designated area chosen to be comfortable for participants. Litoselliti (2003) notes the importance of participants being at ease, away from noise and distractions. The pertinent portion of the focus group lasted approximately an hour. In addition to written observations, the focus group interview was recorded using Apple’s GarageBand software on a laptop. This recording was converted to an MP3 file and then transcribed using TextEdit software. To preserve confidentiality, each participant’s name was replaced with a code based on the order participants joined the conversation. All participants were provided with a copy of the transcript for review. This focus group was conducted to gain information regarding teachers’ perceptions on the consequences of using the DRA with Intermediate students, challenges presented in the administration of the assessment, and suggestions for its use.

Data Analysis

Given the nature of the focus group data, I largely used qualitative techniques. “Sophisticated debates about independent observations, random sampling, and so on do not apply to using descriptive statistics to summarize the content of focus groups” (Morgan, 1997, p. 61). According to Denscombe (2007) there are four steps to interpreting qualitative data: 1) coding the data; 2) categorizing the codes; 3) identifying theme and relationships between the codes and categories; and 4) developing concepts that generalize the analysis. This is an iterative, or spiral process, in which each of the four steps are likely to be revisited during analysis. My initial analysis coded participants’ responses according to the open-ended questions I posed. These five question-based themes and coding were amended to group the data into four
categories: 1) Usefulness of the DRA; 2) Challenges in DRA Administration; 3) Attitudes of Students and Teachers; and 4) What is Not Measured / Recommended Changes.
CHAPTER 4

RESULTS

Little research has examined the use of the Developmental Reading Assessment (DRA) with Intermediate students. Evidence of reliability and validity testing was provided in the DRA Technical Manual (2009) but indicated that mostly primary and junior aged children participated in the DRA developers’ fieldwork. This study investigated two questions: 1) Is the DRA a statistically reliable measure of reading comprehension for Grade 7 and 8 students? 2) Is there evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension? The purpose of the study was achieved by examining: the psychometric properties of the DRA; the correlation of DRA, Report Card and EQAO reading comprehension marks; and by conducting a focus group of experienced Intermediate teachers.

This chapter presents the results of the data analysis for the stated research questions and is organized into two sections. The first section provides quantitative results from a statistical analysis of the internal reliability of the DRA, correlations between DRA subtasks, and a factor analysis to examine how different sections of the DRA account for unique variance. Descriptive statistics for the DRA scores, report card marks and EQAO results are presented to examine how well these measures produced results along a normal distribution curve and resulted in similar decisions regarding student reading achievement. Correlational analysis between these measures is also presented. The second section of this chapter examines responses collected from a focus group of four teachers. These data have been organized into four main themes: 1) Usefulness of the DRA; 2) Challenges in DRA Administration; 3) Attitudes of Students and Teachers; and 4)
What is Not Measured / Recommended Changes. The results chapter concludes with a summary of findings for the two research methods used in this study.

**Quantitative Results**

Reliability analysis was used to study the properties of the DRA and its components and provided information regarding the relationship between individual test items. Students’ DRA scores are determined by adding the Reading Level to the scores from the Student Reading Survey, oral reading passage, and written component. My initial analysis used these four scores to examine inter-item relationships and internal consistency. The correlations provided in Table 1 indicate that the DRA text level and survey questions (wide reading and self assessment) had few statistically significant correlations with the other items. The oral and written items had statistically significant correlations with almost every other item. The strongest of these correlations was between the oral items: expression and phrasing, $r = 0.76$; expression and rate, $r = 0.68$; and phrasing and rate, $r = 0.65$; as well as between literal comprehension and interpretation, $r = 0.67$. 


Table 1

Correlations Between DRA Items Including DRA Text Level and Survey

<table>
<thead>
<tr>
<th></th>
<th>dratl</th>
<th>wr</th>
<th>sa</th>
<th>ex</th>
<th>ph</th>
<th>ra</th>
<th>ac</th>
<th>pr</th>
<th>su</th>
<th>lc</th>
<th>in</th>
<th>re</th>
<th>me</th>
</tr>
</thead>
<tbody>
<tr>
<td>wr</td>
<td>.21</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sa</td>
<td>.23</td>
<td>.44**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ex</td>
<td>.36</td>
<td>.40</td>
<td>.15</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ph</td>
<td>.52</td>
<td>.43</td>
<td>.22</td>
<td>.76**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ra</td>
<td>.23*</td>
<td>.30</td>
<td>.13</td>
<td>.68**</td>
<td>.65**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac</td>
<td>.23</td>
<td>.28*</td>
<td>.13</td>
<td>.46**</td>
<td>.49**</td>
<td>.50**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pr</td>
<td>.06</td>
<td>.40**</td>
<td>.22</td>
<td>.44**</td>
<td>.57**</td>
<td>.48**</td>
<td>.43**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>su</td>
<td>.09</td>
<td>.38**</td>
<td>.27*</td>
<td>.51**</td>
<td>.49**</td>
<td>.51**</td>
<td>.37**</td>
<td>.39**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lc</td>
<td>-.08</td>
<td>.36**</td>
<td>.27*</td>
<td>.46**</td>
<td>.34**</td>
<td>.42**</td>
<td>.24*</td>
<td>.49**</td>
<td>.50**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in</td>
<td>-.10</td>
<td>.35**</td>
<td>.31*</td>
<td>.52**</td>
<td>.36**</td>
<td>.34**</td>
<td>.40**</td>
<td>.34**</td>
<td>.46**</td>
<td>.67**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>re</td>
<td>.11</td>
<td>.46**</td>
<td>.33**</td>
<td>.51**</td>
<td>.51**</td>
<td>.40**</td>
<td>.39**</td>
<td>.43**</td>
<td>.51**</td>
<td>.43**</td>
<td>.54**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>me</td>
<td>.02</td>
<td>.47**</td>
<td>.44**</td>
<td>.27*</td>
<td>.24*</td>
<td>.20</td>
<td>.25*</td>
<td>.30*</td>
<td>.42**</td>
<td>.48**</td>
<td>.47**</td>
<td>.54**</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Sample size varied from 67 - 72. ** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed). dratl = DRA text level; wr = wide reading; sa = self assessment; ex = expression; ph = phrasing; ra = rate; ac = accuracy; pr = prediction; su = summary; lc = literal comprehension; in = interpretation; re = reflection; me = metacognition.

Internal consistency provides an estimate of reliability based on the homogeneity of the responses to items on a test. The higher the value of Cronbach’s alpha, the more the combined items provide consistent scores. The initial interval consistency calculations for all of the DRA items used to compute student scores (text level, survey, oral and written components), resulted in a Cronbach’s alpha of .52. This is generally considered a relatively low degree of
homogeneity. Item-Total Statistics indicated that this value would increase to .89, a very good level of internal consistency, if the DRA text level score was deleted from analysis.

The DRA Technical Manual reports internal consistency between the oral fluency items and between the written measures (Beaver & Carter, 2009, p. 30). However, internal consistency values are not given that include student survey results or DRA reading levels, both of which are used to calculate students’ reading levels for teaching and reporting. The DRA reading level’s omission may be due to this item being on a very different scale to the oral fluency and written measures items. This does not account for the survey’s omission as it uses the same scale as the oral fluency and written items. The survey, reading level, oral fluency and written measures are all used to calculate student reading scores. This missing analysis challenges the ability to determine the quality of critical aspects of the DRA, making it more difficult to assess the validity of decisions made about student reading for teaching and reporting purposes.

Table 2 provides the organization of individual test items into sections of the DRA.
A factor analysis of the DRA scores was conducted to determine the underlying constructs represented by the items to account for the observed pattern of results. This analysis included the survey questions, reading level, oral items, and written items used to estimate students’ overall DRA scores. Using Principal Components Analysis, initial extraction indicated three factors with Eigenvalues greater than 1, accounting for 66.9% of variance. The Oblimin Rotation Method with Kaiser Normalization was used because it was likely that these different components would be associated with each other. The resulting loadings are shown in Table 3 and indicate which items comprise these sections of the DRA.

This analysis produced results differing from those of the DRA developers. Using Principal Components Analysis, the developers found two factors accounting for 58% of
variance at the 40 - 80 (Grades 4 - 8) Reading Levels. No analysis using a more restricted reading level range was provided. The developers found that Oral Fluency items loaded heavily onto one factor and Written items loaded heavily on the other factor, supporting their claim that the DRA has construct validity. However, their analysis omitted the DRA text level and survey scores.

