Mathematics Beyond Computation: An Examination of Assessment Tools and Exploration of an Alternative Item Set

Sarah Angelopoulos

Mount Saint Vincent University

A thesis submitted to the Department of Education

In partial fulfillment

Of the requirements for the degree of

Master of Arts in School Psychology

September 5th, 2008

Abstract

This study examined the role of mathematical understanding in students' overall mathematical knowledge. Participants included a total of two students who were tested with a selection of items from the mathematics subtests of the Wechsler Individual Achievement Test, Second Edition (WIAT-II) and a set of comparison items (Pilot Items). Results suggested that there was a difference between the type of information collected by the two item sets and that mathematical understanding could exist in the face of poor calculation skills. Additionally, an Evaluation Rubric was developed based on the work of the National Council of Teachers of Mathematics (NCTM). Both of the mathematics subtests of the WIAT-II were analyzed using the Evaluation Rubric, as were the Pilot Items. Results suggested that the WIAT-II items were weakest in the areas of Process and Knowledge Representation whereas the Pilot Items were strongest in these areas. Findings are discussed within the context of psychoeducational assessment and mathematics learning disabilities.

Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor, Dr. Geneviève Boulet, Ph.D, Associate Professor, Faculty of Education, Mount Saint Vincent University. Geneviève has been my cheerleader, coach, and friend throughout this process. Without her knowledge, support, and guidance, this work would not have been possible.

I would also like to express my appreciation to my committee member, Dr. Nezihe Elik, Ph.D, Assistant Professor, Faculty of Education, Mount Saint Vincent University. Her assistance has helped me to make this thesis the best work that I am personally capable of producing.

I would also like to thank warmly my internship supervisor, Mr. William O'Leary, Registered Psychologist, who graciously gave of his expertise and kindness.

To Ms. Annie Beart and the staff at the Learning Disabilities Association of Nova Scotia for their dedicated assistance in finding participants to take part in this study. Thank you for your participation. I am also grateful to the children and their families who gave of their time to take part in this study.

I owe my loving thanks to my family members who have devotedly supported me throughout the process of obtaining my education. In particular, thank you to Teena for listening to all of my practice-run presentations, to Maria for all of the advice and the delicious meals, to John for believing in me, to my Mom for all of the prayers and hugs whenever I hit a roadblock, to Dad for working so hard to put me through University and for helping me to keep things in perspective. Last but not least, thank you to Keisha for the walks in the park, and to my husband Paul for your endless encouragement and love.

Table of Contents

TABLES AND FIGURES

Introduction

The essence of mathematics is not to make simple things complicated, but to make complicated things simple. - S. Gudder

Mathematics is supposed to make complicated things simpler, not the other way around. Yet commonly, mathematics is seen as complicated beyond reason and unrelated to use in daily life. One reason why mathematics is viewed in this negative light is due to the general public's understanding of mathematics as the ability to compute number facts quickly and automatically. When children struggle to learn their addition, subtraction, multiplication, and division facts, adults may assume this difficulty to indicate a lack of affinity for mathematics. With this research, I challenge the notion that computation skills are a good overall measure of a person's mathematical ability. I hope to show that computation skills are not a substitute for conceptual understanding and that *understanding* is an important variable to consider when measuring mathematical skill and knowledge.

As a budding school psychologist, I first became interested in this topic when I noticed that despite students' low mathematics scores on academic achievement measures, the focus of remediation was often placed on other areas of academic deficiencies such as reading, commonly believed by educators to be of greater importance to later success in life and of greater detriment if left untreated. Certainly, the importance of reading skills is nowhere debated in this paper. However, a case is made for the importance of mathematics for success in later life and the popular notion of mathematics understood narrowly as calculation is challenged. I also noticed that it is possible in some cases that children may be highly proficient in performing calculations, yet remain unaware of the concepts underlying any given algorithm. By accepting computational proficiency as signifying both skill and understanding without explicitly testing for understanding, educators may overestimate students' true mathematical knowledge. As such, I developed alternative test items based on the work of Ruch and colleagues (1925) aimed at accessing students' mathematical thinking and understanding above and beyond that measured by the ability to utilize mathematics facts effectively. Also, I developed an evaluation rubric based on the National Council of Teachers of Mathematics (NCTM) guidelines (2006) for evaluating large-scale assessment tools by which to examine tools currently used to assess students' mathematical skills.

It is my hope that by assessing critically the tools currently used to measure mathematics and by testing the effectiveness of an alternative item set aimed at measuring mathematical understanding, educational professionals will become more aware of the limitations of mathematics achievement measures and will have a better idea of the types of questions that might do a better job of assessing mathematical understanding. It is my intention that test questions similar to the alternative item set, with a focus on mathematical thinking and understanding, will replace currently used test items that rely heavily on computation to provide an overall measure of mathematical ability, or those which account only for the correctness of the final response, given that knowledge is rarely an all or nothing phenomenon (Carpenter & Leher, 1999; Hiebert & Carpenter, 1992). Therefore, assessments, which rely only on the correctness or incorrectness of student responses, could not capture accurately the full extent of student knowledge and it was hypothesized that the alternative test items would provide a more precise sampling of students' true mathematical skills, knowledge, and understanding.

Chapter 1: The Problem

In 2000, the National Council of Teachers of Mathematics (NCTM) defined the ideal mathematical disposition to be one in which students see mathematics as a sensible activity, a useful tool, and a worthwhile pursuit. Although children may enter school with foundational skills and knowledge in mathematics, school mathematics is often approached as an exercise in the memorization of facts or the application of procedures (Romberg & Kaput, 1999). Unsurprisingly, many children grow to dislike mathematics and see it as disconnected from their everyday experiences (Hiebert, 1984). Some students may perform quite well in mathematics with little understanding of the concepts underlying the rules and procedures they apply (Tobias, 1994). It seems that many people never move beyond superficial experiences with mathematics to develop the ideal disposition as described by the NCTM (2000).

Mathematics, in its real sense, is often misunderstood. The scope of mathematics extends beyond counting and arithmetic, yet, historically, mathematics has been taught in terms of the application and memorization of rules and procedures (Hiebert, 1984; Romberg & Kaput, 1999; Small, 1990). In terms of assessment, measuring only the correctness of the student’s response does not necessarily reveal either knowledge or a lack of knowledge (Cramer & Wyberg, 2007). For example, Cramer and Wyberg (2007) discuss a student who is able to correctly compute the sum of 2/3 + ¼, however, is unable to estimate whether or not the sum would be greater than ½. This example demonstrates the importance of assessing students’ thinking and understanding for capturing a better sample of their knowledge, beyond simply their ability to perform computations. Psycho-educational assessments rarely, if ever, award points for process over correctness of the final answer. For example, within the mathematics subtests of the Wechsler Individual Achievement Test, Second Edition (WIAT-II), responses are scored as either correct or incorrect for which the student receives either one or zero points respectively. Indeed, the standardized nature of many psycho-educational batteries prevents the examiner from querying students at the time of the response as to their thought processes. Although experienced examiners may develop a deeper understanding of mathematics and, upon the completion of a subtest, invite the examinee to elaborate on his or her rationale for a particular response, such an examination procedure is rarely built into assessment tools, despite the quality of information it would derive. This process is also known within the field as testing-of-limits (Sattler, 2001). However, there is no guarantee that even examiners with years of experience ever implement this type of additional assessment. Correctness of the final response is generally presumed to be an indicator of the child's understanding; however, without explicitly assessing students' mathematical thinking and understanding of a given problem, it is difficult to get an accurate idea of their true mathematical knowledge and ability. This is a problem, because academic achievement measures are often used as a primary source of information when determining the presence or absence of a specific learning disability in mathematics, or in other academic areas. Because of the powerful decisions tied to assessment measures, it is imperative that researchers and other professionals continually question, refine, and revise these tools. Since some members of the general public view failure in mathematics as more acceptable than in other academic areas (e.g. the view that people don't use the skills they learn in math class and always have the option of using a calculator; McCloskey, 2007) it is necessary to consider what constitutes mathematical skill, and to examine the components that contribute to successful performance in mathematics. These factors should then be related to psycho-educational assessment batteries as these measures are often relied upon heavily to provide information about an individual student's performance compared with that of a national normative sample. One purpose of these assessments is to provide professionals working with the student with more knowledge of the reasons why a particular student may be struggling with mathematics. In an effort to ensure that students' true mathematical knowledge is accurately and sufficiently captured, the present research project is undertaken.

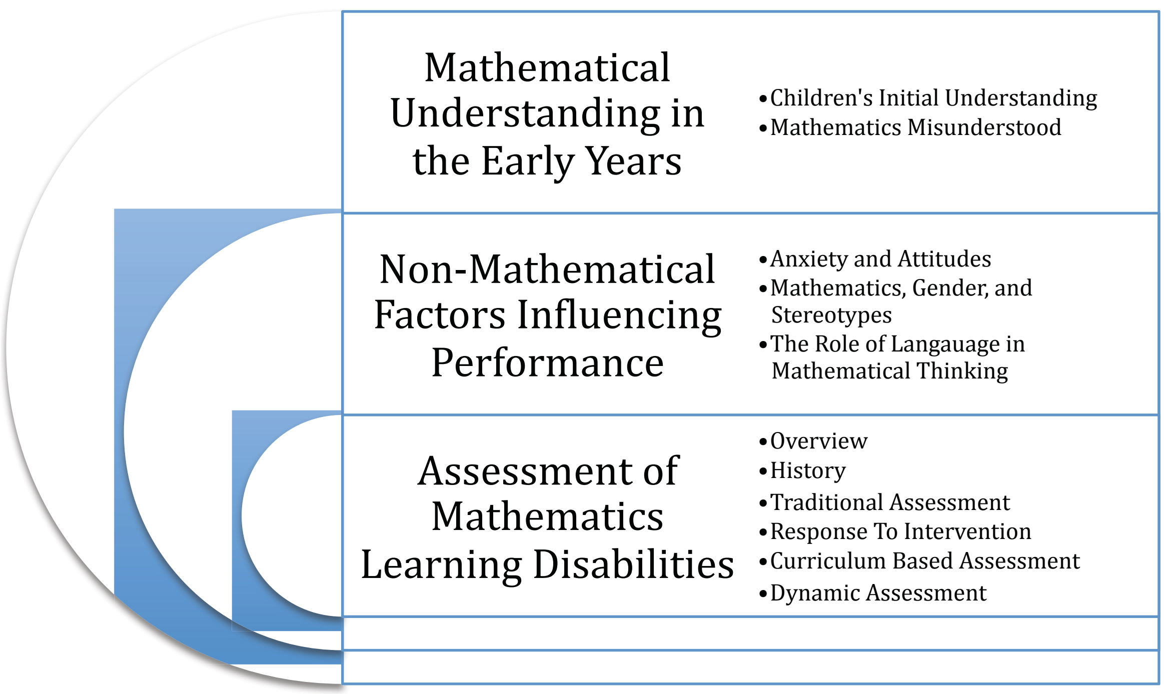
Chapter 2: Literature Review

Introduction

Recently, the importance of mathematical skill and understanding for later academic achievement has been shown with kindergarten-aged children. Clements and Sarama (2008) found that early mathematical knowledge was a strong predictor of later proficiency in mathematics as well as in reading. In addition to knowledge of mathematics, other factors such as anxiety and self-efficacy have been discussed in the literature. In 2001, the National Research Council's description of the target mathematical disposition included elements in which students see mathematics as having use in one's life, of practical importance, a valuable pursuit, and a belief in one's ability to succeed (Kilpatrick, Swafford, & Findell, 2001). There are many factors that affect successful performance in mathematics, many of which I review in the following sections. Starting with the early years, I first discuss students' initial mathematical skills and abilities and consider the transformation of attitude that often occurs in relation to the learning of mathematics in today's classroom. Next, I question a commonly accepted conception of mathematical proficiency and discuss an alternative conceptualization of mathematical thinking, skill, and understanding. The focus of this alternative conceptualization relates more to students' mathematical thinking and understanding that is common in earlier years. For example, the work of Joan Moss and Robbie Case (1999) has demonstrated that by shifting the focus from the teaching of rote application of procedures to using children's existing knowledge bases to build new knowledge that deeper conceptual understanding can be developed.

The role of emotions in learning is so strong that some researchers have posed the question of whether mathematics anxiety is a particular sort of mathematical learning disability (Ashcraft, Krause, & Hopko, 2007). Clearly, the impact of anxiety on mathematics is in need of review in order to gain an appreciation of today's students in mathematics classrooms. Additionally, some research has probed into the effect that gender and stereotypes may have on mathematical learning and performance (e.g. Ben-Zeev, Duncan, & Forbes, 2005; Fennema, 1989; Royer & Walles, 2007). Although less researched, these qualitative factors impact mathematical performance and are in need of consideration in order to fully conceptualize and understand the range of demands that students encounter in the mathematics classroom. Further, since language is inextricably involved in the teaching, learning, and application of mathematics, the impact of language on mathematics cannot be ignored (Boulet, 2007). The role that language plays in the scope of mathematics is in need of consideration when studying mathematics from a broad perspective of factors that may influence performance. Of additional relevance, is the manner in which each of the above mentioned factors (i.e. mathematics education in the classroom, cognition, anxiety, gender/stereotypes, language) might combine to produce specific learning disabilities or difficulties in mathematics. Since assessment tools cannot hope to capture all of the above mentioned factors influencing mathematical performance, educational professionals must possess knowledge of these variables, have the ability to combine it with the data collected through assessment tools, and put it all together to create a coherent picture of the individual student's mathematics learning profile. Psychoeducational assessment batteries play a key role in providing information regarding a student's academic skill set. As such, in the rest of this chapter I review the history of mathematics achievement testing, current assessment models, and up and coming frameworks. As a whole, I hope that the factors chosen for review present the reader with an overview of the variables involved in the successful learning and performance in mathematics and provide a working knowledge of the factors involved in the assessment of mathematical difficulties and disabilities. Finally, I discuss an alternative item set that I believe do a better job of accurately accounting for student's mathematical skills, knowledge, and understanding. Figure 1 presents a pictorial overview of the information to be presented within the current chapter.

Figure 1.
A Pictorial Overview of Topics in the Literature Review



Mathematical Understanding in the Early Years

Children's Initial Understanding

Most children enter school with some informal mathematical understanding and skill (Hiebert, 1984). They demonstrate this knowledge with their ability to reason analytically and to solve real-life, practical problems. For many children, an unfortunate shift occurs in the way mathematical problems are approached. In school, mathematics is often taught in terms of symbols, procedures, and rules. As a result, many children abandon their authentic, innate understanding of mathematics and adopt a superficial approach to solving problems (Devlin, 2000). They no longer see connections between what is learned in the classroom and what exists in their everyday lives. In the context of the present study, it is important that the reader be aware of the changes from early mathematical understanding to one of surface experiences with mathematics that can happen. Also of importance is how this shift impacts assessments of students' mathematical knowledge. Since some assessment measures may not accurately capture the full extent of students' understanding, the reader is invited to join the researcher on a search for the type of assessment questions that may give students greater opportunities to accurately portray their mathematical knowledge and understanding.

In the classroom, an emphasis is often placed on the exactness of procedures and outcomes rather than on process and understanding (Tobias, 1994). When students believe that there is only one correct solution to a problem, or only one correct method of arriving at a solution, they are less likely to use creative or intuitive approaches to problem solving. It seems that the formal manner in which mathematics is traditionally taught moves children away from their initial and more genuine understanding of mathematics instead of guiding them to higher-level thinking (Hiebert, 1984). Consequently, when a student makes a distinction between school-math and real-life math, school-math often becomes the meaningless exercise of recalling facts or applying specific procedures.

William Butler Yeats (n.d.) eloquently stated that, “education is not the filling of a pail, but the lighting of a fire.” Similarly, true mathematical learning is not the recalling of facts or the memorization of procedures. We seek not the type of learning that is briefly committed to memory, only to decay and disappear following an exam, but rather, the type that enhances one’s mind, helps one to think more clearly, or develops some skill. If schools are satisfied with the type of learning that stops at memorization, then they condemn the children to a superficial type of knowing without understanding; surely, this cannot be called true learning. By understanding the true nature of mathematics, which entails knowing that its scope extends far beyond calculation, mathematics can become a meaningful tool that has the potential to assist students in a wide variety of endeavours.