Table 3

Factor Loadings of All DRA Components

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRA Text Level</td>
<td>.39</td>
<td>.82</td>
<td>.24</td>
</tr>
<tr>
<td>Wide Reading</td>
<td>.17</td>
<td>.08</td>
<td>.70</td>
</tr>
<tr>
<td>Self Assessment</td>
<td>-.18</td>
<td>.16</td>
<td>.92</td>
</tr>
<tr>
<td>Expression</td>
<td>.88</td>
<td>.07</td>
<td>-.02</td>
</tr>
<tr>
<td>Phrasing</td>
<td>.88</td>
<td>.26</td>
<td>.08</td>
</tr>
<tr>
<td>Rate</td>
<td>.90</td>
<td>.00</td>
<td>-.17</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.68</td>
<td>.07</td>
<td>-.02</td>
</tr>
<tr>
<td>Prediction</td>
<td>.62</td>
<td>-.21</td>
<td>.08</td>
</tr>
<tr>
<td>Summary</td>
<td>.60</td>
<td>-.35</td>
<td>.18</td>
</tr>
<tr>
<td>Literal Comprehension</td>
<td>.35</td>
<td>-.58</td>
<td>.25</td>
</tr>
<tr>
<td>Interpretation</td>
<td>.35</td>
<td>-.54</td>
<td>.26</td>
</tr>
<tr>
<td>Reflection</td>
<td>.44</td>
<td>-.19</td>
<td>.42</td>
</tr>
<tr>
<td>Metacognition</td>
<td>-.02</td>
<td>-.26</td>
<td>.74</td>
</tr>
</tbody>
</table>

The Oral items (Expression, Phrasing, Rate, and Accuracy) loaded together on Factor 1, but the Written items (Prediction, Summary, Literal Comprehension, Interpretation, Reflection, and Metacognition) were split between Factors 1, 2 and 3. The DRA Text level loaded heavily onto Factor 2. Factor 3 largely contained the survey items and metacognition, a written item that measures a student’s understanding of their reading processes. The Component Correlation Matrix showed a very weak, -0.13 and -0.17, correlations between Factor 2 and the other two factors. A subsequent factor analysis was conducted using only the oral and written items, replicating the Factor Analysis reported by the DRA developers in the Technical Manual. Using Principal Component Analysis and the Oblimin Rotation Method with Kaiser Normalization resulted in a 2-factor solution accounting for 63.5% of variance. The factor loadings shown in Table 4 indicate that items generally loaded heavily onto one of the two factors.
Table 4

Factor Analysis of Oral and Written Items from the DRA

<table>
<thead>
<tr>
<th>DRA items</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Expression</td>
<td>.81</td>
</tr>
<tr>
<td>Phrasing</td>
<td>.92</td>
</tr>
<tr>
<td>Rate</td>
<td>.89</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.69</td>
</tr>
<tr>
<td>Prediction</td>
<td>.59</td>
</tr>
<tr>
<td>Summary</td>
<td>.39</td>
</tr>
<tr>
<td>Literal Comprehension</td>
<td>.09</td>
</tr>
<tr>
<td>Interpretation</td>
<td>.10</td>
</tr>
<tr>
<td>Reflection</td>
<td>.27</td>
</tr>
<tr>
<td>Metacognition</td>
<td>-.20</td>
</tr>
</tbody>
</table>


Oral items 1 - 4 plus Prediction loaded onto Factor 1, Summary writing was split between Factors 1 and 2, and the remaining written items loaded onto Factor 2.

Results provided by the DRA developers indicated that two factors emerged at the 40-80 reading level, accounting for 57.6% of variance. This analysis did not reflect the use of DRA reading level scores, nor the survey questions, both of which are used by teachers to calculate students’ scores. Beaver and Carter’s (2009) analysis displayed factor loadings in which oral component items loaded exclusively onto one factor and written items loaded onto the second factor. The developers used this result to provide evidence for the construct validity of the DRA.
As reported earlier, my initial factor analysis, using all the components used to score the test did not replicate this two factor loading. A subsequent analysis, repeating Beaver and Carter’s procedure, also failed to replicate Beaver and Carter’s results.

A linear regression analysis of the DRA items determined which items were the best predictors of DRA results. Sample sizes can affect the generalizability, and therefore the scientific value of results. Pallant (2007) cites differing guidelines for determining adequate sample sizes for multiple regression analysis, including Stevens (1996, p. 72) “15 participants per predictor” and Tabachnick and Fidell (2007, p. 123) \( N > 50 + 8M \), where \( N \) is the sample size and \( M \) is the number of predictors. Using these guidelines, I decided to reduce the number of predictors by grouping DRA items into four predictors: DRA reading level, survey results, oral items, and written items.

The results of the analysis are shown in Table 5. The results present an evaluation of how well the total DRA results are predicted by each of the four grouped predictors. Written items accounted for most of the variability, \( B = .55 \), followed by oral items, \( B = .42 \), and the survey, \( B = .21 \). Analysis of the DRA text level did not account for any statistically significant variability. These results indicate that success on the written items makes a greater unique contribution to total DRA scores than oral items or the survey, echoing teacher opinions voiced during the focus group, that the DRA is a test of written skills.
Table 5

Regression Coefficients of DRA components

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>writtentotal</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>oraltotal</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>surveytotal</td>
<td>1.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

a. Dependent Variable: dra total

The second question of this study was: Is there evidence that the results from the DRA provide valid interpretations regarding students’ reading skills and comprehension? To answer this, I compared DRA results with other reading measures: the EQAO Reading results and Report Card results for reading. All three claim to measure a single construct (reading comprehension), and should therefore demonstrate a high degree of correlation for individual students. Table 6 illustrates the descriptive characteristics of these measures and Table 7 illustrates the correlation between these measures.
Table 6

Descriptive statistics for the DRA, EQAO results and report card marks

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>% at grade level</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQAO reading overall</td>
<td>67</td>
<td>3.24</td>
<td>.63</td>
<td>89.56</td>
</tr>
<tr>
<td>Report card June 2014</td>
<td>76</td>
<td>74.97</td>
<td>6.87</td>
<td>56.58</td>
</tr>
<tr>
<td>Report card February 2014</td>
<td>74</td>
<td>74.95</td>
<td>6.22</td>
<td>62.16</td>
</tr>
<tr>
<td>Report card June 2013</td>
<td>69</td>
<td>79.06</td>
<td>7.16</td>
<td>75.36</td>
</tr>
<tr>
<td>DRA text level</td>
<td>72</td>
<td>72.22</td>
<td>7.36</td>
<td></td>
</tr>
<tr>
<td>DRA wide reading</td>
<td>68</td>
<td>3.06</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>DRA self assessment</td>
<td>68</td>
<td>2.41</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>DRA expression</td>
<td>71</td>
<td>3.20</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>DRA phrasing</td>
<td>71</td>
<td>3.28</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>DRA rate</td>
<td>71</td>
<td>3.04</td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td>DRA accuracy</td>
<td>71</td>
<td>3.25</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>DRA prediction</td>
<td>72</td>
<td>3.17</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>DRA summary</td>
<td>72</td>
<td>3.00</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td>DRA literal comprehension</td>
<td>72</td>
<td>2.86</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>DRA interpretation</td>
<td>72</td>
<td>2.72</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>DRA reflection</td>
<td>72</td>
<td>2.72</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>DRA metacognition</td>
<td>72</td>
<td>2.57</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>DRA total score</td>
<td>67</td>
<td>107.70</td>
<td>11.26</td>
<td></td>
</tr>
</tbody>
</table>

Analysis for skewness of the distribution produced mostly negative values for the DRA as a whole and for individual items. This deviation from a normal distribution curve indicated
that items were generally too easy, however values were not considered statistically significant for every item (Vogt, 2007). DRA text level was highly skewed at -1.47, SE .28, as were all four oral items ranging from rate -.63, to accuracy 1.11, with a SE of .28 for both. The DRA as a whole was highly skewed at -1.63, SE .29. Kurtosis, the measure of the flatness or pointedness of a distribution curve, was only statistically significant for the DRA reading level at 4.73, SE .56. This value is unsurprising given that the vast majority of participants were given the level 70 or level 80 reading sample. Analysis of report cards and EQAO scores for kurtosis and skewness indicated no statistically significant results.