In its purest sense, mathematics is “a philosophy that simultaneously stresses erudition and common sense, integration through application, and innovation through creativity. Most important, it stresses the creation of knowledge. Against this broad and ambitious view of mathematics, traditional school mathematics appears thin, lifeless, and isolated” (Romberg & Kaput, 1999; p.7). In school, mathematics is frequently taught in a manner at odds with this description. In place of common sense and creativity, there is rote memorization and timed drills; instead of integration and application, there is little real-world context by which to make sense of concepts (Brady & Bowd, 2005). Also both common and problematic in the mathematics classroom is the insufficiently explained mathematical language, teachers' tendency to skim over concepts, and the sequential nature of the mathematics curriculum, which all too often leaves struggling children behind (Brady & Bowd, 2005). It seems there is much work to be done to rectify the manner in which mathematics is taught in schools. Indeed, the National Council of Teachers of Mathematics (NCTM) devotes a full chapter in their 69th Yearbook to describing the phases of mathematics reforms that have taken place, from the drill and practice phase of the 1920's, to the standards, assessment, and accountability phase from the 1990's into the present (Lambdin & Walcott, 2007). Many of the phases of previous generations can still be seen in today's mathematics classrooms.

Commonly, students appear to understand a mathematical concept, yet when a small change is made, they become lost (Moss & Case, 1999). An essential feature of successful mathematical learning is the creation of knowledge relationships, or an understanding of the connections among mathematical concepts (Romberg & Kaput, 1999). When children are unsuccessful at making these connections, mathematical skills are viewed in isolation and related very little, if at all, to everyday life (Carpenter & Leher, 1999). An awareness of interrelationships among mathematical concepts aids in one's ability to solve novel problems. Since it is impossible for teachers to prepare students for every potential problem they may encounter, it is important that students be taught from an early age how to integrate their formal and informal mathematical knowledge.

Mathematics, like other subjects taught in school, is meant to assist students in their everyday lives. The intention is not for it to be an isolated academic pursuit that makes sense only in the context of a classroom. The manner in which mathematics is taught from an early age is clearly one of great importance since it has implications that extend beyond mathematics and includes literacy skills, despite the fact that the two subjects are seemingly unrelated (Clements & Samara, 2008). Generalization of knowledge, as well as diverse yet coherent conceptualizations of what constitutes mathematical success are of paramount importance.

Mathematics Misunderstood

The general public tends to view mathematics differently from the way mathematicians view mathematics. Commonly, mathematics is seen as synonymous with arithmetic (Devlin, 2000). In fact, arithmetic is a low-level branch of mathematics and a person's computational skills does not necessarily indicate mathematical ability or understanding (Devlin, 2000). Often, a student who is able to perform algorithms automatically is seen as mathematically proficient; however, a focus on exactness of the answer does not ensure that the student understands the underlying concepts and can become a replacement for thinking (Tobias, 1994). In the context of the present study, the conceptualization of mathematical proficiency is of major importance. It was hypothesized that the conceptualization of mathematical proficiency as restricted to symbols, numbers, calculation, and algorithms is far too narrow and the usefulness of this idea of mathematics was analyzed.

In contrast to the general public, mathematicians view mathematics as an ability to think, to reason, and to problem solve using mathematics as a tool to assist, not replace, one's thinking (Devlin, 2000). Romberg and Kaput (1999) describe the aim of mathematics education as, "…teaching students to use mathematics to build and communicate ideas, to use it as a powerful analytic and problem-solving tool, and to be fascinated by the patterns it embodies and exposes" (p.16). If mathematics is to be understood broadly and in its truest sense as Romberg and Kaput (1999) describe it, there must be movement away from the memorization and drill techniques that are too frequently used in the elementary school mathematics classroom. For example, Fennema and colleagues (1993) provided a workshop to teachers where they presented an alternative framework for teaching addition and subtraction. Some features of the new framework included building on students' existing knowledge, using real-life problems that were meaningful to the students, and teaching strategies for solving number facts, rather than rote memorization. The researchers discovered that compared to students whose teachers had not attended the workshop, those in the experimental group showed significant gains in number fact knowledge, reported increases in confidence and understanding, and a significant rise in solving word problems. It is interesting to note that these gains were produced, despite the fact that the experimental group teachers spent *less* time in a different type of instruction than control group teachers (Fennema et al., 1993).

Mathematics is not synonymous with arithmetic (i.e. numerical operations). Also, arithmetical proficiency is not necessarily indicative of mathematical understanding. If arithmetical skill does not necessarily indicate mathematical understanding, therefore, a lack of arithmetical proficiency is not necessarily indicative of a *lack* of mathematical understanding. A more accurate conceptualization of mathematical skill is one which accounts for students' skill at using mathematical procedures to assist in *thinking* through a real problem. It is the difference between using an algorithm to assist one in solving a real life problem and absent-mindedly computing a series of arithmetic problems on a worksheet in a classroom. In the former, the mind is engaged, the answer must seem reasonable, and the algorithm is a supporting player, rather than occupying the lead role.

Non-Mathematical Factors Influencing Performance

Mathematics Anxiety and Attitudes

Emotions are known to influence learning (Tobias, 1994). Mathematics anxiety is a negative emotion experienced by some in the mathematics classroom and in real-world contexts and has been identified as a factor known to influence students' performance in situations where mathematical problem solving or calculation must occur (Ashcraft et al., 2007). Due to the impact that mathematics anxiety can have on learning and performance in the area of mathematics, it is in need of consideration. In her book, Tobias (1994) describes a feeling of "sudden death", referring to a belief that one's personal capacity for mathematical understanding has been reached. A common impression exists among lay people, perhaps more than in other academic areas, that some people are good at mathematics while others are not. Some students experience "sudden death" within the first few years of elementary school, prior to studying branches of mathematics that require less monotony as can be the case with the learning of computation. Unfortunately, students may develop false beliefs regarding their aptitude for mathematics prior to encountering mathematics of some substance.

The presence of negative or anxious feelings can make concentration, as well as learning, more difficult (Tobias, 1994).

Past research has established some negative outcomes that mathematics anxiety has on both performance and achievement, including fewer courses enrolled in by the mathematically anxious and lower grades in mathematics courses (Ashcraft & Kirk, 2001). Fennema's (1989) Autonomous Learning Behaviour model describes attitudes as affecting opportunity to develop mathematical mastery. Effectively, the highly mathematically anxious have fewer opportunities to practice their skills, to experience success in mathematics, and this in turn negatively impacts subsequent performance and mathematics self-confidence. In contrast, those with lower mathematics anxiety and more positive attitudes tend to seek out mathematical experiences, thereby increasing their exposure to mathematics, their opportunity to practice, and to experience success (Ashcraft & Kirk, 2001; Fennema, 1989).

Until recently, relatively little was known about the cognitive mechanisms of mathematics anxiety. Ashcraft and Kirk (2001) examined the effect of mathematics anxiety on working memory and performance, based on Eysenck and Calvo's (1992) Processing Efficiency Theory. Ashcraft and Kirk (2001) hypothesized that the intrusive thoughts and worry characteristic of anxiety compete for the limited resources of the working memory system, resulting in poorer performance. The effect of these additional demands on working memory tends to be slower performance and reduced accuracy. Based on a series of experiments testing high and low mathematics anxiety in a variety of situations, Ashcraft and Kirk (2001) add to Fennema's (1989) model of mathematics anxiety and draw a new conclusion. Rather than suggesting poor working memory and a lack of basic competence to explain the differing levels of anxiety, working memory is suggested to be temporarily disrupted by attention to intrusive thoughts and worry. Furthermore, this disruption is hypothesized to be present not only in testing situations, but also in mathematical learning situations in general (Ashcraft & Kirk, 2001). Figures 2 and 3 present diagrams of Fennema's (1989) model and the model proposed by Ashcraft and Kirk (2001).

Figure 2.
The Effect of Mathematical Attitudes and Anxiety on Performance (Fennema, 1989)

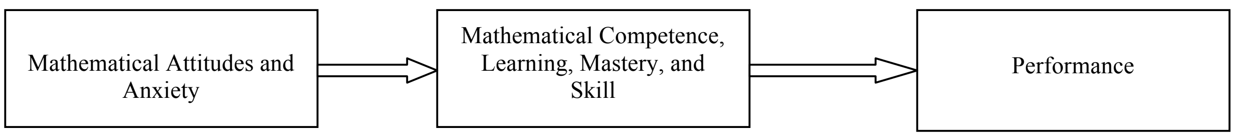
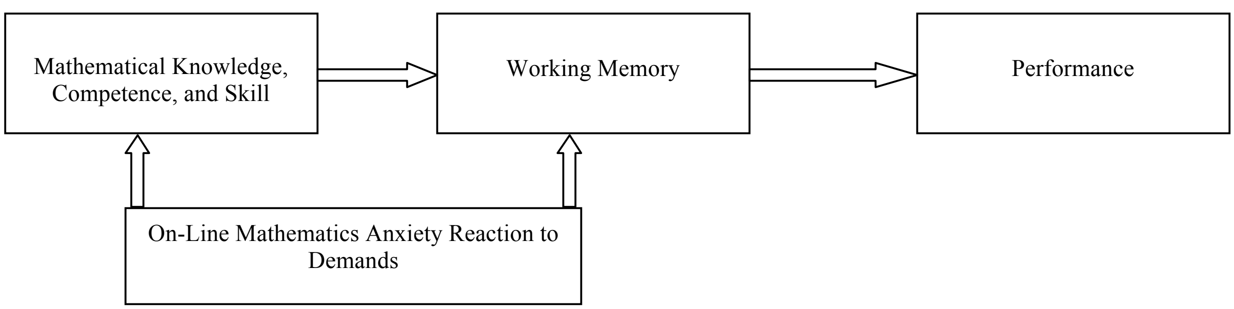


Figure 3.
The Effect of Anxiety on Working Memory and Subsequent Mathematical Learning and Performance (Ashcraft & Kirk, 2001)



One cause of mathematics anxiety include environmental factors, such as a highly math anxious teacher (Trujillo & Hadfield, 1999). From a variety of case studies, pre-service teachers identified as highly mathematically anxious recounted the following types of encounters with their own mathematics teachers: student perception of teacher as nervous or insecure, teacher unwillingness to answer student questions, little explanations provided by the teacher, teacher-focused rather than student focused lessons, little discussion, and lecture followed by independent seat work format (Trujillo & Hadfield, 1999). Interestingly, many of these factors mirror those described by Kersaint (2007) in her discussion of classroom factors characteristic of low performing mathematics programs. Mathematically anxious teachers may unwittingly convey their anxious feelings toward mathematics to their students (Widmer & Chavez, 1982). Past mathematical experiences are cumulative and have an impact on the way one approaches the subject as a teacher and educators teach mathematics in accordance with their personal experiences with the subject (Brady & Bowd, 2005). In some cases, teachers do not feel adequately prepared to teach mathematics, even at a low grade-level (Carpenter & Leher, 1999). Ashcraft (2002) suggests that students whose teachers provide little cognitive or motivational support to students during mathematics lessons (e.g. place a high value on exactness of procedure and outcome, little opportunity for discussion) are at-risk for developing math anxiety.

Royer and Walles (2007) point out that many of the early work on math anxiety continues to be replicated in research on math anxiety today. Royer and Walles (2007) state that one important factor is the finding that mathematics anxiety interferes with both the learning and application of mathematics. The potential impact of mathematics anxiety is long-term and may affect students' career choices (Fennema, 1989). Two models have been suggested to account for the negative impact that mathematics anxiety can have. One, in which the individual attempts to avoid the negative experience of performing poorly and thereby has fewer mathematical experiences upon which to build knowledge (Fennema, 1989), and the other in which working memory is disrupted such that students have fewer cognitive resources by which to learn and execute mathematical thinking and reasoning processes (Ashcraft & Kirk, 2001). Additionally, students may be at risk of developing feelings of mathematics anxiety through secondary sources, such as unresolved teacher feelings of mathematics anxiety being passed on to students (Ashcraft, 2002). The negative impact of mathematics anxiety can be seen from early experiences in the classroom and may continue into adult life.

Mathematics, Gender, and Stereotypes

Students' attitudes toward mathematics and subsequent performance also can be shaped by cultural and gender stereotypes. Research has mostly focused on identifying the cognitive aspects that contribute to learning problems in mathematics, however, factors such as stereotype threat have been found to influence performance in mathematics (Royer &Walles, 2007). In the context of the present study, the impact of gender and stereotype threat was given consideration in an effort to examine critically the variables known to influence performance in mathematics and to include those factors that have historically been given less attention in the research. The effect of gender and stereotype threat should be of prime interest to all educational professionals given the increasing degree of diversity in classrooms across North America today (Meacham, McClellan, Pearse, & Greene, 2003).

Researchers have not found evidence to support the view that one sex is more likely than the other to suffer from a Mathematics Learning Disability, or MD (Shalev, Manor, & Gross-Tsur, 2005). Despite this finding, there are a number of stereotypes that exist regarding mathematics and gender, and disheartening statistics abound. Ben-Zeev and colleagues (2005) reported that females are less likely to participate in mathematical activities, to take higher-level mathematics courses, and are less likely to pursue an undergraduate degree in mathematics or to seek employment in a mathematics-related field. It has also been reported that men outperform women on standardized measures of mathematics, such as the GRE (Brown & Josephs, 1999). These research findings suggest a very real disadvantage for students from historically underpriviledged groups, the effect of which may be long lasting. Some researchers (e.g. Hyde, Fennema, Ryan, & Frost, 1990; Tobias, 1990) describe the impact that social variables can have on the pursuit of higher-level mathematics.

Hyde and colleagues (1990) reports that some variation in male-female entrance to mathematics professions is due to a lack of risk-taking attitudes among females. Risk-taking, as related to mathematical achievement, is dependent upon feelings of personal control and choosing to engage in high-level mathematical tasks (Hyde et al., 1990). Drawing from attribution theory, Tobias (1990) reports that girls are more likely to attribute success in mathematics to effort, whereas boys are more likely to attribute success to ability. It seems that girls who attribute their mathematical success to effort, rather than to ability, may be less likely to persist in mathematics, perhaps due to the hierarchical nature of mathematics education and a belief that eventual "sudden death" would be certain. Fullilove and Treisman (1990) report that isolation is another factor that may contribute to success in mathematics. When females feel isolated in the development of their mathematical thoughts, skills, and understanding, they may be less likely to pursue mathematics (Fennema, 1989). Although there are no biological determinants favouring mathematical success in one sex over the other, it seems that the existence of factors affecting performance in mathematics are in need of consideration when studying mathematics achievement.

Females and other historically underprivileged groups may suffer from stereotype threat in relation to mathematical performance (Ben-Zeev et al., 2005). Stereotype threat is defined as the "extra pressures that can affect the test performance and academic identities of such groups as African Americans and women in math" (Steele, 1998; p.681). Stereotype threat need not be overt to have an effect. High-achieving females' performance on a standardized achievement test was found to decline proportionately to the ratio of males to females in the room; the more isolated the females, the lower their scores were (Inzlicht & Ben-Zeev, 2000). This research shows the delicate nature of stereotypes. It is important to note that the personal meaning of the stereotype is not uniform across all disadvantaged groups (Ben-Zeev et al., 2005). This means that stereotyped individuals are not a homogeneous group and that not all members will respond in the same manner.

Steele and colleagues (1997) showed that performance could be affected by priming individuals with either a facilitating or a hindering message (e.g. by announcing that females tend to perform lower on standardized tests than males). This information is crucial for classroom teachers, who need to be sensitive to the additional pressure that young girls and students from minority groups may be experiencing during mathematics lessons and activities. For students who have internalized a negative self-stereotype regarding expected performance in mathematics, these students are at risk for actually *achieving* lower in mathematics through lower mathematics performance expectancies (Bonnot & Croizet, 2007). This means that if teachers expect less from their stereotyped students, many of those students will live up to their teacher's low expectations (Steele, 1997). This research ties in well with research on the self-fulfilling prophecy hypothesis in which teacher expectations account for more variance in changes in student achievement than prior achievement or motivation (Jussim & Eccles, 1992). It seems that teachers have a tremendous power to influence the state anxiety of their students. By priming students with a message espousing the equal mathematical abilities of girls and boys, teachers may alleviate unnecessary anxiety in their students (Ben-Zeev et al., 2005; Steele, 1997).