Table 7

Correlation between DRA results, report card marks and EQAO reading scores

<table>
<thead>
<tr>
<th></th>
<th>DRA total</th>
<th>report 2014</th>
<th>report 2014feb</th>
<th>report 2013</th>
<th>EQAO overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>report card 2014june</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>report card 2014february</td>
<td></td>
<td>.76**</td>
<td>.82**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>64</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>report card 2013june</td>
<td></td>
<td>.62**</td>
<td>.64**</td>
<td>.61**</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>58</td>
<td>69</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>EQAO overall</td>
<td></td>
<td>.62**</td>
<td>.60**</td>
<td>.66**</td>
<td>.52**</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>57</td>
<td>67</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>EQAO raw score</td>
<td></td>
<td>.68**</td>
<td>.63**</td>
<td>.65**</td>
<td>.61** .88**</td>
</tr>
<tr>
<td>N</td>
<td></td>
<td>57</td>
<td>67</td>
<td>66</td>
<td>66 67</td>
</tr>
</tbody>
</table>

Note: **. Correlation is significant at the 0.01 level (2-tailed)
Moderate to strong statistically significant correlations were found between all of the measures. The EQAO overall score is the published score and is represented by a Level 1, 2, 3, or 4. The EQAO raw score is not publicly available and is reported to 1 decimal place. This score was included in the analysis to provide more precision within each level. EQAO scores are quasi-interval in nature. The distance between points is not equal and this poses a limitation to the significance of any statistical analysis that uses them. The EQAO overall score is derived from the raw score, hence the high correlation between them.

A strong correlation existed between the February, 2014 and June, 2014 report cards which could be explained by halo effect. This halo effect occurs when judgements are made based on general impressions of a student instead of on the specific criteria being evaluated. (Brookhart & Nitko, 2008). This possible bias may also explain the moderate to high correlation between these two reports and the DRA administered by the same teacher. The DRA was administered in the Fall of 2013 and the chronologically closest report to it was June 2013. If they are measuring the same construct, it is surprising that the correlation between these two reading measures was the weakest. Measures performed by a single teacher: the DRA, February, 2014 and June, 2014 report cards were more highly correlated to each other than those scored by other teachers, the EQAO assessment and the June 2013 report card, indicative of a halo effect.

Tables 8 - 10 is a series of two-way contingency tables displaying whether statistical relationships existed between the DRA, EQAO, and report card marks in regards to their ability to determine a minimum competency. *At grade level* was the cut-off score used in this analysis, and varied between measures. For the DRA, a minimum score of 104 for Grade 7 and a score of 114 for Grade 8 indicated *at grade level* (Beaver & Carter, 2004). EQAO grade level is set at
Level 3 and on report cards it has a value of 75%. Students achieving grade level results were assigned “1” while those under this minimum were assigned a value of “0.” The following three tables show the relationships between the measures. Optimally, percentages at 0.00, below grade level for both measures, and 1.00, at grade level for both measures should be high, demonstrating that students categorized as at or below grade level on one measure are categorized the same way on the other measure. Missing data for either the DRA or EQAO scores resulted in a reduction in sample size.

Table 8

*At Grade Level Decision Consistency for EQAO Results and Report Card Marks*

<table>
<thead>
<tr>
<th>EQAO results</th>
<th>Report Card June 2014</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>(85.7%)</td>
<td>(14.3%)</td>
<td></td>
<td>(100.0%)</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>24</td>
<td>36</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>(40.0%)</td>
<td>(60.0%)</td>
<td></td>
<td>(100.0%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>37</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>(44.8%)</td>
<td>(55.2%)</td>
<td></td>
<td>(100.0%)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Pearson $X^2 (2, N = 67) = 5.30, p = .02$, Cramer’s $V = .28$. Percentages within EQAO results appear in parenthesis below observed frequencies.
Table 9

At Grade level Decision Consistency for EQAO Results and DRA Scores

<table>
<thead>
<tr>
<th>DRA scores</th>
<th>EQAO results</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>5 (22.7%)</td>
<td>17 (77.3%)</td>
</tr>
<tr>
<td>1.00</td>
<td>0 (0.0%)</td>
<td>35 (100.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>5 (8.8%)</td>
<td>52 (91.2%)</td>
</tr>
</tbody>
</table>


Table 10

At Grade Level Decision Consistency for DRA Scores and Report Card Marks

<table>
<thead>
<tr>
<th>DRA scores</th>
<th>Report Card June 2014</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>19 (79.2%)</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>1.00</td>
<td>10 (25.0%)</td>
<td>30 (75.0%)</td>
</tr>
<tr>
<td>Total</td>
<td>29 (45.3%)</td>
<td>35 (54.7%)</td>
</tr>
</tbody>
</table>

Note. Pearson $X^2 (2, N = 64) = 17.7$, $p < .001$, Cramer’s $V = .53$. Percentages with DRA scores appears in parenthesis below observed frequencies.
By computing Cramer’s measure of association, we can see that the effect size of decision frequencies was medium for June 2014 report cards versus EQAO, (.28) and for EQAO versus DRA (.36), but was large for DRA versus June 2014 report cards (.51) (Huck, 2008). The proportion of good decisions varied little between the three relationships. Decision consistency between: 1) June 2014 report card and EQAO results = .63; 2) EQAO results and DRA scores = .70; and 3) DRA scores and June 2014 report cards = .77.

Qualitative Results

The focus group questions were designed to explore some of the statements made by the DRA developers in their Technical Manual (2009) and Teacher Guide (2004). These included claims of: ease of use, and administration and “that (it) provides all the information required to individually pinpoint the reading level of a student and subsequently link assessment information with appropriate and individualized instructional steps” (DRA Technical Manual, 2009, p. 5). Within the focus group data, four main themes emerged: (a) Usefulness of the DRA, (b) Challenges in DRA Administration, (c) Attitudes of Students and Teachers, and (d) What is Not Measured / Recommended Changes. Direct quotes were used as section headings to illustrate a commonly expressed view or to provide unique or insightful perspectives. For ease of reading, the participating teachers have been assigned pseudonyms. Alex and Andy are the teachers from the neighbourhood program, Ray teaches in the academic program, and Taylor is the SST.

It’s not a good tool but we don’t have any other tool

A significant difference in the overall usefulness of the DRA was expressed between teachers in different programs. The teacher from the academic program stated that the DRA was used little in programming. Because all of the students coming into the program were very
strong academically, they all scored at least grade level on the DRA. According to Ray, “That’s one of the limitations, is there’s a good possibility that when our students come to us in grade 7, they’re reading at grade 8 level.” This teacher found that some students’ DRA scores dropped over the course of Grade 8. “...when I did it at the end of the year I was finding very little difference, in fact some kids actually go backwards. I think that wasn’t the case of my teaching but it (the DRA) didn’t really reflect what I was teaching.” It was posited that one cause of this phenomena was that the content of the DRA was not aligned with what is taught in his academic reading program. He suggested that more higher order thinking questions would provide him with better information about his students.

In contrast, the regular program teachers and Student Support Teacher (SST) found administering the DRA was useful at the start of the school year. Taylor mentioned using it to group students for guided reading groups. She also found DRA scores an effective indicator of students that may need special accommodations or modifications to their programs. Andy suggested that the DRA is more useful in classrooms where there is a greater variety of reading comprehension levels than what occurs in the academic classes.

Taylor, Andy, and Alex agreed that they were disappointed in the DRA because it measured a student’s ability to write an answer, not if a student could read the text. They felt that DRA scores did not truly reflect a student’s ability to read and that students would score differently in assessment of their guided reading conversations. According to Andy, “I think the biggest difficulty with the DRA is the pen”.

The DRA Teacher Guide does not provide any information regarding accommodations or modifications to the written items for students with expressive language learning difficulties.
other than to follow the recommendations of the Individual Education Plan (IEP). An interesting line of inquiry might look at the discrepancy between written answers to the DRA items and oral answers to the same questions. As a follow-up, Taylor questioned administering procedures for students identified with learning disabilities. She wondered whether students with Language Disabilities should be given DRA texts at their expressive language level, or at their receptive language level. Often, what these students comprehend is very different from what they can express. When describing a student with a reading disability, Taylor stated, “it’s going to come out that he’s reading at like Grade 2 level ... but intellectually he’s thinking at Grade 7 or 8 level so are we going to penalize him or are we going to give him Grade 3 reading material? or are we just going to give him the accommodations he needs to be successful at Grade 7 or 8?”

Participants became more animated when discussing the Reading Survey section of the DRA. Student disengagement and apathy in the regular and academic programs was mentioned by Ray and Andy to especially affect students’ willingness to provide a detailed answer about their reading habits and goals. According to Ray, “even our (academic) kids are too apathetic to, in a lot of cases, really put detailed answers to that (the survey).” Taylor and Ray confided that they coached their classes on how to answer this section. The participants felt that aside from apathy, many students are not aware of what they are reading. Students may not regard reading on-line, in chat rooms, on blogs, in newspapers, graphic novels, comics, and magazines as real reading. The DRA survey asks students to list the books they have read, not reading from any other sources. This omission of non-book reading could lead to under-reporting of reading activities, with a low Wide Reading score as the consequence. Difficulties with marking the survey will be further detailed, but to quote one of the participants, “The Goal Setting (subtask),
it’s terrible.” In fact, the participants agreed that the Reading Survey was, “... the least profitable section” of the DRA.

The Oral Reading subtask was regarded more favourably by all of the participants except the academic program teacher. Ray found that the subtask had no discriminatory power because, “... most kids score, almost 95% of them, score really, really high on it.” This ceiling effect was less noticeable amongst the other teachers. Taylor and Alex mentioned the usefulness of determining, through one-on-one sessions, the reading comprehension strategies used by students, and whether the students were reading for meaning.

Regular program teachers and the SST all stated that while the written section of the DRA did not necessarily reflect students’ reading comprehension abilities, it was still a useful measurement tool. They noted that this section of the DRA displayed a student’s perseverance, learning skills, and willingness to take risks. As Alex noted, “I think it (the written section) gives me a lot of information about their learning skills, and I’ve come to understand that what is written there (in the DRA) doesn’t necessarily reflect the sum total of their comprehension .” Ray felt that unlike the oral Reading subtasks, the written Inferring and Metacognition questions best assessed student reading levels, while the Summarizing section did not. The participants agreed that not surprisingly, students scored lower on the more difficult questions, as stated by Andy, “We already know, I think, where the difficulties are going to be (open-ended questions, inferencing, higher-order thinking) ...They can all recall facts.”

Yes, there are a lot of challenges

All participants noted difficulties administering the DRA in their Intermediate classrooms and their responses illustrated the lack of consistent procedures across classrooms. The one-on-
one oral reading subtask proved the most troublesome, regardless of administration procedure. For teachers who removed themselves from the classroom, by setting up desks in the doorway and hallway, classroom behaviour management of the remaining students was mentioned. “I guarantee you there hasn’t been a set (of DRAs) that I have done where I haven’t had to stop a kid in the middle of a reading record ... because the class, someone in the class or the class as a whole is completely gone squirrelly while your attention is elsewhere” (Ray). Teachers that were able to stay in the room still had some behavioural issues because of the amount of time students were expected to work independently. Alex described administering the DRA, “where students are not always focussed and on task and independent workers ... for the required amount of time is very difficult.” Each oral reading session takes between 6 - 10 minutes (Beaver & Carter, 2004), so a class of 28 students needs to keep itself quietly occupied for 168 to 280 minutes. Andy noted his shock when realizing how long the administration and marking of the DRA would take. Beaver and Carter (2004) acknowledge that scoring the written items of the DRA is labourious. They suggest that with practice, teachers will reduce this time to 10 to 12 minutes per pupil. For a class of 28, that is another 280 to 320 minutes. Aside from this required time, Ray noted that the timing of the DRA administration in the early fall made setting classroom tone difficult, “… and it’s frustrating when you are trying to establish those routines, to have lost that momentum so early.”

Additionally, there were other challenges to scoring the DRA written items. Taylor, Alex and Andy were unaware of the existence of the Teacher’s Manual that contained the marking exemplars and procedures for administration of the DRA. Ray was familiar with the manual and noted the absence of exemplars for all texts at each reading level and achievement level. He
remedied this omission by collecting actual student examples which he believed represented achievement Level 4. The lack of complete exemplars and discrepancies in marking procedure call into question the DRA’s reliability.

The Reading Survey asks teachers to assign marks based on the breadth of reading across genres and reading levels. Alex noted that it was very difficult to be familiar with everything that students list and that assigning a grade level, “Often times it’s guess work.” All of the participants felt that the survey was the least helpful part of the DRA, that it was difficult to evaluate, and Taylor stated, “I hate that about that!” Alex and Taylor felt they would get a much better assessment of a student’s reading level by letting them answer the written component orally, but that this was not possible given how long it would take.

Text selection provided another source of differing administration procedures between participants. Taylor let students choose between the three to four texts available at each reading level. Alex, Ray, and Andy, for ease of marking and tracking, had every student read the same text within a reading level. Alex noted that when teachers choose the reading level to give, “You guess” especially when testing students who may be below grade level. Taylor stated that in other schools in this school board, all students in a class would receive the same level of text, regardless of whether they were on modified programs for reading difficulties. It was noted by Andy that some students had read all of the available texts by the time they reached Grades 7 and 8. These students were being assessed on texts with which they were already familiar, representing a challenge to the DRA results’ validity.

When asked directly for their “gut feeling” about the DRA’s reliability and the validity of its assessment, Taylor stated that it was a “good predictor” of reading ability for all students but
especially useful for identifying students who need modified reading programmes. Alex felt that, “it’s (the DRA) definitely good value for time for a select group of my students... I always get a few surprises in there, like, “Wow, you were much lower that I thought you would be.” However, all participants noted issues which they felt might affect the DRA’s ability to assess students’ reading comprehension.

Even with exemplar use, Taylor called the marking “pretty subjective.” Andy noted that a discrepancy occurs between a student’s reading level in June, as assessed by one teacher, and then the following September, as assessed by a different teacher. Maturation effects were discounted by the participants. Inter-rater reliability was also questioned as the DRA is not administered by the classroom teacher for some select students, but by other people such as Student Support Teachers and Educational Assistants.

Other procedural issues were seen as challenges to the DRA’s validity. These challenges included: (a) stopping a student midway through the Oral Reading for classroom management (b) student anxiety from being singled out into the hallway or the back of the classroom, (c) coaching by some teachers, and (d) differences in text choice. One participant noted concern that results from the DRA coloured his marking of students’ work for the rest of the year.

_Do we really have to do this again?_

As reported previously, participants stated that student disengagement and apathy affected the thoroughness of students’ answers, especially during the written sections of the DRA. The results did not truly reflect students’ reading abilities, thereby questioning the validity of the DRA results. Andy suggested that the timing of the DRA in September exacerbated this attitude, “And so the first thing we do is hit them with this four or five diagnostic thing, like,
“I’m going to do page one, but I’m not doing 2, 3 4.” Unique characteristics of Intermediate students was also mentioned by Taylor as a factor. “I think that there is so much more going on when they are 12, 13, and 14 ....” such as puberty, and adolescence,. Alex noted, “They don’t care about pleasing us (teachers) anymore”, to which Andy added, “especially if they’ve done it (the DRA) five or six years in a row by now.” Social issues at the target school, such as conflict at home and absenteeism, were cited as outside factors affecting student attitudes towards school and the DRA results in particular, “But they are not engaged even when they come to school in September after Labour Day.” according to Andy. The focus group came to the consensus that the DRA did not take into consideration that Grades 7 and 8 is, “a very sensitive age group.”

I’m doing it because I’ve been told to do it

Participants stated that they administered the DRA simply because it had been mandated. According to Andy, “I’m doing it because I’ve been told to do it. And I’ll give you (students’) scores and then we’ll move on.” Andy, Taylor, and Ray also questioned the timing and frequency of DRA administration, “I (Andy) don’t see the point in me doing it again in June, and then teacher over here (gesturing to Alex) doing it again in September ... I don’t think the kids are going to read a whole lot in the nine weeks in between that are going to help them.”