For some, the presence of stereotype threat results in anxiety symptoms (Ben-Zeev et al., 2005), such as higher levels of physiological arousal, such as heart palpitations, perspiration or rapid breathing (Ashcraft, 2002). Yerkes-Dodson law states that performance is at its peak when levels of arousal are moderate, and performance decreases when arousal is either too high or too low (Keller, 2007). Inzlicht and Ben-Zeev (2000) have suggested that anxiety and its accompanying physiological symptoms may rise when an individual is faced with a stereotype threat. For example, stating that males tend to outperform females on a particular mathematics test would likely increase anxiety symptoms in some females beyond optimal levels and negatively affects their performance. This theory of how stereotype threat affects performance also ties in well with Ashcraft and Kirk's (2001) theory of how the worry and negative thoughts associated with anxiety interfere with optimal mathematical performance. The good news is that these factors influencing the mathematics performance of females and other stereotyped individuals are malleable; therefore, changing the negative features of mathematical learning environments can positively impact performance (Ben-Zeev et al., 2005).

Today's classrooms are more diverse than ever before. Some students may experience disadvantages in the form of negative beliefs regarding their abilities. Knowledge of the effect of stereotype threat is empowering to teachers because they have the power to positively influence students who may be suffering from stereotype threat. Failure to do so can have a lasting negative impact for students from disadvantaged groups. Another factor for teachers to be mindful of in their mathematics classrooms is the type of language being used to present and discuss mathematical ideas (Boulet, 2007; Chapin & O'Connor, 2007).

The Role of Language in Mathematical Thinking

The use of language to discuss mathematical concepts is closely linked to students' understanding (Chapin & O'Connor, 2007). For the purposes of the present study, review of the literature relating to the role that language plays in mathematical learning is necessary in order to consider the impact of language on the learning of mathematics, to evaluate currently employed assessment tools, and to aid in the construction of alternative test items.

In Western culture, language and number systems are so closely tied together that it is difficult to tease them apart. Vaidya (2004) states that the conceptual understanding of numerosity is bound to the language system in that symbolic representations are accessed in the mind through language. In a study to determine infants' understanding of number as distinct from language, Wynn (1992) found that infants as young as three days old are able to discriminate among numerosities ranging between one and three, but no greater. Citing the work of Starkey (1992) and Wynn (1992), Case and colleagues (1999) describe the quantitative capacities of infants as being "…born with a natural sensitivity to number and can distinguish actions that preserve the numerical value of small arrays from those that change it" (Case, Griffin, & Kelly, 1999, p.135) By the age of four, most children develop a uniquely human understanding of the concepts that adding gives more, taking away leaves less, and that no action leaves the same quantity (Case, 1999; Starkey, 1992). The development of hypotheses of separate number modules, one involving an approximate system (e.g. more, less, same) and the other involving a precise system (e.g. one, two, three, four, five) followed. The precise system for processing quantities has been suggested to be more reliant on language than the approximate system (Chiappe, 2005; Gelman & Butterworth, 2005). It has been suggested that it is the development of *linguistic* number systems which allows human beings to work with numbers greater than four or five (Carey, 2001). Chiappe (2005) states that there is consensus within the field that the analog (approximate) system does not handle quantities greater than three. To test these ideas, Pica and colleagues (2004) conducted studies of numerical concepts and abilities with adults from cultures where linguistic number words were less precise (e.g. one, two, few, many). Two important findings were uncovered with these studies. First, despite the fact that participants utilized an approximate number system, the participants were able to engage in number tasks involving integers as great as 80. Secondly, participants quickly developed counting rules when placed in situations where they were paid for their labour, suggesting that the underlying concepts for a precise number system were already present but required more precise language in order to work efficiently with them. Gelman and Butterworth (2005) insightfully call to attention that it is difficult to imagine how such rules could be developed so quickly if the underlying concepts had not already been present. To a certain point, the role of language seems unnecessary in the development of numerical concepts, however, without language, mathematical thinking clearly remains basic (Gelman & Butterworth, 2005). Thus, mathematical language can be of critical importance in students' learning (NCTM, 2000). It is not any kind of language, but rather, mathematical language, which is a powerful tool that can be used to help students learn complex material (Chapin & O'Connor, 2007).

Mathematical language can be quite complex and some students struggle to understand it. Unlike reading connected text, the meaning of a mathematical symbol cannot easily be inferred by context (Pedrotty Bryant, 2005a). For example, to decode the mathematical phrase (2 + 4) x (6 + 3), students must possess prior knowledge of each of the number and operation symbols, whereas one may be able to infer the meaning of an unknown word based on context clues in a literary text. However, Boulet (2007) states that, "reading mathematical texts should be similar to reading ordinary texts, in that the reader must transcend the code and comprehend the text's meaning" (p.6). This requires that the students and the teacher have a good grasp of numerical symbols, operations, and concepts. One difficulty arises when educators use confusing or ambiguous language to discuss mathematical ideas (Barwell, 2005). In her article, Boulet provides examples of students who correctly apply mathematical procedures, but who have very little understanding of the work they complete because they lack conceptual knowledge. Particularly memorable, was Maxwell who wondered whether the phrase, "goes into", in reference to a long division problem was similar to a numbers car crash (Boulet, 2007). This example brings to light the way the language used to discuss mathematics can either elucidate or conceal the underlying meaning. Algorithms are not taught to frustrate students or to test their memory skills. The purpose is that the algorithms assist students in solving real life problems. Language is used to provide a meaning and purpose for mathematics. As Boulet (2007) suggests, in Maxwell's case, the teacher could have provided a real-world example along with clearer and more meaningful language in order to help him make sense of the problem. Language should provide the opportunity to make sense of algorithms, rather than computing them mechanically like robots. As Kaput (1988) stated, "the language of mathematics is both a means of communication and an instrument of thought" (as cited in Boulet, 2007; p.4).

Language is required to facilitate the learning of mathematics with quantities greater than four or five (Carey, 2001). Moreover, the use of language to discuss mathematical procedures and concepts can either hinder or facilitate students' understanding. As such, educators need to carefully scrutinize their use of language to determine whether they are elucidating or concealing mathematical concepts with their words.

Assessment of Mathematics Learning Disabilities

Overview of Mathematics Learning Disabilities

A learning disability in mathematics has many names. Often, the terms dyscalculia, mathematics disorder, and mathematics learning disability (MD) are used interchangeably (Fletcher, Lyon, Fuchs, & Barnes, 2007). The use of a variety of terms to discuss a single construct contributes to confusion in understanding the nature of learning problems in mathematics (Berch & Mazzocco, 2007; Fletcher et al., 2007). Dyscalculia is an older, less-used term and refers to a person who has an inability to perform calculations (Pedrotty Bryant, 2005a). Practitioners have long relied on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) criteria to assist them in making important diagnosis decisions (APA, 2000). (See Table 1 for the DMS-IV-TR diagnostic criteria relating to MD).

Table 1.
DSM-IV-TR Diagnostic Criteria for Mathematics Learning Disability

A. Mathematical ability, as measured by individually administered standardized tests, is substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education.
B. The disturbance in Criterion A significantly interferes with academic achievement or activities of daily living that require mathematical ability.
C. If a sensory deficit is present, the difficulties in mathematical ability are in excess of those usually associated with it. Coding note: If a general medical (e.g., neurological) condition or sensory deficit is present, code the condition on Axis III.

In Canada, there is not a federal definition of a Learning Disability (LD), since educational legislation and policy falls under the jurisdiction of each individual province and territory (Kozey & Siegel, 2008). Kozey and Siegel (2008) state that the trend among the provinces and territories has been the adoption of features of the Learning Disabilities Association of Canada's (LDAC) definition of LD. Within Nova Scotia's (1996) Special Education Policy Manual (amended in October of 2007), it is made clear that each school board has a responsibility to develop a process by which at-risk students are identified, and assessed. Many Canadian provinces and territories, including Nova Scotia, maintain that a discrepancy between IQ and achievement is a key feature of LD, despite that the discrepancy definition has largely been discredited by recent research and is no longer considered a best practice for the identification of LD (Kozey & Siegel, 2008; National Association of School Psychologists, 2003) which will be reviewed in later sections. One unique feature within Nova Scotia's special education documents is the inclusion of a Response to Intervention (RTI) component for the identification of LD (Kozey & Siegel, 2008; discussed in more detail in a following section), which is an alternative to the traditional discrepancy model.

For the purposes of this paper, the terms mathematics learning disability and MD (for short) will be used to refer to cognitive deficits inherent in children of average intelligence who experience significant, specific, and pervasive difficulty in mathematics, where said difficulty cannot be better explained by hearing and/or vision problems, socio-economic factors, cultural or linguistic differences, lack of motivation or ineffective teaching, although these factors may further complicate the challenges faced by individuals with MD (Learning Disabilities Association of Nova Scotia, n.d). Approximately 5-8% of the population has some form of cognitive deficit that interferes with the learning of mathematics which may be sufficient to result in a diagnosis of MD (Geary, 2004). MD may or may not be experienced in combination with another disorder, such as Attention Deficit Hyperactivity Disorder (ADHD) or dyslexia (Pedrotty Bryant, 2005b). There is some evidence that MD has a genetic component. Children are ten times more likely to be diagnosed if one or both parents also carry a diagnosis of MD (Fletcher et al., 2007; Geary, 2004). In addition, studies involving twins have shown a substantial genetic factor for MD, especially for monozygotic (i.e. identical) twins, and less of an environmental factor (Fletcher et al., 2007).

Assessment tools, such as the Wechsler and Woodcock scales, are used to provide additional information about a student's cognitive and academic strengths and weaknesses. If assessment tools focus too narrowly on students' calculation proficiency and their skill at applying mathematical procedures, without specifically examining their understanding of the problem, then it remains unknown whether or not students truly understand the procedures they apply. Since the information garnered from assessment tools are often given prime consideration for determination of a learning disability, it is necessary to examine what has been written about MD. Following, this information was related to the assessment tools used to inform educational professionals of a student's strengths and weaknesses in order to see whether or not the tools actually assess the disorder in a valid way.

Some studies have suggested that working memory problems and severe language deficits underlie mathematics disabilities (Ackerman & Dykman, 1995; Fletcher et al., 2002; Geary, Hoard, & Hamson, 1999). However, the conceptualization of mathematics disabilities has typically been in terms of numbers and arithmetical skill; the belief being that proficiency in computation is necessary to master prior to advancing in mathematics (Fletcher et al., 2002). As noted in previous sections, mathematics has a much broader scope than computation alone (Fletcher et al., 2007). The way to define MD is crucial, because the working definition has a strong impact on the research, assessment, and intervention strategies pertaining to MD (Fletcher et al., 2002). Even the name "dyscalculia", meaning an inability to perform mathematical calculations as a result of brain disorder (McKean, 2005), provides insight into the way learning problems in mathematics have been defined.

Unlike the field of reading disabilities, in which phonological processing has been identified as a core deficit and predictive of later reading achievement (Torgesen, 2002), mathematics researchers have yet to identify the underlying cognitive structures that support success in higher-level mathematics, such as algebra (Geary, 2005). Difficulties in mathematics are considerably less understood than those in the reading disabilities field (Pedrotty Bryant, 2005b). Mathematics has grown in the last hundred years to include between 60 and 70 distinct branches (Devlin, 2000). Due to the vast span within the mathematics field, it is unlikely that a single cognitive deficit could explain mathematics disabilities as a whole (Chiappe, 2005; Fletcher et al., 2007).

Much research on MD has focused on the roles of working memory, automaticity, and counting (Geary & Brown, 1991; Geary & Hoard, 2001). Concentrating on automaticity and counting is troublesome because it relies on studies of low-level mathematical skills (i.e. arithmetic) to inform researchers of the nature of mathematics disabilities in general. Fletcher and colleagues (2007) suggest that due to the broad scope of the field of mathematics, researchers should specify the area of mathematics being studied, as the underlying cognitive correlates likely vary depending on which area is under review (e.g. arithmetic, mathematical reasoning). Researchers within the field (e.g. Fletcher and colleagues, 2002; 2007) have acknowledged that the domain of mathematics extends beyond computation. However, Fletcher (2007) goes on to propose that it is due to the complex nature of calculation skills that researchers have been unable to identify the core cognitive correlates underlying MD. One of the fundamental propositions of this paper shall be to challenge the notion that arithmetical skill is the best indicator of mathematical ability, and that it therefore is unlikely to be the best area of study to determine the core deficits of MD.

Research on counting skills has emphasized the student's progression from using low-level strategies and building up to the automatic retrieval of number facts (Geary, 2004; Geary & Hoard, 2001). In the early stages of learning addition, children tend to employ a relatively simplistic strategy; they generally add numbers by counting from one, known as "counting all." Following mastery of this strategy, students begin using a "counting-on" technique, whereby counting is begun from the highest addend. This advancement in strategy normally occurs in the absence of explicit instruction on counting procedures (Geary, 2004; Geary & Hoard, 2001). With practice and repetition, many students eventually recall number facts quickly and automatically from memory. The underlying assumption, though not explicitly stated, seems to be that, changes in strategy, building up to automaticity, corresponds with greater understanding and this progression constitutes an aptitude for mastering higher-level mathematical concepts.

Some researchers believe that automatic retrieval of facts is necessary in order for students to understand higher-level mathematics (Fletcher et al., 2002; Gersten, Jordan, & Flojo, 2005). According to these researchers, it is thought that if students are unable to retrieve number facts quickly and accurately from memory, these retrieval deficits are detrimental to developing their ability to learn more complex mathematical concepts. Gersten and colleagues (2005) describe the importance of the quick retrieval of number facts for the importance of learning higher-level concepts as follows. As students progress in their school years, they encounter mathematics problems in which the focus is no longer on algorithms. However, correct completion of an algorithm may be necessary to solve a given problem. Teachers may assume students have already computed a number fact and therefore continue a lesson, using the number fact as an anchor. A student who remains stuck in the algorithm portion of the problem is often unable to engage in or follow the remainder of the lesson. However, being a computation wizard is not essential to success in mathematics (Devlin, 2000). Indeed, many mathematicians have admittedly low computation skills and mechanically following rules and procedures has little to do with the way mathematicians use mathematics (Devlin, 2000; Romberg & Kaput, 1999). It is not speed of computation nor skill in performing arithmetic calculations that predicts mathematical understanding and knowledge, rather the ability to reason analytically, to measure, and to approach problems creatively (Romberg & Kaput, 1999).

There is consensus in the field of MD that compared to research on reading disabilities; MD research is still in its infancy (Fletcher et al., 2002). As such, there has been some interest in using results of studies on reading disabilities to inform research on mathematics disabilities (Chiappe, 2005; Kulak, 1993). Drawing from reading research, one factor known to influence the development of successful reading skills is the ability to read words quickly and accurately (Savage et al., 2005). It seems that researchers in the field of mathematics have attempted to identify factors that mirror the cognitive correlates known to influence reading skill. Within mathematics research, there has been a focus on the development and progression of counting strategies, culminating in automatic retrieval of number facts (Pedrotty Bryant, 2005a). This assumption that fluid retrieval of number facts is a foundation for building proficiency in mathematics seems to parallel the research on the importance of quick and accurate word recognition for the ultimate goal of comprehending text. In both cases, children rely on lower-level skills (i.e. phonological knowledge and procedural knowledge) to work through unknown words and number facts, eventually building up to automatic retrieval, allowing for more cognitive resources to be spent on higher-level tasks such as reading comprehension or solving word problems (Kulak, 1993). While it is wise to use reading research to assist in the development of studies of mathematics disability, it is important to keep in mind that learning problems in mathematics do not necessarily share the same underlying cognitive processes. For example, Fletcher and colleagues (2007) report that the phonological processing deficits typically found in children with both MD and reading disabilities (RD) cannot account for the difficulty of retrieving math facts from long-term memory. It seems then, that MD is in need of conceptualization and research projects tailored to the specific nature of mathematics.