Conversely, Alex thought DRA administration was useful in September, “I think that mixed with reliable professional judgement, it’s fine, it is a great way to start off the year. Would it be great, would it be even greater if there was something better? Absolutely.”

In general, the focus group participants were in favour of using the DRA with Intermediate students but qualified their support. As expressed Alex,

Because what else is there? ...it’s not a good tool but we don’t have any
other tool... So if there was something else to replace it (the DRA), that would meet our needs in a better way, absolutely we’d jump ship and head to that right away, ...

it’s a great measuring stick cuz it’s the only stick we have.

Alex, Taylor, and Andy felt that in concert with other assessments, the DRA was a fairly comprehensive test of reading. Ray noted that the DRA is not well linked to the Ontario curriculum and does not measure elements of style, such as symbolism or connections. Teachers would like to see a mechanism to measure strategies used in the Oral Reading section. Currently, student responses are counted as correct or incorrect, without measuring the effectiveness of strategies used on unfamiliar words pronounced incorrectly.

*The biggest difficulty with the DRA is the pen.*

Participants noted that the DRA procedures do not follow what are considered good teaching practices. Alex wanted to talk about the text first in literature circles and guided reading groups so that students would better engage with the text and demonstrate a better understanding of it. Extra release time, allotted by the school board for one-on-one student feedback sessions was recommended. Taylor and Alex would have appreciated the option, especially for those identified with learning exceptionalities, for students to answer the written portion of the DRA orally, or as part of a guided reading conversation. Identified students have access to computers with software such as Dragon Naturally Speaking and Kurzweil to allow them to produce their best work. Taylor would like to see the same accommodations and modifications listed in Student Individual Education Plans (IEPs) used during DRA administration. All of the participants were unaware that the Teacher Guide (2004) mentions that IEP recommendations should be followed. This is not surprising as some of the participants were unaware of the
Teacher Guide’s existence. Alex and Ray stated that they received no training on the use of the DRA and were self-taught. Taylor recalled that her training focused on keeping running records during the oral tasks, not on assessing the written questions. Andy recalled that he had “pretty comprehensive” training from a school board official.

Participants felt that an exemplar for every text and at every achievement level should be included in the Teacher Manual. They also wanted to see more texts available at the high reading levels because often students had already read them all, especially by Grade 8. Taylor, felt the texts at the Intermediate level were quite interesting, Alex did not.
The purpose of this study was to explore the use of the Developmental Reading Assessment (DRA) with intermediate students. This study contributes to reading comprehension assessment literature as it is the first time that research has examined the psychometric properties of the DRA with exclusively intermediate students. Second, this study compared DRA data to results from EQAO results and report card data to examine the validity of decisions based on DRA scores. This comparison adds to the body of knowledge regarding alignment of grades and assessments. Third, this study explored the challenges and concerns identified by teachers using the DRA in Grade 7 and Grade 8 classrooms. In the following sections, research findings are discussed in relation to the two research questions posed in the introduction. Next, implications of this study for both reading comprehension assessment and school board policy are presented. Finally, limitations of this study and avenues for further research are presented, followed by a discussion of conclusions to be drawn from this research.

**Is the DRA statistically reliable?**

The first research question was concerned with determining whether the DRA is a statistically reliable measure of reading comprehension for Grades 7 and 8 students. Teachers in Ontario are required to use diagnostic assessments to measure students’ reading comprehension but are allowed to use their professional judgement in the selection of which measure to administer. The DRA is commonly used in my school board for students from Grade 4 to Grade
8, however field testing during the DRA development was conducted primarily with younger children (DRA K-8 Technical Manual, Beaver and Carter, 2009).

The results of an internal consistency analysis provided an estimate of the DRA’s reliability based on the homogeneity of responses to items on the test. Initial calculations using all of the DRA items used for computing students’ scores (text level, survey, oral, and written components), resulted in a Cronbach’s alpha of .52. This is a relatively low degree of homogeneity and an indication of low test reliability. Item-Total Statistics indicated that this value would increase to .89, a very good level of internal consistency, if the DRA text level was deleted from analysis. The DRA’s reliability as a reading assessment is more consistent when the text level is not part of the score. This text level may serve other functions, but it does not help to provide a consistent measure of students’ reading when used in combination with the other DRA scores. The DRA should be used very successfully if the reading level is removed.

Does the DRA provide valid interpretations?

The second research question asked whether there is evidence that the DRA provides valid interpretations of students’ reading skills and comprehension. To be technically adequate, a test must demonstrate reliability and validity. Test validity is more than measuring what the test is supposed to measure, “Validity is a situation-specific concept: It is assessed depending on the purpose, population, and environmental characteristics in which measurement takes place. A test can therefore be valid in one situation and invalid in another” (McMillan & Schumacher, 2010, p. 173). Two methods were used to examine the validity of decisions produced by the DRA: (a) a statistical analysis, and (b) a focus group of experienced teachers.
A factor analysis of test items was conducted in an attempt to replicate the results reported by the DRA developers and to examine the relationships between items. Beaver and Carter (2009) claimed that the DRA had construct validity because its items loaded onto two distinct factors: comprehension and oral fluency. My initial factor analysis used all of the items used to calculated students’ scores, unlike the developers who omitted the text level and survey questions, and failed to replicate this two factor solution. A subsequent factor analysis, using only the oral and written items, also failed to replicate the authors’ results. A two-factor solution was found with oral items loaded together, however written items were split between the two factors. The split in written item loadings indicated that these items do not measure the same underlying construct of reading comprehension and call into question the construct validity of the DRA. A linear regression of the DRA items determined which items best predicted DRA results. Due to the sample size, items were grouped into four variables: DRA reading level, survey results, oral skills, and written skills. Written items accounted for the greatest variability ($B = .55$), followed by oral items ($B = .42$) and the survey ($B = .21$). DRA text levels did not account for any statistically significant variability after controlling for the other predictors. These results indicated that success on the written items made a greater unique contribution to total DRA scores than the survey or oral fluency scores, echoing teacher opinions that the DRA is a test of written skills. The DRA uses written evidence as the proxy to measure reading comprehension, but may introduce a confounding variable. Writing is an expressive language skill, but reading is a receptive skill and many students, especially those identified with language exceptionalities, display significant differences in these two skills. The results of the linear regression may indicate a problem with instrumentation, challenging the DRA’s internal validity.
Beaver and Carter (2009) examined the DRA’s concurrent validity by comparing scores on the DRA to several other well-known tests. This study compared DRA scores to report card data and Grade 6 EQAO reading scores. Analysis for skewness and kurtosis was conducted on these three measures. No statistically significant results of skewness or kurtosis were found for report cards marks or EQAO scores. Significant skewness values were mostly negative for the DRA, indicating a potential ceiling effect (that items may have been too easy). In particular, the DRA test level, all four oral items, and the DRA as a whole were highly skewed. This result implies that the DRA, for these intermediate students, is less useful to differentiate between students and could instead be used as a measure of minimum competency. Moderate to strong statistically significant correlations were found between the DRA scores, report cards, and EQAO results. The strongest correlations existed between the two 2014 report cards (.82) and between these cards and the DRA (February = .76, June = .73). These report card marks and DRA scores were assessed by the same teacher for each student, raising the question of halo effect bias. This form of confirmatory bias is the tendency, for instance when evaluating students’ work, to be influenced by another characteristic or general impression of the students. “The halo effect can seriously reduce the validity of marks assigned by teachers to essay tests ... When essay tests are not read anonymously, the halo effect will seriously confound the results.” (Hopkins, 1998, p. 191). The survey and written component of the DRA are essay-like questions and thus their validity is challenged by this effect. Changes in administration procedures could reduce the risk of halo effect. Students could be assigned identification numbers instead of writing their names on the answer booklets. Each section of the DRA: survey, oral skills, and written skills, could be evaluated independently, with these sections’
results not collated until the end. The other correlations between DRA, EQAO and report cards were moderate, between .52 and .68, indicating that they had shared variance, and likely exhibited commonalities in the constructs they measured.

Test validity is the extent to which inferences or decisions made on the basis of the test are appropriate (McMillan & Schumacher, 2010). Two-way contingency tables determined the decision consistency for the DRA, report cards, and EQAO to determine who read at grade level. Results varied little between the measures, .60 to .64, indicating that they were equivalent in their ability to determine a minimum competency.