Butterworth (1999) suggested the possible existence of a “number module”, referring to a section of the brain responsible for processing quantity (as cited in Chiappe, 2005). Later research (Gelman & Butterworth, 2005) has suggested that there are two number modules; one which is responsible for processing exact calculations (the exact system), and the other for approximation (the analog system). The exact system is thought to be influenced more by language and culture than the analog system (Chiappe, 2005; Gelman & Butterworth, 2005). Interestingly, results from brain imaging studies have found that during estimation tasks, the horizontal segment of the intraparietal sulcus and the posterior superior parietal sulcus are activated (Gelman & Butterworth, 2005). This area of the brain is a fair distance from classical language processing areas (Gelman & Butterworth, 2005). The exact number module is thought to be responsible for processing numbers greater than four and essential in the development of mature number representations and in understanding basic mathematical concepts (Chiappe, 2005). This evidence provides support for the importance of language in the learning of mathematics.

Although there have been numerous research projects dedicated to the study of MD, a consensus has yet to be reached regarding academic skill deficits as well as core cognitive deficits, although MD diagnoses typically fall into either the computation/arithmetic category or the math reasoning category (Fletcher et al., 2007). In light of this, there is still some variability in the practice of diagnosing MD. One such difficulty in diagnosing MD is the lack of specifically designed assessment materials. Often, children are diagnosed with MD based on intelligence and academic achievement test scores taken together with other information including samples of classroom work, and parent and teacher reports (Fletcher et al., 2002). A reliance on academic achievement measures is problematic because the composition of mathematical subtests may not be sufficiently sensitive to detect MD (Geary, 2005; Pedrotty Bryant, 2005b). Without knowledge of the underlying cognitive deficits of MD, it is difficult to develop assessment tools that accurately test for the presence of MD. Furthermore, children's scores on mathematics clusters of academic achievement tests can be inconsistent from year to year (Geary, 2004). It may be that inconsistencies reflect difficulty with a specific concept or a different teaching approach, rather than the presence of a learning disability (Geary, 2004).

In addition to the problems inherent in relying on intelligence and academic achievement tests to detect the presence of MD, the use of conventional standardized assessment instruments have many well-known drawbacks. Not only have people of diverse backgrounds been consistently over-diagnosed (Fletcher et al., 2004), there are also issues relating to the proper use of statistical procedures to determine the presence of a learning disability, as well as the difficulty in drawing a line between "garden variety" low achievers, those who suffer from low achievement without neurobiological or other exclusionary causes, and those with true learning disabilities (Rivera, 1997). A lack of subject matter specificity combined with the drawbacks inherent in standardized testing increases the possibility of misdiagnosis.

The search for core academic and cognitive deficits affecting and explaining MD lags behind the reading disabilities research base. Fletcher and colleagues (2002) call attention to the fact that MD has rarely been studied beyond the scope of arithmetic and numbers. Much past research has focused on the study of mathematics learning disabilities by using numbers and arithmetical skill to inform researchers of the nature of MD.

The study of MD has often focused on numbers and arithmetical skill. Since mathematics is such a broad field, if researchers continue to rely on information obtained through studies of calculation proficiency without examining knowledge of the underlying processes or alternative formats of expressing knowledge, then the information obtained from assessment tools will not accurately capture the full extent of a student's mathematical knowledge. If an assessment of students' mathematical understanding is not undertaken, the full extent of a student's mathematical knowledge is not tapped, and it may be unethical to diagnose a learning disability. An examination of the history of intelligence testing, the different frameworks by which psychologists collect information to determine learning disability status, and possible future directions is necessary in order for the reader to grasp the present state of the practice of diagnosing learning disabilities in general and mathematics disabilities in particular.

History of Mathematics Achievement Testing

Achievement testing has been happening since as early as 1894. As one can imagine, many changes have taken place since that time. In education, it is often said that as the pendulum swings from one generation to the next, often what appear to be new ideas are simply old ideas repackaged (Linn, 2001). By studying the history of achievement testing, indeed the history of anything, we come to understand better how the present state came to be (Stearns, 1998). By taking the time to examine the history of achievement testing, information is found on how far the discipline has come and reminds us of good ideas that have been forgotten.

As early as 1894, records document the existence of mathematics achievement testing. Between 1894 and 1911, there were at least four known psychological tests that included an arithmetic component, developed by J.M. Rice, Alfred Binet, C.W. Stone, and S.A. Courtis (Bryant & Pedrotty Rivera, 1997). Following this period, interest in mathematics achievement testing continued to grow, with over 40 tests published between 1915 and 1925 (Bryant & Pedrotty Rivera, 1997). Among the tests published in this period, included is the Compass Diagnostic Tests in Arithmetic, developed by Ruch, Knight, Greene, and Studenbaker in 1925. A noteworthy element of this test is the volume of data it gathers. Contrary to some of today's tests which consider only the correctness of the final answer, the Compass collected information on the student's comprehension of the problem, the student's knowledge of information provided within the problem, the student's ability to pull out meaningful parts of the problem, the student's ability to estimate the answer, and finally, the correctness of the student's final answer (Bryant & Pedrotty Rivera, 1997). It seems that some of today's testing tools may lack the type of qualitative information gathered by tests like the Compass.

One purpose of assessment is to develop interventions tailored to the specific needs of the student (Dawson & Guare, 2004). Some psychological tests, such as the Wechsler Individual Achievement Test, Second Edition (WIAT-II) provide information outlining the underlying skill each question seeks to measure. This type of information helps school psychologists and other educational consultants to identify the type(s) of questions that identify the student's general strengths and weaknesses, but researchers have cautioned against their use in the development of instructional plans due to over-reliance on a small item set, disagreement over which skill items truly measure, and that interventions are not tested for effectiveness prior to being recommended (Bryant & Pedrotty Rivera, 1997; Hammil & Bryant, 1991; Parmar, Frazita, & Crawley, 1996; Salvia & Yesseldyke, 1995; Taylor, Tindal, Fuchs, & Bryant, 1993). Other mathematics achievement tests (e.g. Woodcock Johnson Tests of Academic Achievement, Third Edition [WJ-III] Math Fluency subtest) focus on the speed at which the student produces answers to calculation questions. This type of test reveals how quickly a student retrieves an answer from long-term memory, or the speed at which the student calculates an answer, but does not facilitate suggestions for remediation. Bryant and Pedrotty Rivera (1997) describe the purpose of norm-referenced tests in general as being developed to compare an individual student's performance to that of a national sample. These tests rely on a small number of test items to represent a student's skill level in a broad range of problem-types. Furthermore, they suggest that because norm-referenced tests are not comprehensive assessments, they are ill used for constructing instructional plans. This suggestion is in keeping with the National Council of Teachers of Mathematics' (NCTM, 1995) call for the use of a range of assessment methods for evaluating students' mathematical prowess (Bryant & Pedrotty Rivera, 1997). In today's reality of scarce resources and waiting lists, it is easy to see why professionals have attempted to develop measures that may quickly and accurately provide information about a student's difficulties in a particular academic area. However, it seems that somewhere along the way, assessment tools became less explicit in their ability to facilitate an examiner's ability to identify highly specific knowledge gaps and areas of difficulty and link this information to ready-made interventions. This type of testing represents a clear shift from the types of tests used in days past.

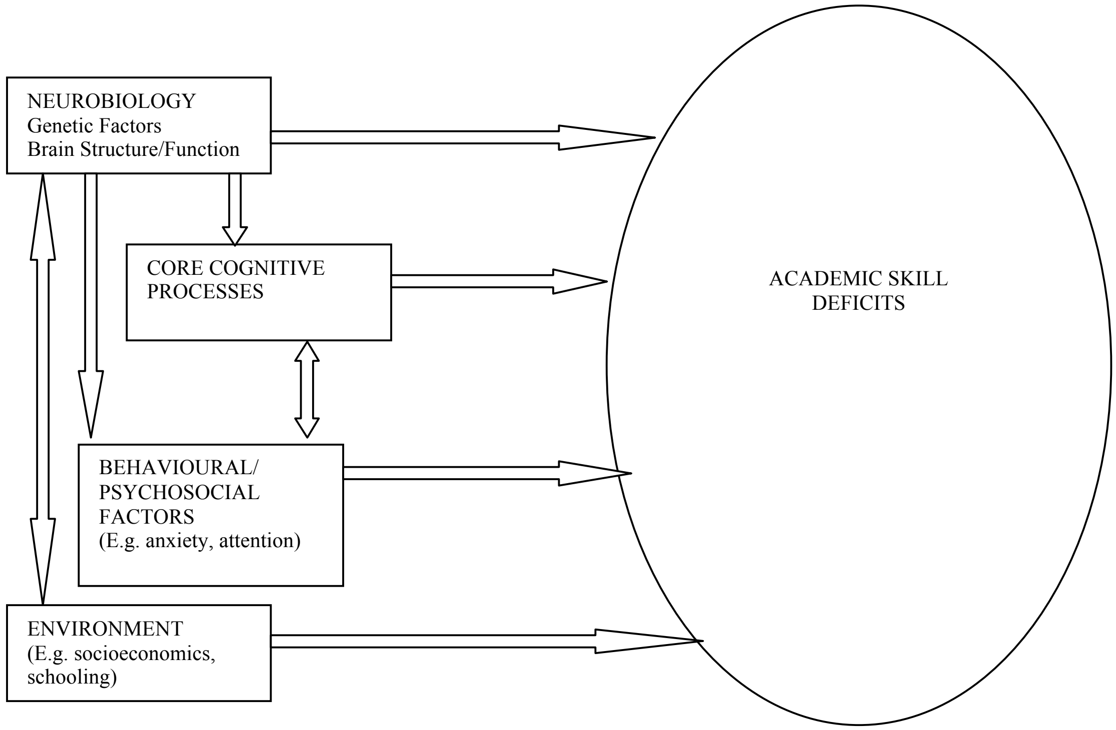
Valuable information may be gained by studying the history of any subject. By reviewing the history of achievement testing, it is uncovered that former tests of arithmetical ability appear to be less mechanical than some of those in use today. In particular, it seems that previous tests examined factors beyond the student's ability to produce a correct answer and used errors to identify and remediate a student's difficulties (Bryant & Pedrotty Rivera, 1997). Current achievement tests in widespread use do not appear to facilitate educational professionals' understanding of students' difficulties nor do they appear to link well to remediation plans. Next, a commonly used and well-debated framework for assessing learning disabilities is discussed.

Traditional Assessment of Specific Learning Disabilities

There is much debate within the field of learning disabilities as to how a learning disability is best defined (Fletcher et al., 2007). The classification model that a psychologist selects determines the types of tests and procedures he or she uses to determine a student's learning disability status. In the context of the present study, the use and perhaps over-reliance on tests that are heavily based on calculation to inform practitioners of a student's mathematical skills and areas of weakness is called to attention in an effort to encourage practitioners to think critically about the way mathematics is defined within their chosen assessment tools, and to encourage them to select assessment measures that conceptualize mathematics in the broader sense.

Part of the DSM-IV-TR criteria outlines discrepancy analysis that involves making sure that a child's academic achievement is significantly lower than would be predicted based on ability as a primary method for determination of learning disability status. According to Fletcher and colleagues (2004), there are four basic tenets of discrepancy analysis: (1) Discrepancy, (2) Heterogeneity, (3) Exclusion, and (4) Presumption. Usually, ability is measured by an IQ test and achievement is measured by an academic achievement test, such as the Wechsler or Woodcock-Johnson cognitive and academic scales. Academic achievement scores are predicted on the basis of one's IQ scores (i.e. ability) and compared to the scores of the normative population. A clinically significant discrepancy is determined by a statistical formula, whereby a student's academic achievement falls approximately two standard deviations below what would be predicted based on performance on the intelligence test. *Heterogeneity* refers to the various academic areas that a student may struggle in. A child may have a learning disability in one or more of the following specific areas: reading, mathematics, written expression, and language (Fletcher et al., 2004). The *exclusionary* tenet describes various other causes that cannot be at the root of the problem in order for the problem to be classified as a learning disability. For learning disability status, the cause of the learning problem(s) may not be due to a "sensory disorder, mental deficiency, emotional disturbance, economic disadvantage, linguistic diversity, or inadequate instruction" (Fletcher et al., 2004). The mental deficiency exclusion prevents anyone with below average intelligence to be classified as learning disabled. The *presumption* is that achievement problems are believed to lie within the child, with neurological problems that should not impact overall IQ at the root of the problems (Fletcher et al., 2004). However, "it is widely recognized that the presence of IQ-achievement discrepancy, an academic difficulty, and absence of the exclusions does not mean that the child has a neurobiological disorder" (Fletcher et al., 2004; p. 310). Figure 4 shows a graphic view of some factors that may affect academic achievement.

Figure 4.
Factors Affecting Academic Skill Deficits (Fletcher et al., 2007)



The discrepancy model has been heavily criticized in recent years. Of primary concern is the typically lengthy passage of time that passes prior to a student obtaining learning disability status, and therefore receiving access to additional supports and services through the public school system, making the discrepancy model known as the "wait to fail" model (Fletcher et al., 2004). The name comes from the gap that is required between IQ and achievement before a student is considered sufficiently below his or her peers, usually 2-3 grade levels behind the student's current grade placement. Clearly, this gap is frustrating for parents, school professionals, and the student. Another common criticism of the discrepancy model is the high cost associated with it because it requires a psychoeducational assessment, with an average psychoeducational assessment costing thousands of dollars (Fletcher et al., 2004).

There has been some movement away from a traditional discrepancy model towards a model that does not rely on testing to provide access to treatment. Fletcher and colleagues (2004) believe that many students' difficulties may be remediated with high quality, evidence-based intervention without psychoeducational testing. Fletcher and colleagues (2004) suggest that the discrepancy model can be harmful to students due to the fact that it delays intervention until performance is sufficiently low enough to produce an IQ-achievement discrepancy. Torgesen and colleagues (2001) point out that for many students, a learning disability is not identified with the use of the discrepancy model until a later age, at a point where it is difficult for a student to catch up to their same age peers. Since average-developing children are continuing to add to their knowledge bases, a struggling child essentially has to work twice as hard in order to catch up (Torgesen, 2002).

Those who uphold the discrepancy model argue that a learning disability cannot be identified without some measure of cognitive functioning. Fletcher and colleagues (2004) point out that continuing to propagate the use of the discrepancy model would be to ignore the consensus reports from four major bodies of research within the special education field, namely, the National Research Council's report on the over-representation of minority students in special education (Donovan & Cross, 2002), the Fordham Foundation and the Progressive Policy Institute's (2001) report on "Rethinking Special Education", the Learning Disabilities Summit by the US office of Special Education Programs (2002), and the President's Commission on Excellence in Special Education (2002).

Fletcher and colleagues (2007) propose a hybrid model for assessment of learning disabilities, including a response to intervention (RTI) component, norm-referenced academic achievement measures, and assessment of other non-academic factors that may negatively impact a student's achievement (Fletcher et al., 2007). Fletcher and colleagues (2007) point out that assessment of adequate instructional opportunities plays a key role in their hybrid approach. Diagnoses of learning disabilities are made only after systematic instruction in the deficient area is delivered. By contrast, a learning disability could be diagnosed using the discrepancy model based on the results of an IQ test, achievement test, and with other qualitative information, such as parent and teacher reports (Fletcher et al., 2007).

Standardized mathematical achievement tests are typically used to screen children struggling with mathematics, but are not sophisticated enough to distinguish between children whose struggles are of a cognitive nature or due to other factors, such as inadequate instruction (Geary, 2005). Achievement tests differ from aptitude tests in that the former measure a child's current levels of academic achievement whereas the latter measure a child's current cognitive functioning. Therefore, we cannot conclude that children who perform poorly on achievement tests lack the ability to understand mathematical concepts or that their difficulties will persist. Indeed, Geary and colleagues (1991) found that a single assessment finding low achievement did not reliably predict low achievement scores in the future. Many students who performed poorly on an initial assessment later went on to score within the average range. Geary and colleagues (1991) conclude that low achievement scores in math in one year do not necessarily indicate MD. In sum, we cannot equate a low achievement score with a disability (Fletcher et al., 2007), nor can we equate it with an inability to perform or to comprehend mathematics; rather, low achievement scores are a cue to educational professionals to look further into the potential causes of the problem rather than to conclude that the problem lies within the child.

Stanovich (2005) argues that the use of the discrepancy model for identification of learning disabilities is pseudoscientific, given that there is a lack of evidence to support its continued use. Stanovich (2005) reviews the evidence in the learning disabilities field relating to reading disabilities and shows that there is not a difference in skill deficits between high IQ and lower IQ students. This point is important because children with lower IQ's are less likely to show a discrepancy between aptitude and achievement, and thereby not receive a diagnosis of LD, which may exclude them from receiving special education services (Stuebing et al., 2002). Also, children with lower IQ's and higher IQ's have both been shown to respond to high-quality, evidenced-based instruction (Stanovich, 2005). Further, Stanovich (2005) states that IQ tests are not necessarily a good measure of aptitude.

As mentioned before, the discrepancy model has been criticized primarily due to the lengthy passage of time that sometimes occurs prior to a student's academic achievement scores being sufficiently low to produce a statistically significant difference based on IQ test performance (Stuebing et al., 2002). This passage of time is frustrating for all those involved with the struggling child, as well as the child. Waiting to deliver high-quality, evidence-based intervention seems counterintuitive, especially since we know that academic achievement problems that are not resolved early are typically extremely difficult to remediate (Torgeson, 2001). Those who uphold the discrepancy model argue that a learning disability cannot be identified without a test of cognitive ability. In terms of academic achievement, Geary's (1991) research points out that academic achievement scores may fluctuate from year to year and may not necessarily indicate a learning disability. In order to address some of the problems inherent in the discrepancy model, a newer framework, known as response to intervention (RTI) has been adopted in some areas and is discussed in the next section.

Response to Intervention (RTI)

With the 2004 American re-authorization of the Individuals with Disabilities Education Act (IDEA), each state has the right to include an RTI component in state definitions of learning disabilities (Hale, Flanagan, & Naglieri, 2008; Stanovich, 2005). Since most definitions of a learning disability include access to quality instruction as one of the exclusionary criteria (Fletcher et al., 2004), the RTI model provides a framework from which to assess students' response to instruction. Fletcher and colleagues (2004) describe the process as follows. Students are screened early in their school career, either in kindergarten or grade 1. Those who score in the bottom 25-30^{th}percentile are identified as at-risk, and provided with more intensive instruction. Progress is carefully monitored, and those who do not benefit from increasingly intense instruction may be identified as learning disabled. From the reading disabilities research, neuroimaging studies have found that non-responders fail to activate certain left-hemisphere areas known to be implicated in reading and instead show predominant right hemisphere activation (Fletcher et al., 2007). The RTI model provides a framework from which to assess a student's access and response to quality research-based instruction, whereas the discrepancy model does not provide a ready means for assessing this important component (Fletcher et al., 2004). The RTI literature contributes to the present study in terms of knowledge of possible future directions within the learning disabilities field.

There are many advantages to the RTI model. Recall that parents and families of children with learning problems report being frustrated with the time-consuming process of the discrepancy model. In the RTI framework, the focus changes from determining eligibility for special education services to providing evidence-based instruction to at-risk children (Fletcher et al., 2004). Additionally, the RTI model may reduce the number of teacher-biased referrals. Researchers have noted the over-population of minorities and male students in special education referrals (Burns & Senesac, 2005; Fletcher et al., 2004; Hosp & Reschly, 2004). The RTI model reduces the problem of teacher biases by screening all children rather than only those referred by classroom teachers and other school professionals (Fletcher et al., 2004). For example, using the discrepancy model, students from African American, Hispanic, and First Nations groups are identified as LD at a disproportionate rate as compared to Caucasian and Asian American children (Hosp & Reschly, 2004). However, within the RTI framework, there is a more representative proportion of the population who are identified as LD (Burns & Senesac, 2005). In addition, the RTI framework offers a ready means by which to evaluate a student's response to quality instruction, thereby satisfying one of the exclusionary criteria. Namely, that the source of a student's academic difficulties cannot be due to inadequate instruction for the diagnosis of LD (Fletcher et al., 2004). "Including RTI as one of the criteria for identification allows educators and parents to immediately provide students with well-targeted and much needed intervention rather than waiting for extensive, time-consuming assessments that offer little or no information to inform instruction" (Fletcher et al., 2004, p. 312). Traditional discrepancy based models have little usefulness in providing information for planning interventions whereas inclusion of an RTI component allows more of a focus to be on providing quality instruction to students, rather than on determining eligibility for services (Fletcher et al., 2004).

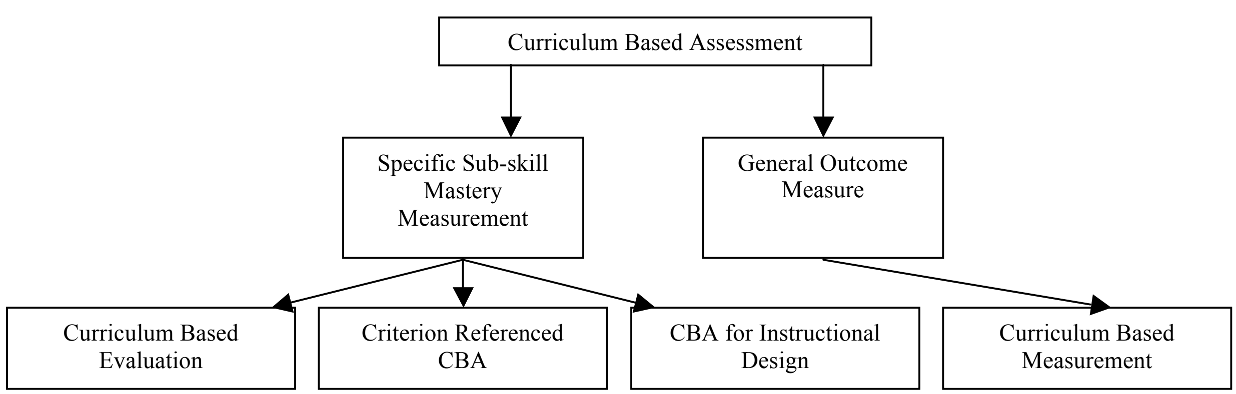
In conclusion, the RTI model appears to offer a viable solution to providing struggling students with access to evidence-based intervention at a time when their difficulties are more likely to be remediated successfully. This model also provides a ready means for testing one of the exclusionary criteria within the definition of most learning disabilities: access to adequate instruction. Additionally, the RTI model may provide practitioners with more valuable information for planning interventions than that typically available via norm-referenced measures. The RTI framework also offers a potential solution to the problem of referral bias. The RTI model appears to be very promising and it will be interesting to track the progress and obstacles encountered by areas that have already implemented RTI into their existing frameworks.

Curriculum Based Assessment (CBA)

Curriculum Based Assessment (CBA) can be used more frequently than standardized tests. Due to potential practice effects, where a child may superficially inflate his/her score due to familiarity with a particular item, rather than due to increased skill, generally, a child may only be administered the same academic achievement test within one year of the original testing date and within two years for a test of cognitive ability. CBA was of interest to the present study not only because it represents an up and coming framework that is being used with increasing frequency, but also to examine the types of questions and information it provides in comparison to norm-referenced tests.

Curriculum based assessment tools are often employed when short-term progress is monitored. The general purpose of CBA is to measure “a student’s skill development toward (more) general outcomes” (Hintze, Christ, & Methe, 2006, p. 45). Most forms of CBA break the curriculum into smaller chunks, focusing on a more specific set of subskills. According to Hintze and colleagues, (2006), there are four types of CBA: Curriculum-based assessment related to designing instructional programs (CBA-DI); Criterion-referenced CBA; Curriculum-based evaluation (CBE); and Curriculum-based Measurement (CMB). Practitioners select the appropriate subtype based on the referral question. As Figure 5 shows, CBM is most often used for longer-term tracking of academic goals, whereas the other three types of CBA are more useful for measuring short-term progress.

Figure 5.
Different Types of Curriculum Based Assessment (CBA) Frameworks (Hintze et al., 2006)



The rationale behind CBD-DI recognizes the academic diversity in the classroom. Since a classroom is full of students with differing entry skills, using a single teaching pace and instructional level almost guarantees that some students will be left behind. The goal of CBA-DI assessments is to find the best instructional level for the child (Hintze et al., 2006). Hintze and colleagues (2006) describe the procedure as follows: Material is selected from the curriculum, usually something that is currently being taught. Information is then collected on the student's skill levels on various sub-skills (e.g. counting skills, basic facts). The new information is then used to determine the best instructional level for the student, such that he or she can be successful about 95% of the time when working independently. Progress is carefully monitored and instruction continues until the skill is mastered.

Using criterion-referenced CBA, students are tested over a three-day period, using alternate test forms representative of the curriculum outcomes (Hintze et al., 2006). The student's performance is then compared to other students in the classroom. This format is sometimes preferred since it provides information regarding an individual student's performance compared to other students in his or her class, rather than compared to a national sample (Hintze et al., 2006). Other types of information derived from criterion-referenced CBA include a description of the student's skills in the academic areas of interest, and error analyses (Freeman & Miller, 2001). Some studies have found that teachers prefer the information derived from CBA than from traditional norm-referenced tests (Chafouleas, Riley-Tillman, & Eckert, 2004).

When using curriculum-based evaluation (CBE), a general survey of the area of difficulty is given, followed by more in-depth investigations. The goal is to identify which tasks the student is ready to learn based on his or her current knowledge base (Hintze et al., 2006). According to Hintze and colleagues (2006), knowledge is broken down into *subtasks* and *strategies*. Subtasks refer to the background knowledge and skills a student must possess in order to engage in the task. "If a student is missing a subtask component, he or she is missing one of the essential building blocks of the task. If a student is missing a strategy component, he or she may have all the material necessary to succeed at the task, but not know how to assemble it" (Hintze et al., 2006, p. 50). Strategies represent the rules and procedures that a student must be knowledgeable of in order to successfully engage in a given area. This type of analysis breaks an area of difficulty down into smaller sub-parts and provides practitioners with valuable information for understanding the students' problems and what types of knowledge gaps exist, in order to facilitate remediation.

Since curriculum based measurements (CBM) are less sensitive to smaller gains within specific sub-skill areas, they are often used to measure long-term progress (Hintze et al., 2006). When combined with other types of CBA, curriculum based measurements can be part of a larger battery of assessments used to identify, remediate, and track students' progress. In CBM, students' performance is compared to local norms. By sampling a student's proficiency across a range of items representative of the curriculum, CBM can be used as a tool to assist practitioners in selecting the appropriate level of instructional materials for students (Shapiro, Angello, & Eckert, 2004). CBM provides information that can help practitioners to make a variety of educational decisions (Deno, 2003).

Shapiro and colleague's (2004) research suggests that CBA is being used with increasing frequency for several reasons. First, some variations of CBA can be administered more often to the student, and therefore provide more information on growth in smaller areas that may not be captured with norm-referenced tests. Second, some forms of CBA provide practitioners with information on how a student is performing within the context of his/her own curriculum. This is in contrast to norm-referenced tests which offer a rough approximation of curriculum, given that a student's performance is compared with a *national*, rather than *local* sample. Third, the information garnered using CBA can provide educational professionals with information regarding the best instructional level for the individual, thereby providing a ready means for linking assessment to intervention. Some drawbacks to using CBA include the time commitment involved in creating separate forms for repeated testing, as well as when the emphasis becomes on rote learning rather than understanding (Sattler, 2001). Another framework which has been gaining increasing attention is that of dynamic assessment.

Dynamic Assessment/Brief Experimental Analysis

Dynamic Assessment measures, and a related type of measurement known as Brief Experimental Analysis, are used to inform intervention by endeavouring to discover which instructional methods work for the individual child (Chafouleas et al., 2004; Freeman & Miller, 2001). Rather than basing recommendations for remediation on the students' cognitive and academic profile alone, dynamic assessments test the effectiveness of a particular intervention prior to making recommendations. Brief Experimental Analysis measures are usually combined with CBM in order to identify the specific areas of weakness for the student with the ultimate goal of determining a successful intervention (Chafouleas et al., 2004). The roots of Brief Experimental Analysis come from functional behavioural analysis in which the effect of the interventions are monitored to determine which will produce the greatest effect for the student (Chafouleas et al., 2004), representing a systematic and longer-term assessment strategy over that of limits testing. Within the context of the present study, dynamic assessment models represent a different approach to testing that was in need of consideration in order to think fully about how best to assess students' mathematical understanding.

Dynamic assessment tools are less used than traditional norm-referenced assessment measures, and therefore the results have been found to be less familiar to educational professionals. However, in Freeman and Miller's (2001) study, special education coordinators reported that the information garnered from dynamic assessment tools were more useful for understanding students' difficulties and for planning intervention than norm-referenced measures. In fact, although norm-referenced measures were rated by educational professionals as producing the most familiar type of assessment results, the information was ranked as least helpful in understanding students' difficulties, and in intervention planning (Freeman & Miller, 2001). Educators' ratings of the information provided by curriculum-based assessments fell in between that of traditional norm-referenced assessments and dynamic assessment models. Typically, the types of information collected from dynamic assessment measures includes information regarding specific teaching strategies that work for the student, a description of deficient thinking skills, information about strategies the child used prior to intervention, and information about how much instruction the child required in order to succeed (Freeman & Miller, 2001).

Although dynamic assessment measures are used less often than other types of assessment tools and the results are less familiar than those garnered by norm-referenced tests, the results appear to be valuable to those who work closely with students. By testing the effectiveness of a given intervention prior to offering recommendations, psychologists are in a better position to be confident that their recommendations will be successful since the effectiveness has already been tested with the child. In addition, dynamic assessment models are useful for distinguishing between learning problems and LD among English as a Second Language (ESL) students (Sanez & Huer, 2003).

Conclusion

I have discussed children's early mathematical knowledge and how this authentic approach to using mathematics to solve real-life problems is sometimes replaced by a superficial approach to mathematics, in terms of symbols, rules, and procedures. Mathematics encompasses the ability to think and to approach problems creatively, to see and to work with patterns, and to use mathematics as a tool to aid in one's thinking. Clearly, the scope of mathematics and mathematical ability encompass more than numbers and algorithms. There are many factors that affect performance in mathematics, including anxiety and gender or cultural stereotypes. These are real, though often invisible, threats that negatively impact mathematical performance for some (Ashcraft et al., 2007; Royer & Walles, 2007).

There are a wide variety of factors that affect learning and successful performance in mathematics. There are also different assessment approaches used within the field of educational psychology to determine the nature of students' learning difficulties. Academic skill deficits are often the first clue to teachers and other educational professionals that a child may have a learning disability, however, as discussed, there are other potential causes for the lagging skills that need to be explored, in addition to the learning disability hypothesis. Some researchers (see Fletcher et al., 2004 for a review) have suggested that the discrepancy model should be abandoned in favour of other classification systems for identifying LD. For example, within the RTI model, there is a significantly less emphasis on IQ testing and more on progress monitoring. Interestingly, this conceptualization of learning disabilities does not change the fundamental concepts of discrepancy and unexpected underachievement. The difference lies in the conceptualization of discrepancy as a failure to achieve *in spite of adequate instruction* rather than due to discrepant ability and achievement, given the expectation that most students can achieve when exposed to high quality instruction (Fletcher et al., 2004). However, there are additional issues that require further research, such as distinguishing a child with developmental delay from LD and determining if it is possible for a developmentally delayed child to have LD. In light of the problem with conceptualizing and assessing mathematical ability in terms of calculation skill, the next section presents research questions aimed at discovering new information to add to the field.

Chapter 3: Research Questions

Purpose of the Study

Given that MD has largely been studied in terms of calculation skill and that the ability to compute numbers efficiently is not necessarily indicative of mathematical understanding (Devlin, 2000), the aim of the present research project was to review critically an existing mathematics achievement measure and to develop and pilot seven alternative test items that aim to measure mathematical thinking and understanding. Past research has suggested that the ability to retrieve basic mathematical facts automatically from long-term memory facilitates the learner's ability to engage in higher-level mathematics (Fletcher et al., 2002, Gersten, Jordan, & Flojo, 2005). A focus on speeded retrieval of mathematics facts has permeated both the classroom and the psychoeducational assessment culture (Fletcher et al., 2002). While certainly the importance of knowing mathematics facts is clear, it is less clear whether psychoeducational assessment measures relying heavily on calculation provide practitioners with sufficient evidence on which to base, even partially, a diagnosis of MD (Parmar et al., 1996). It was hypothesized that the alternative test items would provide educational professionals with a greater quality of information on which to base important decisions, such as educational programming and determination of learning disability status.