This study compared DRA data to report card marks and EQAO scores from intermediate students at my school. Examining the DRA’s internal structure, its content, and relations to other measures, did not provide evidence that the DRA is a valid measure of reading comprehension for intermediate students. The DRA’s authors used a survey of teachers to determine face validity of the measure. This study used a focus group of experienced teachers to explore their perspectives on the usefulness of the DRA, challenges in its administration, attitudes towards the measure, and recommended changes. Views expressed during the focus group echoed many of the challenges to validity found in the statistical analysis of the DRA. Generally, participants found that the DRA, although flawed, was somewhat useful, “...it’s not a great tool but we don’t have any other tool.” Three of the four participants did not view DRA results as a true reflection of a student’s ability to read, given the emphasis on writing ability, “I think the biggest difficulty with the DRA is the pen.” These views mirrored the results of the linear regression.

The teachers felt that the written section of the DRA was more an indicator of perseverance and learning skills, than reading skill, providing a possible explanation for the split
factor loadings for the written items found during the factor analysis. The academic program’s teacher suggested that the oral reading section was too easy and had no discriminatory power, echoing the high negative skewness values calculated for this portion of the DRA. The other three teachers favoured the use of the oral reading section, but all four agreed that the reading survey was, “... the least profitable section” on the DRA. The participants noted several issues as challenges to the validity of inferences made from DRA scores: (a) interrupting student reading assessment to deal with classroom management issues, (b) student anxiety from reading one-on-one, (c) inconsistencies in administration procedures between teachers, (d) incomplete exemplars for grading, (e) lack of training in proper administration procedures, and (f) subjectivity of the grading system.

The use of a focus group of teachers provided the opportunity to gain unique perspectives on how the DRA is used in actual intermediate classrooms. Unlike the survey ratings published by the DRA developers (Beaver & Carter, 2009), these experienced teachers did not strongly endorse the content and usefulness of the DRA, challenging its claims of face validity.

**Research Implications**

Arguably, the most important contribution of this study to the literature is that it provides the first larger scale analysis of the DRA’s psychometric properties with exclusively intermediate students. Results from the statistical analyses and the focus group defined several issues challenging the technical adequacy of the DRA as a measure of reading comprehension for intermediate students. The discrepancy between this study’s results and the authors’ results with mostly primary and junior aged children continues the work of Cutting and Scarborough (2006),
and Keenan, Bjeman, and Olson (2008), that tests vary in sensitivity across different ages and across different skill levels. This study also contributes to the line of study examining the alignment of test scores and report card marks by researchers such as, Willingham et al. (2002), Ross and Gray (2008), and Ross and Kostuch (2011).

This study contributes to an examination of DRA administration procedures. As previously mentioned, test reliability may increase with a revised scoring system that removes text level from the score calculation. Focus group participants found little use for the student survey and the Technical Manual offered no evidence of technical adequacy for this section. Policy makers within school boards may examine whether this section should be included in the assessment. Other proposed changes to DRA administration include: standardized training for all teachers, especially in the use of accommodations and modifications, might alleviate some of the discrepancies in DRA administration, thereby enhancing the validity of its interpretations; an increased number of available student reader texts at each level would reduce the number of students reading the same text over and over; scoring exemplars for every question, of every text, and at each achievement level might improve scoring reliability; revising the scoring scheme of the oral reading subtasks, to make the task more difficult, would improve its discriminatory power and alleviate skewness of results; and assigning identification numbers instead of name writing on the written section booklets would provide anonymity and reduce scorer bias.

Limitations and Avenues for Future Research

A number of limitations must be kept in mind when interpreting the results of this study. While an attempt was made to collect data from all of the Grades 7 and 8 students, only 27%
took part in the study and the majority (63%) of these students came from the artistic program. This over-representation of artistically inclined students may have resulted from me being a teacher in this program. The generalizability of the results to all Grades 7 and 8 students may be problematic, especially if the non-participants differed on any important variable. The low participation rate amongst the other three programs did not allow for a comparison of the relationships between EQAO results, report card grades, and DRA results between programs. Another limitation of this study was the non-participation of a teacher from either the artistic program or the learning difficulties program during the focus group. The Student Support Teacher who works with students from these programs did participate, however homeroom teachers from these programs administer the DRA and may potentially have shared unique perspectives during the group discussion.

Further research might investigate the psychometric properties of the DRA with a greater number of participants. This would allow the aforementioned comparison of relationships between program cohorts. It would also allow regression analysis by individual DRA items instead of grouping them. Item Response Theory (IRT) analysis may also be possible with increased participants. These additional analyses would examine the difficulty of individual DRA items as well as their ability to discriminate. Research using participants from Primary, Junior, and Intermediate grades could attempt to confirm the results published by the DRA developers. A comparison of results after implementation of some of the suggested changes to DRA structure and administration could be another interesting line of inquiry.

In conclusion, the results of the statistical analyses and focus group are in agreement answering this study’s two questions. This study found little evidence that the DRA provides
valid interpretations regarding students' reading skills and comprehension. Factor analysis, skewness and kurtosis, and linear regression analyses produced results challenging the DRA’s internal validity and teachers’ views during the focus group challenged the DRA’s claim of validity. Indicated changes to the structure and administration procedures of the DRA may ameliorate some of these issues.
References


Education Quality and Accountability Office. (2014). EQAO’s technical report for the 2012-2013 assessments: Assessments of reading, writing and mathematics, primary division (grade 1-3) and junior division (grade 4-6); grade 9 assessment of mathematics and Ontario secondary school literacy test. Toronto: Queen’s Printer for Ontario. Retrieved from


Limestone District School Board. (2013a). *Diagnostic assessment toolkit*.

Limestone District School Board. (2013b). *Growth for every student*.


Appendix A

Achievement Chart for Language, Ontario Ministry of Education, 2006

<table>
<thead>
<tr>
<th>Categories</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Knowledge and Understanding</strong> -- Subject-specific content acquired in each grade (knowledge), and the comprehension of its meaning and significance (understanding)**</td>
<td>The student:</td>
<td>The student:</td>
<td>The student:</td>
<td>The student:</td>
</tr>
<tr>
<td><strong>Knowledge of content</strong> (e.g., forms of text; strategies associated with reading, writing, speaking, and listening; elements of style; terminology; conventions)</td>
<td>demonstrates limited knowledge of content</td>
<td>demonstrates some knowledge of content</td>
<td>demonstrates considerable knowledge of content</td>
<td>demonstrates thorough knowledge of content</td>
</tr>
<tr>
<td><strong>Understanding of content</strong> (e.g., concepts; ideas; opinions; relationships among facts, ideas, concepts, themes)</td>
<td>demonstrates limited understanding of content</td>
<td>demonstrates some understanding of content</td>
<td>demonstrates considerable understanding of content</td>
<td>demonstrates thorough understanding of content</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Thinking</strong> -- The use of critical and creative thinking skills and/or processes</th>
<th>The student:</th>
<th>The student:</th>
<th>The student:</th>
<th>The student:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of planning skills</strong> (e.g., generating ideas, gathering information, focusing research, organizing information)</td>
<td>uses planning skills with limited effectiveness</td>
<td>uses planning skills with some effectiveness</td>
<td>uses planning skills with considerable effectiveness</td>
<td>uses planning skills with a high degree of effectiveness</td>
</tr>
<tr>
<td><strong>Use of processing skills</strong> (e.g., making inferences, interpreting, analysing, detecting bias, synthesizing, evaluating, forming conclusions)</td>
<td>uses processing skills with limited effectiveness</td>
<td>uses processing skills with some effectiveness</td>
<td>uses processing skills with considerable effectiveness</td>
<td>uses processing skills with a high degree of effectiveness</td>
</tr>
<tr>
<td><strong>Use of critical/creative thinking processes</strong> (e.g., reading process, writing process, oral discourse, research, critical/creative analysis, critical literacy, metacognition, invention)</td>
<td>uses critical/creative thinking processes with limited effectiveness</td>
<td>uses critical/creative thinking processes with some effectiveness</td>
<td>uses critical/creative thinking processes with considerable effectiveness</td>
<td>uses critical/creative thinking processes with a high degree of effectiveness</td>
</tr>
<tr>
<td>Categories</td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
<td>Level 4</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Communication</strong> – The conveying of meaning through various forms</td>
<td>The student:</td>
<td>The student:</td>
<td>The student:</td>
<td>The student:</td>
</tr>
<tr>
<td><strong>Expression and organization of ideas and information</strong> (e.g., clear expression, logical organization) in oral, visual, and written forms, including media forms</td>
<td>expresses and organizes ideas and information with limited effectiveness</td>
<td>expresses and organizes ideas and information with some effectiveness</td>
<td>expresses and organizes ideas and information with considerable effectiveness</td>
<td>expresses and organizes ideas and information with a high degree of effectiveness</td>
</tr>
<tr>
<td><strong>Communication for different audiences and purposes</strong> (e.g., use of appropriate style, voice, point of view, tone) in oral, visual, and written forms, including media forms</td>
<td>communicates for different audiences and purposes with limited effectiveness</td>
<td>communicates for different audiences and purposes with some effectiveness</td>
<td>communicates for different audiences and purposes with considerable effectiveness</td>
<td>communicates for different audiences and purposes with a high degree of effectiveness</td>
</tr>
<tr>
<td><strong>Use of conventions</strong> (e.g., grammar, spelling, punctuation, usage), vocabulary, and terminology of the discipline in oral, visual, and written forms, including media forms</td>
<td>uses conventions, vocabulary, and terminology of the discipline with limited effectiveness</td>
<td>uses conventions, vocabulary, and terminology of the discipline with some effectiveness</td>
<td>uses conventions, vocabulary, and terminology of the discipline with considerable effectiveness</td>
<td>uses conventions, vocabulary, and terminology of the discipline with a high degree of effectiveness</td>
</tr>
</tbody>
</table>