Less is known about mathematics difficulties and disabilities in comparison to the reading disabilities research base. The aim of the present study served two basic purposes. First, I critically examined a diagnostic tool used to assess mathematical difficulties and MD. Past studies have found traditional mathematics assessment measures to be over-reliant on calculation-based problems in order to inform educational professionals about the nature and severity of students' mathematical difficulties (Parmar et al., 1996). For example, in a review of mathematics assessment tools, Parmar and colleagues (1996) found that of a mathematics inventory comprised of 243 items and 8 subtests, 73% focused on arithmetical problems (Sequential Assessment of Mathematics Inventory – SARI). They concluded that "the majority of instruments and procedures used for assessment of students in special education do not reflect important mathematics, nor do they provide information that will drive or guide curriculum or instructional decisions (p.128)." The second aim of the present study was to create and to pilot seven alternative test items, representing the type of questions that access students' mathematical thinking and understanding, beyond that captured by measuring calculation skills. By evaluating the quality of information garnered by assessment tools and contrasting it with that gathered by the pilot items, researchers and practitioners are a step closer to understanding the complexities of learning problems in mathematics and what types of questions are most apt to assess mathematical knowledge.

Research Questions

The first research question seeks to test for any significant differences between selected items from the WIAT-II (Numerical Operations and Math Reasoning subtests) and the Pilot Items. By doing so, the researcher is able to determine whether or not the WIAT-II items and the Pilot Items effectively measure the same or different constructs. If there were not a statistically significant difference between the two item sets, this result would suggest that the Pilot Items do not measure a construct that is different in any meaningful way from what is measured by using the WIAT-II items. However, if a significant difference were found between the WIAT-II items and the Pilot Items, this result would suggest that the Pilot Items are tapping into a skill that is not captured by the WIAT-II items. The validity of the Pilot Items was established by being evaluated and approved by an experienced mathematics educator within the field and by an experienced school psychologist who is knowledgeable of standardized assessments. Also, the items were constructed with the NCTM criteria in mind, which additionally contributes to the validity of the pilot items. A significant difference between the WIAT-II items and the Pilot Items would be noteworthy in the context of the present study, because it would suggest that the Pilot Items represent a set of test items that could be used to refine current mathematical achievement tests such that a greater emphasis would be placed on mathematical thinking and understanding.

If the Pilot Items are found to measure a construct that is suggested to be significantly different from that of the WIAT-II items, then it would be safe to conclude that the two item sets measure different constructs. The WIAT-II Numerical Operations items were developed to measure calculation skills whereas the WIAT-II Math Reasoning items were developed to measure the application of math skills in context, such as solving word problems. However, the work of Parmar and colleagues (1996) suggests that some of the WIAT-II Math Reasoning items are improperly classified. The Pilot Items were adapted from the existing WIAT-II items and modified to measure mathematical thinking and understanding based on a review of the literature and in particular the work of Ruch and colleagues (1925), Nancy Jordan (2006), David Geary (1999, 2001, 2005), Lambdin and Walcott (2007), Jitendra and DiPipi (2003), Carpenter and Leher (1999), and the National Council of Teachers of Mathematics (1995, 2000, 2006).

The second research question seeks to determine the relationship between calculation proficiency and mathematical understanding. The answer to this question was of great interest in that it shed light on the importance of understanding in students' overall mathematical knowledge.

The third research question seeks to discover the extent to which the WIAT-II mathematics subtests (Numerical Operations and Math Reasoning) align with the assessment standards of the National Council of Teachers of Mathematics (NCTM). As such, a critical examination of both the WIAT-II mathematics subtests and the Pilot Items was undertaken using an evaluation rubric developed based on the NCTM's Assessment Standards (1995) and rubric for evaluating large-scale assessment tools (2006).

The fourth research question seeks to determine if any of the Pilot Items measure the areas described by the NCTM and detailed in the rubric. The Pilot Items were compared to the evaluation rubric and results are reported qualitatively. By evaluating both item sets, new information relating to areas known to be of importance to learning and successful performance in mathematics will be brought to light. In the following section, I will describe the procedures that were followed in order to answer the research questions.

Chapter 4: Methodology

Introduction

In the next section, I outline the research design for testing the research questions and describe the manner in which the Pilot Items and Evaluation Rubric were constructed. Following, I describe the ethical considerations that pertain to the present study, the selection of participants, and the method by which the data were analyzed. Finally, I close the method section with a summary of the chapter.

Research Design

The suggested test items were a modification of existing items, extracted from the WIAT-II Numerical Operations and Mathematical Reasoning subtests. The original WIAT-II items were selected on the basis that they represent typical test items measuring basic calculation skills or word problems involving simple algorithms. The level of difficulty for the items ranges from grade 4 to grade 6, according to the Nova Scotia Mathematics Curriculum outcomes. One reason for testing older students is that although IQ's can change over time, a certain amount of stability is usually found after age 6 (Sattler, 2001). In light of this, the population of interest for the current research study is upper elementary aged children, since this is the time frame in which children are typically referred for psychoeducational testing. Testing time is an important variable to be mindful of in today's reality of scarce resources. Given that the Pilot Items do not comprise a comprehensive battery, it is difficult to determine how they would compare time-wise to an assessment with the full mathematics subtests of the WIAT-II. However, it was observed that the Pilot Items were completed within a typical time frame. Future research will need to confirm this finding.

Construction of the Pilot Items

The construction of the Pilot Items was based on the work of Ruch, Knight, Greene, and Studebaker who developed the (1925) Compass Diagnostic Tests in Arithmetic. This test contains a component which allows practitioners to assess the student's general understanding of the problem, the overall process, and the solution. The Compass provides the examiner a moment-to-moment standardized procedure which allows him or her to acquire information about the student's thought processes while working on a problem, rather than waiting until the end of a subtest before being able to probe further into a child's reasoning. Information is collected for each item and compared to a set of 32 common error types in an effort to facilitate both understanding of the student's difficulties and the remediation of specific deficits. In addition to the correctness of the solution, the Compass collects information on the quality of the student's estimation, as well as the caliber of the student's comprehension of the problem to inform practitioners of the nature of the student's specific difficulties.

The work put forth by Ruch and colleagues (1925) provides the framework for the development of the Pilot Items. The items developed by Ruch and colleagues tie in well with present-day mathematical constructs, widely recognized to influence mathematical knowledge, achievement, and understanding. For instance, the ability to estimate the reasonableness of one's solution relates to the concept of number sense (Berch, 2005). Number sense is difficult to define, however, Nancy Jordan (2006), a respected researcher within the field of mathematics learning disabilities, and her colleagues are currently conducting a longitudinal study regarding number sense in kindergarten-aged children. They assess number sense along the following constructs:

counting skills, knowledge of relationships among numbers, nonverbal calculations, estimation, and story problems (Jordan et al., 2006). Witzel, Ferguson, and Brown (2007) state that number sense is to mathematics as reading comprehension is to literacy; that is, the ultimate goal is one of understanding.

Children with lower-level counting skills have been found to have more difficulty inhibiting irrelevant information when solving mathematics problems (Geary, et al., 1999). For example, when retrieving a basic math fact, other facts may be triggered in the student's long-term memory. This additional information makes it more difficult for the student to direct focus to the most important information (Geary et al., 1999). For example, when solving the problem 4+5, the student may retrieve the correct answer of 9, or potentially 6, being the next number in the sequence, or 20, the product of 4 x 5. On this basis, some of the Pilot Items include questions that tap into a student's ability to inhibit unessential pieces of information.

Geary (2005) states that assessment tools developed to measure children's mathematical understanding and ability are intimately linked to developmental theories and children's emerging competencies in these areas. Piagetian concepts that are critical foundations of mathematical concepts and relations include the concepts of object permanence (the understanding that an object continues to exist, although it may be outside of one's vision) and seriation (the ability to sort objects according to a common characteristic; Rathus, 1988). These concepts occur in the concrete operational stage, between the ages of 7 and 12, such that students in the present study presumably have reached a sufficient developmental stage to solve the mathematics problems selected (Rathus, 1988).

Constructivist theorists believe that students develop their own math attack strategies, which may be quite different from those presented by the teacher (Lambdin& Walcott, 2007). Proponents believe that the students' chosen strategies can be more conducive to the individual's thinking and reasoning style, and that much can be learned from the manner in which the student chooses to approach the problem (Lambdin& Walcott, 2007). Jitendra and DiPipi (2003) report that students with disabilities tend to experience considerable difficulty solving word problems, particularly with problem representation and identifying relevant information, along with difficulties in identifying operations. This finding is not surprising given the mechanical drill-and-practice approach common still in many classrooms. As such, questions are developed for the pilot items that invite students to represent their understanding of problems by drawing a picture, by selecting the appropriate operation from a list, and by reporting which information is important to know in order to solve a given problem.

The generation of a word problem from an algorithm requires the ability to link abstract symbols with real-world concepts and suggests a deeper level of understanding (Small, 1990). Hiebert (1984) states that this connection between school-math and real-world context is commonly missing in many mathematics classrooms. Children who are unsuccessful at making connections among mathematical concepts view mathematical concepts in isolation (Carpenter & Leher, 1999) and presumably would be unable to or have difficulty with transposing an algorithm into a real-world problem. A student who is able to create a word problem which suits the given algorithm demonstrates knowledge of and facility with the concept beyond that implied by basic calculation (Small, 1990). Similarly, a student who is able to extract the correct numerical and operational symbols from a word problem and represent these as an algorithm understands more about the concept than the student who is able to compute the algorithm yet is unable to work with the problem in new ways (Small, 1990). As such, questions are included in the Pilot Items that explore a student's ability to generate a real-world question from an algorithm and ability to create an algorithm from a given word problem. Please see Appendix A for a copy of the Pilot Items used in this study.

To serve as a comparison basis, five items were selected from the WIAT-II mathematics subtests. Two items were selected from the Numerical Operations subtest measuring basic addition and subtraction skills using whole numbers (e.g. 135-22). Also, three items from the Math Reasoning subtest measuring one's ability to create and solve multiplication and division problems using whole numbers were selected (e.g. determining an average number of scores from a given word problem). The Pilot Items were also developed based on these five WIAT-II items. The information gathered from the WIAT-II items was compared to the information gathered with the Pilot Items.

Construction of the Evaluation Rubric

The rubric for evaluating the pilot items and WIAT-II mathematics subtests was constructed based on aspects of the NCTM's Large Scale Assessment Tool (2006). The first factor refers to the *content of the measure*. On their website, NCTM (2006) states that assessment tools should measure a broad range of mathematical processes, such as problem solving and reasoning. For example, Parmar and colleagues (1996) reported that of the mathematical achievement tests reviewed in the context of their study, approximately 70% of test items measured calculation skills. Clearly, decisions regarding a student's mathematical skills should not be made on the basis of tests that primarily sample one domain of mathematics. As such, reference is made within the rubric to the range of content contained in the assessment tools.

The second factor pertains to the extent to which the assessment tool facilitates the *examiner's awareness of the student's process*. The NCTM (2006) refers to this area as accounting for the "messier" aspects of doing mathematics. Inspecting the student's process can be thought of as "error analysis." Error analysis is concerned not only with the final answer, but also with the quality and type of errors made (Borasi, 1994). This type of investigation can provide information about a student's mathematical thinking and can be used to isolate and remediate problems (Borasi, 1994). Reference is made in the rubric to examining the process a student undertook to solve the problem and takes into account whether or not students are awarded points on this basis.

The third factor deals with the extent to which students are provided with *multiple options for representing their knowledge and understanding of problems*. The NCTM (2006) outlines this standard as one in which students are provided with test items that maximize the likelihood that they are able to express what they know. Examples of this listed by NCTM include things like providing enough white space for students to record their responses, or the use of simple vocabulary. Taking this standard one step further would include evaluating the extent to which items provide multiple opportunities for students to represent their knowledge in non-traditional formats. This concept refers to the degree of flexibility the test allows for communicating one's knowledge. For example, a test which allowed for students to demonstrate their facility with symbolic representation or their ability to produce a picture to aid in problem solving is one which examines mathematical thinking beyond that captured by a student's ability to correctly apply a standard procedure.

The fourth factor looks at the *time constraints* imposed by the assessment tool. The NCTM (2006) states that students should be allowed sufficient time to complete their work, unless determination of the student's efficiency is one of the goals of assessment. As such, the influence of time constraints on students' performance was considered.

Finally, the fifth factor assists the examiner in determining the *ease with which assessment results are linked with intervention* (NCTM, 2006). This is a key matter since assessment does not end with diagnosis, but rather with the design and implementation of meaningful interventions aimed at remediation of students' areas of difficulty (Small, 1990). Frameworks such as dynamic assessment models represent the ideal method for linking assessment to intervention since strategies are tested for effectiveness during the assessment. Therefore, an informed intervention may be made, including knowledge of strategies that are likely to be successful (or unsuccessful) for a particular student. As such, I evaluated the measures to determine if any dynamic assessment procedures are utilized as part of the standard procedure. Please see Appendix B for a copy of the Evaluation Rubric used in this study.

Administration of Test Items

Testing was conducted in a quiet room with minimal distractions. The researcher and student sat at the corner of a table such that the researcher was easily able to observe and converse with the student. Once rapport was established, the researcher introduced the test items using a standard introduction such as that used in the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). For example, “I will be asking you to answer a number of different questions today, such as estimating an answer to a math problem, or drawing a picture to show your understanding of a math problem. Some of the questions may be really easy, but others may be more difficult. It’s important that you try your best on both the easy and the harder items. Do you have any questions?” During testing, the researcher gave verbal praise to students, such as saying “I can see you’re working really hard!” such that students were recognized for their efforts. Each test item was administered precisely according to the format and directions described in the Pilot Items and in the WIAT-II manual. Upon completion of the test items, the researcher thanked the student for participating and answered any further questions.

Ethical Considerations

Participation in the present research project was completely voluntary and participants were able to remove their consent at any time. Following approval from Mount Saint Vincent University’s Ethics Review Board, initial contact was made with two schools for students with Learning Disabilities in Nova Scotia. Due to the time of year, both schools declined to participate in the study. Following, contact was made with the Learning Disabilities Association of Nova Scotia (LDANS), which is an organization for students with special needs in Nova Scotia. The purpose of the study was presented to the Executive Director at LDANS and administrative consent was received. LDANS staff identified students who fit the study’s criteria and approached families about the possibility of participating in the present study. Interested parents/legal guardians had the opportunity to ask questions of the researcher and signed consent forms to allow their children to participate in the study. Participation in the study was both confidential and voluntary. Although the researcher knows the identity of the participants, the final thesis does not contain information that in any way reveals the identity of a participant. Upon completion of data collection, raw data was coded such that no information that could identify a participant was present on the working data forms. Upon completion of the study, data will be stored in a locked cabinet in a locked office on the premises of Mount Saint Vincent University for five years.

Once parental consent was received by LDANS, the researcher contacted the parents/legal guardians to schedule times for the researcher to meet with the students. The purpose of the research was presented to the students in age-appropriate language and they were given the opportunity to ask questions. If the student wished to participate in the study, he/she signed an assent form documenting his/her desire to participate. Students understood that their participation was completely voluntary, that there would be no penalty for refusing to participate, nor for withdrawing their participation at any time during the study.

Selection of Participants

Randomization serves the purpose of creating roughly equal groups. However, when using a sample population of less than 5 participants, randomization may have insufficient power to produce roughly equal groups (Smith & Davis, 2003). In the present study, using a sample of 5 participants was the target and was justified due to time and availability constraints of participants, as well as the purpose of the study.

Recall that the purpose of the study was to test examples of assessment items that access a student's level of mathematical understanding. Therefore, a within-subjects comparison was preferable since it would hold variables such as IQ and exposure to quality instruction constant, since these factors remained constant throughout each of the measures. Also, this controls for between-subjects factors that could account for the within-subjects variance.

All participants were enrolled in an organization for students with special needs. Piloting the test items with students diagnosed with learning disabilities was preferred since doing so ensured IQ to be in the Average range of ability (i.e. Standard Score FSIQ = 90 – 109). Since current local diagnostic practices pertaining to Learning Disabilities employs the Ability-Achievement Discrepancy model, a student's intellectual functioning must fall within the Average range in order to attain Learning Disability (LD) status. Also, if participants have been diagnosed with LD and are able to perform well on the Pilot Items but not the original WIAT-II items, this finding would demonstrate students' understanding of mathematical concepts that is not captured with this measure. The grade level ranged from completion of grades 5 and 6 within the Halifax Regional Municipality.

In order to consider generalizability of the results in terms of gender, a sample, which included both male and female participants, was preferred. Additional demographic information collected includes gender, age, ethnicity, and any diagnosed medical or mental health disorders (e.g. LD, ADHD). The parents of the participants provided the demographic information. Past research has suggested that Japanese and Chinese children outperform their European and North American counterparts on tests of mathematical skills (Devlin, 2000). One reason suggested for this phenomenon is that the counting rules of the Chinese and Japanese languages appears to facilitate the learning of place value, counting, and computation more so than the English, French, or Spanish languages (Devlin, 2000). As such, it was of interest to collect data regarding the ethnic background of each of the participants in the study. In terms of age, this information is of interest in that older students may perform better than younger students, perhaps due to the greater number of years of instruction and experience with the subject matter. Although the Pilot Items are designed to be accessible to students ranging in grade from 4 to 6, age is an important variable that may influence the results of the study and therefore was in need of consideration. Information on diagnosed medical and/or mental health conditions allowed the researcher to consider the implications that diagnoses may have on the results of the study. A diagnosis of MD was preferred, as it would have allowed for a purer sample and increased the strength of the research findings, however, was unavailable.

A within subjects comparison helps to control for error variability, thereby allowing the researcher to attribute more of the cause of the difference in type of information collected to the independent variable (Smith & Davis, 2003). Error variability refers to other factors, such as individual differences between participants that may cause a change in the dependent variable (Smith & Davis, 2003). In this case, because the same subjects are tested with two different measures, and their individual characteristics remain constant over both measures, any variability in the quality of information garnered is presumed to be due to the test items, rather than due to other factors. Testing two separate groups of students and comparing the data gathered by each of the testing tools would not provide a reliable source of information, since the variability in the type of information collected could easily be due to individual differences between the participants. In a within-subjects comparison design, these factors remain constant, thereby allowing the researcher to attribute more of the variance to the independent variable, in this case, the type of assessment measure.

Data Analysis

The first research question seeks to determine whether or not there is a difference between the types of information collected by each of the assessment tools. The independent variable is the type of assessment measure and the dependent variable is the information gathered by the assessment tools. Researchers often prefer a single independent variable when new areas are being investigated (Smith & Davis, 2003). Two types of measures were used to answer this question. In the first measure, participants were administered five selected items from the WIAT-II Numerical Operations and Math Reasoning subtests. In the second measure, participants were administered the newly developed Pilot Items. A paired samples *t* test is implicated for this type of experimental design, since there is one independent variable and two dependent variables that are on a continuous scale. Since participants were tested twice, and the tests yield continuous data, a paired *t* test is most appropriate to analyze the data collected. However, due to the limited number of students who participated in the study, the results of the *t* test were rendered meaningless. As such, differences between participants' performance on each of the item sets are reported qualitatively. Table 2 presents a pictorial summary of the research design for the first research question.

Table 2.
Experimental Design of Research Question 1
Independent Variable (Assessment Tool)

Measure 1	Measure 2
Original WIAT-II Items	Pilot Items

The second research question seeks to determine the relationship between calculation proficiency (as measured by the WIAT-II items) and mathematical understanding (as measured by the pilot items). To measure this relationship, a Pearson product moment correlation would be most appropriate. If the analysis revealed that the Pilot Items and WIAT-II items had a strong positive correlation, this would suggest that both item sets measure the same construct and indicate convergent validity between the assessment tools. Convergent validity occurs when two different measures assessing the same construct are highly correlated (Crocker & Algina, 1986). In the context of this study, a moderate to strong negative correlation would indicate that calculation proficiency does not necessarily demonstrate mathematical understanding. As with the first research question, due to the small number of participants who took part in the study, the results of the correlation became of little value. As a result, the relationship between the two item sets is discussed qualitatively.

The third research question seeks to assess critically the mathematics subtests, Numerical Operations and Math Reasoning, from the WIAT-II. This measure of academic achievement is widely used by school psychologists in the Halifax Regional School Board, across the province of Nova Scotia, and throughout Canada and the United States. To evaluate these subtests, a rubric was developed based on the NCTM's Large Scale Assessment Tool as well as the NCTM's (1995) Assessment Standards. The

WIAT-II mathematics subtests were evaluated based on adherence to the standards outlined by the NCTM. Results are reported qualitatively.

The fourth research question seeks to determine if any of the Pilot Items measure the areas described by the NCTM and detailed in the rubric. The Pilot Items were compared to the Evaluation Rubric and results are reported qualitatively.

Conclusion

It is believed that both the Pilot Items and the Evaluation Rubric are consistent with today's efforts to emphasize understanding in the mathematics classroom. Contrasted with previous generations' tendency to focus on the correct answer, today's mathematical classroom is as much focused on mathematical understanding as it is on finding producing correct solutions (NCTM, 2000). It was hypothesized that this research would show that assessment tools lag behind today's classroom in terms of the emphasis unduly placed on computation problems to represent students' overall mathematical skills, knowledge, and understanding. A review of existing assessment tools for diagnosing MD suggests there is a need for test items that tap into students' mathematical thinking and understanding. As such, the present study seeks to pilot seven alternative test items, developed from the existing WIAT-II items and modified based on the work of Ruch and colleagues (1925), Nancy Jordan (2006), David Geary (1999, 2001, 2005), Lambdin and Walcott (2007), Jitendra and DiPipi (2003), Carpenter and Leher (1999), and the National Council of Teachers of Mathematics (1995, 2000, 2006). The Pilot Items are aimed at measuring not only students' ability to solve calculation problems correctly, but also their conceptual knowledge and understanding of the problem. It was hypothesized that these new items would assist practitioners in three ways: (1) they would provide more detailed information and facilitate understanding of a child's particular mathematical strengths and weaknesses; (2) they would also, therefore, better assist in program and intervention planning, and (3) they would help practitioners call to mind the limitations of information gathered by current testing tools, thereby better distinguishing between computational difficulty and MD. The Pilot Items test the effectiveness and feasibility of an easily accessible item set in order to provide a greater wealth and depth of information for the practitioner and ultimately to serve better the needs of the child. The present research project also answers questions regarding the relationship between calculation proficiency and mathematical understanding, the type of information provided by the WIAT-II mathematics subtests compared to the guidelines of the NCTM, and the extent to which the Pilot Items measure up against the NCTM standards articulated in the rubric.

Chapter 5: Results

Introduction

With the present study, I endeavoured to answer four research questions relating to mathematical thinking and understanding. First, I describe the demographic characteristics of the participants who took part in the study and then I outline the findings for each of the research questions. With the first research question, I wanted to determine if there was a meaningful difference between the information collected by the WIAT-II items and the Pilot Items. In order to define further the knowledge between these two assessment measures, I attempted to establish the relationship between mathematical thinking and understanding and calculation skills. Then, I sought to discover the extent to which the WIAT-II mathematics subtests and the Pilot Items aligned with the assessment standards of the NCTM, as detailed in the Evaluation Rubric. I present my findings in the following section.

Participant Characteristics

A total of two participants took part in the present study. Participant A was a Caucasian female, aged 12 years 8 months who had been diagnosed with Tourette's syndrome and a Language-Based Learning Disability. Total testing time for Participant A was 10 minutes for both subtests. Participant B was a Caucasian male, aged 11 years 4 months who had been diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) and Developmental Delay. Total testing time for Participant B was 16 minutes for both subtests. A summary of participant characteristics is presented in Table 3.

Table 3.
Participant Characteristics

	Age	Sex	Ethnic Background	Diagnosed Medical/Mental Health Disorders	Length of Testing Time
Participant A	12 years, 8 months	Female	Caucasian	Tourette's Syndrome, Language-Based Learning Disability	10 minutes
Participant B	11 years, 4 months	Male	Caucasian	Developmental Delay, ADHD	16 minutes

With regard to the participants, two unexpected limitations occurred. First, there were fewer participants in the study than was originally planned, making it impossible to conduct the anticipated statistical analyses. Second, although the criteria for the study originally designated a diagnosed Learning Disability as a necessary component for participation, a communication error resulted in testing being completed with Participant B who had been diagnosed with Developmental Delay rather than a Learning Disability. The error was not discovered until the testing had been completed. Upon consideration of the situation, it was determined that although the data was not from the anticipated population, there would nevertheless be value to include it in the study. Despite the fact that Participant B was diagnosed with Developmental Delay rather than LD, it was of interest to ascertain if differences would be detected between performance on the Pilot Items compared to performance on the WIAT-II items.

Research Question 1

With the first research question, I sought to determine if there was a significant difference between the type of information collected by five items selected from the WIAT-II mathematics subtests (Numerical Operations and Math Reasoning) and the seven newly developed Pilot Items. Several qualitative differences appeared to exist between the two item sets, including the target skills measured by each item, the manner in which credit is awarded, and the manner in which problems are presented to students. Whereas the Pilot Items were developed with the goal of measuring mathematical thinking and understanding (e.g. describe in your own words what subtraction means), the selected WIAT-II items were aimed at measuring basic paper-and-pencil computation skills (e.g. 42 + 15), and solving word problems involving simple calculation (e.g. Johnny had 5 apples. His mother gave him 10 more. How many did he have then?). Items from the WIAT-II subtest are coded as either 1 (correct) or 0 (incorrect) whereas the Pilot Items offered students the opportunity to receive partial credit and were coded as 2 (best possible response), 1 (partially correct but missing information), or 0 (incorrect). Also, the Pilot Items were presented to students in a manner that allowed them to demonstrate and receive credit for partial understanding of a given problem (e.g. identifying the correct operation necessary for solving 120-17) whereas the WIAT-II items tended to focus on the final answer (e.g. producing the correct response for the problem 120-17). Due to the differences described above between the two item sets, it was suspected that a significant difference would be uncovered.

Conversely, the results of the first research question produced a statistically insignificant result. Data from each participant was separated into two groups relating to performance on the WIAT-II items and the Pilot Items. The average of each participant's score was calculated for both groups respectively. A paired samples *t*-test was conducted to determine if a meaningful difference between the two groups existed. A significant difference between the two groups (Average score on Pilot Items vs. Average score on WIAT-II items) was not found. The difference between participants' performance on the Pilot Items and the WIAT-II items did not reach statistical significance, t (1) = 3.174, p > .05. However, due to the small number of participants, it was determined that this result could be misleading and thus the data were reanalyzed qualitatively.

Participants' performances are analyzed item-by-item and patterns will be discussed in the following section. Table 4, found on page 86, presents a report of participants' scores on each of the items, including a description of the type of skill assessed by each question. The assessment process began with the Pilot Items, as it was suspected that these types of questions would put participants more at ease given that they are not traditional mathematics test items.

Item 1.1 asks participants to look at the problem 47 + 64 and to describe in their own words what the problem is asking them to do. A two-point response would have included a response that made reference to each of the numbers, the operation, and indicated that the two numbers should be combined to find a total. A one-point response would emphasize procedure (e.g. first you add the right column). To receive zero points, the participant would describe an operation other than addition. Both participants received a score of 1/2 on this item since responses indicated that adding was the necessary operation but emphasized the procedure for solving the problem. One difficulty with this item is that younger children might not possess the language skills necessary to articulate a complete description of what the given problem is asking, thereby masking understanding. In the future, this type of problem could be avoided by having students select the most appropriate response from a list of response choices.

Item 1.2 asks students to select a good estimate from a list of response choices to the algorithm in the previous question (47 + 64). One of the distractor items is visually similar to the correct response and was selected by Participant B, while Participant A chose the correct estimate. This item is useful in that more distractor items could be created and linked to common error types that students make. The pattern of responses could then be linked to appropriate interventions.

Item 2.1 simply asks the student to name the necessary operation for a given problem from a list of response choices. Both participants correctly responded to this item. Item 2.2 delves deeper in asking students to provide a definition of the given operation in their own words. Participant A received full credit for this item while Participant B received partial credit. This item may also be negatively impacted by difficulties by a lack of verbal skill by which to articulate understanding. This problem might be revised in the future to have students select the most appropriate response from a list of choices, as with the first item. Although there is merit in having students produce their own definition of a given operation, the language demands may be too great for younger students. For Item 2.3, students are asked to create a word problem based on the algorithm from Item 2.1 (110-17). Participant A provided an excellent response that concisely demonstrates the target quantities and operation, “She had 110 apples and she shared 17 of them”. Use of the word *shared* clearly shows knowledge of the intended operation. Participant B’s response earned him partial credit for correctly representing the operation, although not the given numbers. Both participants’ responses demonstrate some facility with the concept of subtracting, despite the fact that neither was able to correctly solve a traditional subtraction problem. This duality lends further support to the suggestion that the data collected by the Pilot Items and WIAT-II items may indeed measure different constructs.

For the remainder of the items, Participant B was unable to produce either a partially or fully correct response. Item 3.1 is a basic word problem involving averages, except that it contains some extraneous information. It is not intended that students solve the problem, but rather indicate what information is *necessary* to solve the problem. To receive 2 points, a student would be required to indicate the necessary information and not name any unessential information. Participant A knew that the numbers needed to be added but could not provide any further details. This item could perhaps be improved by asking the student to identify the piece of information that is *not* necessary for solving the problem, as leaving the question as it is might emphasize procedural knowledge moreso than understanding.

The final Pilot Item invites students to represent their understanding of a given word problem either by drawing a picture or by writing a number sentence (e.g. 104 ÷ 8). Participant A chose to represent her knowledge by using a number sentence and completed the operation and part of the quantities correctly but chose the wrong divisor (104 ÷ 15). As such, she received partial credit. The nature of the word problem suggests that students who chose to draw would likely draw a calendar, thereby facilitating the examiner's awareness of their thinking procedure. However, further scoring options would probably arise once the results of a standardization sample were analyzed.

Both participants struggled with the WIAT-II items. Between the two participants, only one item was answered correctly. The examiner's record form indicates that the skill measured by this particular item is one in which the student creates and solves a division problem (Wechsler, 2005). Participant A provided a correct solution to this problem. In some ways, this result is surprising given that division is generally considered to be the most difficult of the four operations, therefore, it is interesting that Participant A correctly responded to a division problem but not to traditional multiplication, subtraction, or addition problems. One possible explanation could be that since Participant A had completed grade 6 in the preceding few weeks, she had likely spent time practicing division skills more recently than multiplication, subtraction, or addition. This result suggests that despite the great amount of time typically spent on the development of calculation skills (Romberg & Kaput, 1999), students are still leaving elementary school without basic computation knowledge.

An interesting phenomenon is revealed when the responses of Participant B are studied closely. Participant B's answers to each of the WIAT-II items received a score of 0/5 whereas solutions on the Pilot Items received a total score of 5/14. Although neither score is strong, the question becomes one of which assessment tool provides more meaningful or actionable information. Performance on the WIAT-II items suggests a complete lack of knowledge while performance on the Pilot Items clearly demonstrates *some* knowledge, albeit incomplete. For example, Participant B was unable to solve the problem 120-15 correctly. However, on the Pilot Items, he correctly identified that subtraction was the operation needed to solve the problem 110-17. Also, he was able to describe in his own words what subtraction meant ("taking away"), for which he received a score of 1 out of a possible 2 points. Further, he was able to make up a word problem using the algorithm 110-17 in which he correctly represented the necessary operation, although not the given quantities. This example demonstrates that Participant B did not have a complete lack of knowledge regarding subtraction problems, as suggested by performance on the WIAT-II items, but rather that knowledge could be uncovered by using an alternative set of test questions, such as the Pilot Items. When it comes to planning interventions, which, recall is a main outcome of conducting assessments in the first place, it would be beneficial to have information regarding what a student *does* know, as much as it is helpful to have information about what a student *does not* know. By more closely examining Participant B's responses to both the WIAT-II items and the Pilot Items, it becomes obvious that there is in fact a difference between the types of information collected by these two item sets.