| **Application** – The use of knowledge and skills to make connections within and between various contexts | The student:                                                           | The student:                                                           | The student:                                                           | The student:                                                           |
| Application of knowledge and skills** (e.g., concepts, strategies, processes) in familiar contexts** | applies knowledge and skills in familiar contexts with limited effectiveness | applies knowledge and skills in familiar contexts with some effectiveness | applies knowledge and skills in familiar contexts with considerable effectiveness | applies knowledge and skills in familiar contexts with a high degree of effectiveness |
| Transfer of knowledge and skills** (e.g., concepts, strategies, processes) to new contexts** | transfers knowledge and skills to new contexts with limited effectiveness | transfers knowledge and skills to new contexts with some effectiveness | transfers knowledge and skills to new contexts with considerable effectiveness | transfers knowledge and skills to new contexts with a high degree of effectiveness |
| Making connections within and between various contexts** (e.g., between the text and personal knowledge or experience, other texts, and the world outside the school; between disciplines)** | makes connections within and between various contexts with limited effectiveness | makes connections within and between various contexts with some effectiveness | makes connections within and between various contexts with considerable effectiveness | makes connections within and between various contexts with a high degree of effectiveness |
# Appendix B

Reading Rubric, EQAO Junior Division, 2015

<table>
<thead>
<tr>
<th>Code</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Blank: nothing written or drawn in the space provided</td>
</tr>
</tbody>
</table>
| I    | * Illegible: cannot be read; completely crossed out / erased; not written in English  
      * Irrelevant content: does not attempt assigned question  
      * Off topic: no relationship of written work to the question  
Typical responses:  
* do not attempt to answer the question OR  
* restate the question (e.g.,) |
| 10   | Response attempts to explain why Joseph-Armand Bombardier’s age is mentioned in paragraphs 1–5.  
The response either:  
* answers an aspect of the question OR  
* does not refer to the reading selection OR  
* provides inaccurate support |
| 20   | Response indicates a partial understanding of why Joseph-Armand Bombardier’s age is mentioned in paragraphs 1–5.  
The response provides:  
* irrelevant support from the reading selection OR  
* vague support from the reading selection OR  
* limited support from the reading selection  
The response usually requires the reader to connect the support to what it is intended to prove. |
| 30   | Response indicates an understanding by explaining why Joseph-Armand Bombardier’s age is mentioned in paragraphs 1–5.  
The response includes:  
* some accurate and relevant support and  
* some vague or underdeveloped support  
The response requires the reader to make some connections between the support and what it is intended to prove. |
| 40   | Response indicates an understanding by explaining fully why Joseph-Armand Bombardier’s age is mentioned in paragraphs 1–5 and provides specific and relevant support. |

Appendix C

Appendix C


<table>
<thead>
<tr>
<th>Teacher Observation Guide</th>
<th>Surveys—Birth of an Island</th>
<th>Level 8B, Page 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Engagement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wide Reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Genre/materials (e.g., chapter books, content or 5 tests)</td>
<td>2. Textually below-grade-level, limited reading experiences</td>
<td>3. Some titles within 3-5 grades or multiple books within a genre, generally on-grade-level text</td>
</tr>
<tr>
<td>2. Vocational and/or goal, may not be directly related to reading</td>
<td>3. Vocational and/or goal(s) identified, vague or no plan</td>
<td>4. Most goals and/or goal related to the reading process or behavior, relevant plan</td>
</tr>
<tr>
<td>Self-Assessment and Goal Setting</td>
<td>5. Multiple strengths and specific goals related to the reading process or behaviors, multiple strategies</td>
<td></td>
</tr>
<tr>
<td><strong>Score</strong></td>
<td>1 2 3</td>
<td>4 5 6</td>
</tr>
<tr>
<td>Oral Reading/Fluency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>1. Link expression, recontextualize</td>
<td>2. Some expression borrowed or recontextualized</td>
</tr>
<tr>
<td>2. Link expression, recontextualize</td>
<td>3. Expressions emphasizing key phrases and words at different times</td>
<td></td>
</tr>
<tr>
<td>3. Longer, meaningful phrases more than 80% of the time</td>
<td>4. Consistently longer, meaningful phrases</td>
<td></td>
</tr>
<tr>
<td>4. Average with a few phrases and/or expressions</td>
<td>5. Very good</td>
<td></td>
</tr>
<tr>
<td>Accuracy Rate</td>
<td>1. 90-94%</td>
<td>2 95-98%</td>
</tr>
<tr>
<td>3. 97-99%</td>
<td>4 99-100%</td>
<td></td>
</tr>
<tr>
<td><strong>Comprehension/Strategy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictions</td>
<td>1. Logical or incorrect predictions and/or questions</td>
<td>2. 1 or 2 reasonable predictions and/or questions related to the text</td>
</tr>
<tr>
<td>3. Several reasonable predictions and/or questions related to the text</td>
<td>4. Several thought-provoking predictions and claims related to the text</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>1. 1 or 2 facts in own language and/or copied text, may include incorrect information</td>
<td>2. Partial summary, generally in own language, some important facts, may include misinterpretation</td>
</tr>
<tr>
<td>3. Adequate summary in own language, many important ideas, some details/facts and vocabulary</td>
<td>4. Adequate summary, generally in own language, many important ideas, some details/facts and vocabulary</td>
<td></td>
</tr>
<tr>
<td>Literal Comprehension</td>
<td>1. Limit information from the text and/or included incorrect information</td>
<td>2. Some information from the text, may include misinterpretation</td>
</tr>
<tr>
<td>3. Adequate information from the text that accurately responds to question(s) or prompts(s)</td>
<td>4. Adequate information from the text that accurately responds to question(s) or prompts(s)</td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>1. Little or no understanding of important text implications</td>
<td>2. Some understanding of important text implications, relevant supporting details</td>
</tr>
<tr>
<td>3. Understands implications, relevant supporting details</td>
<td>4. Thoughtful understanding of important text implications, important supporting details</td>
<td></td>
</tr>
<tr>
<td>Reflection</td>
<td>1. Insufficient or inaccurate analysis of information</td>
<td>2. Vague analysis of some important information, relevant or no statements</td>
</tr>
<tr>
<td>3. Important message or information, relevant analysis or support</td>
<td>4. Significant message or information, relevant analysis on support or opinion</td>
<td></td>
</tr>
<tr>
<td>Metacognitive Awareness</td>
<td>1. Vague explanation of the use of 1 strategy or automated response</td>
<td>2. Brief explanation of the use of 1 or more strategies, general statement(s)</td>
</tr>
<tr>
<td>3. Adequate explanation of the use of 1 or more strategies, general statement(s)</td>
<td>4. Effective explanation of the use of more than 1 strategy, examples from the text</td>
<td></td>
</tr>
<tr>
<td><strong>Score</strong></td>
<td>5 6 7 8 9 10</td>
<td>11 12 13 14</td>
</tr>
</tbody>
</table>

Add the scaled numbers in each section together to obtain scores for reading engagement, oral reading fluency, and comprehension skills/strategies.
Appendix D

LETTER OF INFORMATION for STUDENTS & PARENTS/GUARDIANS

This research is being conducted by Rachel Stewart (Master of Education, Candidate) and Grade 8 LEAP teacher under the supervision of Dr. Don A. Klinger, in the Faculty of Education at Queen’s University. This study has been approved by the Limestone District School Board and granted clearance according to the recommended principles of Canadian ethics guidelines and Queen’s policies.

What is this study about? Teachers use a variety of assessment techniques throughout the year. One of the reading assessments used widely in the United States and Canada, including the Limestone District School Board is the Developmental Reading Assessment (DRA, Beaver & Carter, 2004). The developers of the DRA conducted their field testing and analysis using mostly Primary and Junior aged children (DRA2, K-8 Technical Manual, Beaver & Carter, 2009). It hasn’t been made clear whether the DRA is an appropriate measure of reading comprehension for Intermediate students. The purpose of my study is two part: to conduct a statistical analysis of DRA results from Intermediate students to calculate test reliability; and to determine if relationships between DRA results, report cards, and EQAO results remain stable across the four different program groups of students: ATLAS, Challenge, Core, and LEAP. This information is important to verify that the correct tests are being administered for each program group at Calvin Park, and to streamline testing.

What will this study require from you? Nothing except signed consent to use your child’s data. I will be examining DRA data already collected by the classroom teachers as part of their normal teaching, and report card and EQAO data from the LDSB’s database. In this study I will not be contacting students or parents.

Is participation voluntary? You may request to have your child’s data removed by contacting me directly. There are no known risks to participation and this research will not affect your child’s academic standing. Your child’s data will be separated from his/her name as soon as all three test scores (DRA, report card, EQAO) are merged into one encrypted file. Before that time, you are welcome to withdraw your child’s scores if you wish by emailing me at 7ras1@queensu.ca, or by leaving a message with your child’s name and class at (613) 533-3028.

What will happen to the data? None of the data will contain students’ names or grade. Data will be secured in a locked cabinet and confidentiality will be protected to the extent possible. In accordance with the Queen’s University policy, the data may be used in subsequent research, and may result in publication of various types, including journal articles or other professional publications.

What if I have concerns? Any questions about the study or a request to withdraw from the study may be directed to Rachel Stewart at 7ras1@queensu.ca, or my supervisor Dr. Don Klinger at (613) 533-3028; don.klinger@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at (613) 533-6081 or chair.GREB@queensu.ca.

Thank you. Your interest in participating in this research study is greatly appreciated. Please sign below indicating your willingness to allow the use of your child’s DRA, EQAO, and report card data. Please retain the second copy of this letter for your records.
Please sign one copy of this LOI/Consent Form and return to Rachel Stewart. Retain the second copy for your records.

I have read the above statements and had any questions answered. I freely consent to my child’s data been used in this study.

Participant’s Signature: ______________________________________________________

Child’s Name :__________________________________________________________

Date: __________________________ E-mail address: _______________________________
Appendix E

Letter of Information and Consent form for Teachers

Intermediate Language Teachers: The Psychometric Properties of the DRA Across Diverse Populations

LETTER OF INFORMATION/CONSENT FORM for TEACHERS

This research is being conducted by Rachel Stewart (Master of Education, Candidate) under the supervision of Dr. Don A. Klinger, of the Faculty of Education at Queen’s University in Kingston, Ontario. This study has been approved by the LDSB and granted clearance according to the recommended principles of Canadian ethics guidelines and Queen’s policies.

What is this study about? In education, accountability, as determined by assessment results is prevalent. High stakes, large-scale assessments, such as EQAO in Ontario, have an extensive line of research regarding their reliability and the validity of the decisions made from their results. In-class assessments, such as the Developmental Reading Assessment (DRA, Beaver & Carter, 2004) do not include the same statistical analysis. The developers of the DRA conducted their field testing and analysis using mostly Primary and Junior aged children (DRA2, K-8 Technical Manual, Beaver & Carter, 2009). It is unclear whether the DRA is an appropriate measure of reading comprehension for Intermediate students. The purpose of my study is two part: to conduct a statistical analysis of DRA results from Intermediate students to calculate test reliability; and to determine if relationships between DRA results, report cards, and EQAO results remain stable across four different program groups of students.

What will this study require? In your role as a teacher, you have important insights and beliefs regarding the DRA. I would like to invite you to participate in a 30 minute focus group. With your permission, the focus group will be audio-taped. The results will be used to support my research. At no time, will these results be used in any way to monitor teacher performance.

Is participation voluntary? Your participation is completely voluntary and choosing not to participate will not result in any adverse consequences. There are no known physical, psychological, economic, or social risks associated with this study. Further, you are free to choose, without reason or consequence, to refuse to answer any questions or withdraw from the focus group at any time.

What will happen to my responses? The focus group interview recording will be transcribed after which the recording will be destroyed. Rachel Stewart and Dr. Klinger will be the only people with access to the transcript. All electronic files will be password protected. Paper and audio data will be secured in a locked cabinet. Your responses during the focus group will be kept anonymous. Comments you make during the focus group may be used to inform my work but at no time will your name or personal information be used or published. Thus I will protect your confidentiality to the extent possible. The data will be used to inform my research. Although very unlikely, I may also publish or present my findings in professional or academic journals and conferences. In accordance with the Queen’s University policy, I will maintain copies of the transcripts for a minimum of 5 years and may use the anonymous data in subsequent research.

What if I have concerns? Any questions about study participation or a request to withdraw from the study may be directed to Rachel Stewart at 7ras1@queensu.ca, or my supervisor Dr. Don Klinger at (613) 533-3028 at don.klinger@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at chair.GREB@queensu.ca or (613) 533-6081.

Thank you. Your interest in participating in this research study is greatly appreciated. Please sign below indicating your willingness to participate in this research. Please retain the second copy of this letter for your records.

I am willing to participate in the focus group for the purposes of the research, and have my comments recorded.
Please sign one copy of this LOI/Consent Form and return to Rachel Stewart. Retain the second copy for your records.

I have read the above statements and had any questions answered. I freely consent to participate in this study.

Participant’s Signature: ___________________________________________________ Date:
__________________________ E-mail address: _______________________________
Appendix F

Teacher Focus Group Questions:

1) How have you used the DRA in your classroom?

2) Does administration of the DRA pose any challenges or difficulties?

3) Are there subtasks of the DRA that you find especially useful or especially not useful?

4) Would you change any sections of the DRA? What would you do?

5) Do you feel that the results of the DRA are good value for the time required to administer it?
Appendix G

Research Ethics Board Approval

March 21, 2016

Mrs. Rachel Stewart
Master’s Student
Faculty of Education
Queen’s University
Duncan McArthur Hall
511 Union Street West
Kingston, ON, K7M 5R7

GREB Romeo #: 6012165
Title: "GEDUC-721-14 The Psychometric Properties of the DRA Across Diverse Populations"

Dear Mrs. Stewart:

The General Research Ethics Board (GREB) has reviewed and approved your request for renewal of ethics clearance for the above-named study. This renewal is valid for one year from April 4, 2016. Prior to the next renewal date you will be sent a reminder memo and the link to ROMEO to renew for another year.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period. An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours. To submit an adverse event report, access the application at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Adverse Event Form".

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes in study procedures or implementation of new aspects into the study procedures. Your request for protocol changes will be forwarded to the appropriate GREB reviewers and/or the GREB Chair. To submit an amendment form, access the application at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Request for Amendment of Approved Studies".

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Yours sincerely,

John Freeman, Ph.D.
Chair
General Research Ethics Board

c.: Dr. Don Klinger, Faculty Supervisor
Dr. Liying Cheng, Chair, GREB
Ms. Erin Wicklam, Dept. Admin.