Participant A showed a similar pattern of responses to that of Participant B. On the WIAT-II items, Participant A received a total score of 1/5 while a score of 11/14 is earned on the Pilot Items. Participant A received at least partial credit for each of the Pilot Items, while only one correct response was produced on the WIAT-II items. On the WIAT-II items, Participant A was unable to correctly solve the problems 37+54 or 120-15, suggesting a lack of knowledge for both addition and subtraction problems.

However, on the Pilot Items, Participant A was able to describe the process of addition, to provide a reasonable estimate to a given addition problem, to identify the correct operation for a subtraction problem, to provide an acceptable definition of subtraction in her own words, to make up her own word problem where she correctly represented both quantity and operation, to identify important information in a given word problem involving averages, and to produce a partially correct algorithm for a word problem involving division. Clearly, knowledge of addition and subtraction existed and was accessed only when an alternative type of question was asked. Both participants received higher scores on the Pilot Items than they did on the WIAT-II items, lending support to the hypothesis that there is a difference in the type of data collected by each of the assessment tools. Table 4 presents an item-by-item report of the skills assessed and participants' performance on each item.

Table 4.
Item-by-Item Report of Participants' Scores on the Pilot Items and WIAT-II Items

Skill	Item #	Participant A	Participant B
Ability to demonstrate understanding of a problem beyond simple computation	1.1	1/2	1/2
Ability to estimate the reasonableness of one's response	1.2	2/2	0/2
Ability to identify the appropriate operation to solve a given problem	2.1	2/2	2/2
Ability to demonstrate understanding of a problem beyond simple computation	2.2	2/2	1/2
Ability to generate a word problem from given algorithm	2.3	2/2	1/2
Ability to inhibit unessential pieces of information	3.1	1/2	0/2
Ability to represent a given problem using correct	4.1	1/2	0/2

number and operation, or by drawing a picture			
Addition – multi digits, with renaming	5.1	0/1	0/1
Subtraction – multi digits, with regrouping	6.1	0/1	0/1
Create and solve multiplication problem using whole numbers	7.1	0/1	0/1
Create and solve division problem using whole numbers	8.1	1/1	0/1
Create and solve multiplication and division problems using whole numbers	9.1	0/1	0/1

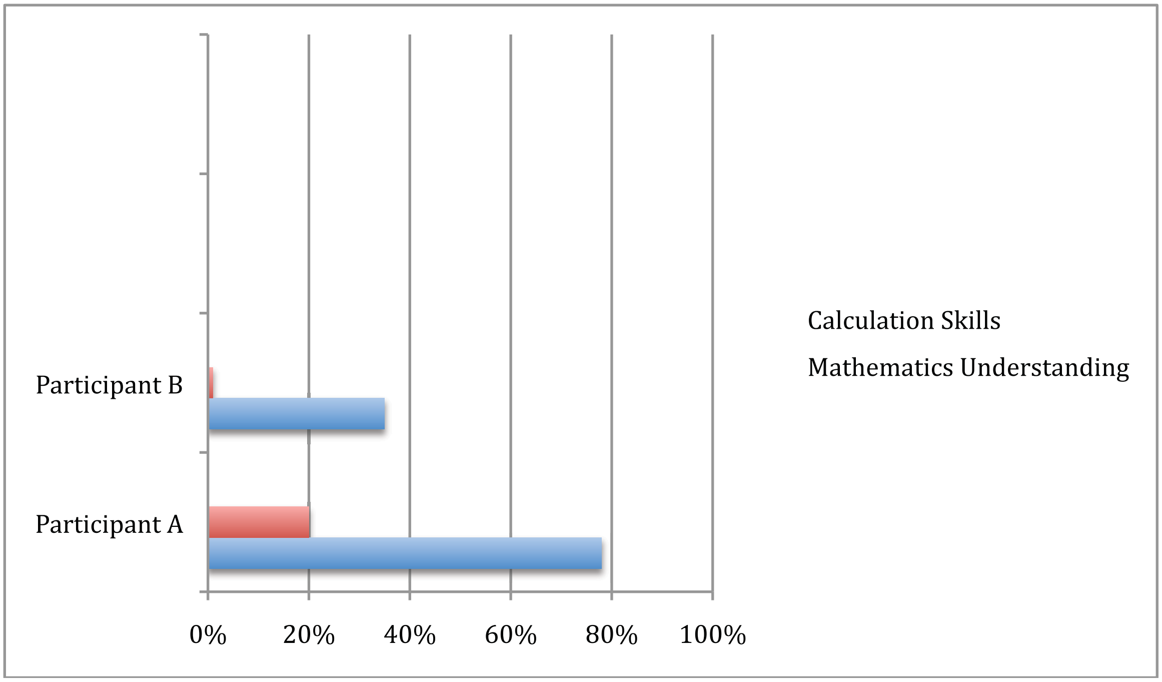
As a result of the differences in types of questions asked, mathematical knowledge was either highlighted or concealed. The Pilot Items allowed students to demonstrate their mathematical knowledge beyond that accounted for by traditional questions, such as those selected from the WIAT-II. Performance on the WIAT-II suggested a severe lack of skill in the case of both participants; however, performance on the Pilot Items verified that knowledge could be accessed by sub-dividing questions into smaller sub-skills, thereby allowing students to show some understanding, despite that knowledge being imperfect. Based on the results of the two participants in the present study, there does appear to be a difference between the types of information collected by each of the assessment tools. However, further research would need to confirm this finding with a larger sample size. From the results of the present study, it seems that a great deal of care is necessary when mathematical test questions are developed in order to sufficiently capture students' mathematical skill, knowledge, and understanding.

Research Question 2

The purpose of the second research question was to determine the relationship between mathematical understanding (as measured by the Pilot Items) and calculation skills (as measured by the WIAT-II items). With the first research question, I attempted to determine *if* there was a difference between the two item sets, and with the second research question, I endeavoured to *refine* our understanding of the relationship between the two item sets. The result from the first research question suggests that the Pilot Items and WIAT-II items do measure different constructs but more study is required in order to draw a firm conclusion. In order to investigate further the relationship between the two item sets, a Pearson product moment correlation was calculated. The statistical analyses revealed that the relationship between the WIAT-II items and the Pilot Items had a perfect positive correlation, r (1) = 1.00, $p < .05$. Very likely, this result was strongly influenced by the small sample size. It would be interesting to determine what result would be produced in a larger study. The result of this analysis suggests that both understanding and calculation prowess are important aspects of overall mathematical knowledge and skill and we would expect them to be at least moderately correlated.

With data from only two participants, it becomes difficult to provide a meaningful conclusion to this research question. However, some hypotheses may be developed as an outcome of the results at hand. Participant A answered only 20% of the WIAT-II questions correctly while responses to 78% of the Pilot Items received either partial or full credit. Conversely, Participant B answered none of the WIAT-II items correctly and 35% of the Pilot Items correctly. From these results, a hypothesis could be developed that as calculation skills increase (represented by performance on the WIAT-II items), so does mathematical understanding (as represented by performance on the Pilot Items) given that the student who performed better on the calculation problems also showed a significant jump in understanding. Alternatively, the data seem to support the hypothesis that strong mathematical understanding is not necessarily translated into strong calculation skills. It seems that mathematical understanding may exist in the face of poor calculation skills, suggesting that there can be a difference in these two areas. The implications in programming for students who show this pattern of strong or some mathematical understanding with little calculation skill would likely be different from a student who showed strong calculation skills but poor understanding. While certainly it is difficult to draw any firm conclusions from such a small sample size, it would appear that the greater wealth of information provided by a measure such as the Pilot Items puts educational professionals in a better position to successfully remediate the individual differences faced by students struggling with mathematics. Figure 6 presents a pictorial view of the results obtained by participants on the WIAT-II items compared to performance on the Pilot Items.

Figure 6.
A Pictorial View of Performance on the WIAT-II Items Compared to Performance on the Pilot Items



At the outset of this research project, I posed the question as to what constitutes mathematical skill and challenged the notion that calculation skills were a good overall measure of mathematical knowledge, skill, and understanding. By examining participants' responses to problems from both item sets, a pattern began to emerge. The pattern is one in which poor calculation skills are evident, and yet, there is a greater wealth of mathematical understanding.

Romberg and Kaput (1999) state that within the first seven to eight years of mathematics education, students are exposed to little mathematics beyond computation. While clearly participants have developed some conceptual knowledge, the finding that students who are entering grades 6 and 7 respectively are unable to solve basic addition, subtraction, and multiplication problems correctly is cause for concern. However, given that Participant B had been diagnosed with Developmental

Delay, this could explain the poor calculation skills. Perhaps a sample of students who did not have LD would provide an interesting comparison group. Despite the fact that so much time is spent on developing calculation skills, still, some students at the late elementary and early junior high levels remain unable to perform even the most basic arithmetic. Previous research (e.g. Moss & Case, 1999; Fennema et al., 1993) has demonstrated that computation skills can be improved when teachers work to augment students' existing knowledge bases and when strategies are taught for solving calculation problems rather than the development of rote knowledge. The results of the present study suggest that the time has come for a shift in teaching strategies. While the development of computation skills should not be the only goal, most students should nevertheless be competent in solving such problems by the time they reach late elementary or junior high school.

With the second research question, it was my goal to describe in more detail the relationship between calculation skills and mathematical understanding. Although the small number of participants prevented any meaningful statistics from being calculated, the data nevertheless yielded some interesting results. Both participants performed better on the measure of mathematical understanding compared to performance on calculation skills. This result demonstrates that mathematical understanding may exist despite sub-par calculation skills. The specific nature of students' mathematical difficulties is important information for educational professionals to possess in order to effectively remediate students' mathematical difficulties. Although the second research question was not carried out exactly according to plan, the data show that mathematical understanding may exist despite poor calculation skills.

Research Question 3

With the third research question, I sought to determine the extent to which the WIAT-II mathematics subtests (Numerical Operations and Math Reasoning) aligned with the (1995) Assessment Standards of the NCTM and Large Scale Assessment Tool (NCTM, 2006). This goal was accomplished by applying the factors contained within the Evaluation Rubric to the mathematics subtests of the WIAT-II. Five factors emerged as pertinent to the present study; namely, Content, Process, Knowledge Representation, Time Constraints, and Implications for Learners. Each subtest was evaluated along the five above-mentioned strands and given an overall rating of either Exceeds Expectation, Meets Expectation, or Needs Improvement. Results relating to the Numerical Operations subtest will be reported first, followed by the results for the Math Reasoning subtest.

The first factor relates to the Content of the subtest, referring to the range (i.e. broad, basic, or limited) of mathematical processes and concepts measured by the test. For the purposes of this project, the construct of mathematical processes refers to the execution of calculations whereas the construct of mathematical concepts refers to the understanding of a mathematical idea. For example, a student correctly carries out the procedure for solving a long-division equation (mathematical process) and also has a solid understanding of division itself (mathematical concept). Of course, one of the major premises of this paper is the idea that a student may possess the skill to apply a mathematical procedure yet have little or no understanding of the concept. Alternatively, conceptual knowledge may exist despite a failure to effectuate a mathematical procedure correctly. The best test will be one which accounts for both mathematical processes *and* concepts.

In the examiner's manual, the WIAT-II test developer describes the Numerical Operations subtest as assessing "the ability to identify and write numbers, count using 1:1 correspondence, and solve written calculation problems and simple equations involving the basic operations of addition, subtraction, multiplication, and division" (Wechsler, 2005, p. 59). Upon reviewing the Numerical Operations subtest, it was determined that a broad range of mathematical processes are indeed assessed using this tool. In concordance with the test developer's description of the Numerical Operations subtest, fewer mathematical concepts appear to be assessed with this subtest. While a rudimentary conceptual understanding is presumed to be necessary to complete the problems correctly, it is conceivable that a student with very little understanding is able to complete the problems correctly. In this scenario, the test would overestimate a child's true mathematical prowess. A list of the skills measured by the Numerical Operations subtest, as described in the WIAT-II protocol, appear in Table 5. As can be seen upon viewing Table 5, a drawback to this subtest may be the relatively small number of items measuring each skill. This phenomenon has been reported in previous research (Parmar et al., 1996). Due to the broad range of processes yet relatively smaller range of concepts measured by the Numerical Operations subtest, it receives a judgement of Meets Expectation (measures basic mathematical processes and concepts).

Table 5.
Skills Measured by WIAT-II Numerical Operations Subtest

Broad Skill	Specific Skill	Number of Items Measuring Skill
Number discrimination		1
Identifying missing number in rote count		1
Writing single and double digit numbers		1
Counting to 8 by rote		1
Writing number to correspond with rote counting		1
Addition	Basic facts	2
	Multi-digits, no renaming	1
	Multi-digits, with renaming	3
	Using single digit decimals	1
	Negative integers	1
Subtraction	Basic facts	2
	Multi-digits, no regrouping	1
	Multi-digits, with regrouping	2
	With regrouping using decimals	2
	Simple fractions with common denominators	1
	Simple fractions with different denominators	1
Multiplication	Basic facts	1
	Multi-digit and single digit	1
	Multi-digits	1
	Simple fractions	1
	Using decimals	1
	Simple fractions and whole numbers	1
Division	Basic facts	1
	Single digit divisor, no regrouping	1

	Single digit divisor, with regrouping	1
	Multi-digits, with regrouping	1
	Simple fractions	1
	Simple fraction by whole number	1
	With decimals	1
Calculating with exponents		2
Calculating square root		1
Calculating percent		2
Solving simple algebraic equations		2
Calculating pi		1
Calculating after applying order of operations		1
Solving complex algebraic equations		1
Calculating square root of exponents		1
Using basic geometry		1

With the second factor, the variable of interest is Process, or the extent to which the test facilitates the examiner's understanding of the student's process for solving mathematics problems. Ideally, the test would enhance such knowledge and would award points for a student's demonstration of correct reasoning. In the least favourable scenario, the test would award points solely for the correct response. To receive an intermediate-level ranking, the test would provide an informal or qualitative assessment of the student's reasoning. The WIAT-II protocol encourages examiners to record qualitative observations by using the provided checklist, reproduced in Table 6.

Table 6.
WIAT-II Numerical Operations Qualitative Observations Checklist

Writes incorrectly formed or reversed numerals
Uses fingers/aids for counting or calculating
Demonstrates automatized math facts when completing calculations
Converts math problem from horizontal to vertical presentation prior to calculation
Uses place value correctly during calculation so as to avoid "spatial" errors
Makes sequential errors in procedures for multi-step calculations
Follows sequential procedures correctly, but makes math fact errors

The student's process is of interest to examiners in that it may shed light on aspects of the problem that are well-known by the student and others which require further development. This information is valuable for intervention planning because time may be spent more effectively when it is focused on building students' knowledge based on what is already known and on targeting specific problem areas (for examples, see Moss & Case, 1999). Further, "understanding the reason for the weakness is critical in determining the most effective intervention" (Dawson & Guare, 2007; p.11). Although a qualitative observations checklist is both provided and encouraged by the test developer, the extent to which it enhances the examiner's awareness of the student's process is limited. While the information provided may assist the examiner in planning certain interventions, overall, the checklist is not specific enough to each item as to the most common types of pitfalls that students encounter. For example, a student who struggles to form numerals is clearly spending undue effort on an unintended task. In contrast, a student who is unable to solve multi-digit subtraction problems involving regrouping may be plagued by a number of difficulties not found in the checklist. Since addition, subtraction, multiplication, and division problems encompass the bulk of the Numerical Operations subtest, it would be of benefit to examiners, and ultimately to students, if more specific information were provided regarding common error-types in these domains. Further, if suggestions for remediation were included based on error-types, this would be exemplary. In addition, the checklist does not provide prompts that help direct the examiner to probe further into the student's rationale, which could potentially highlight strengths and weaknesses. Beyond the qualitative checklist, the Numerical Operations subtest does little to enhance the examiner's understanding of the student's process nor does it credit students for correct reasoning. As such, it earns a rating of Needs Improvement (awards points only for correctness of final answer).

The third factor of interest is Knowledge Representation, or the extent to which the test provides opportunities for students to demonstrate their mathematical knowledge, skill, and understanding in non-traditional formats (e.g. drawing a picture). This variable is of interest in that it may provide a venue by which to tap into students' understanding and conceptual knowledge. Non-traditional formats for question response provide greater opportunities for students to demonstrate knowledge, even if the knowledge is incomplete. Given that the purpose of the Numerical Operations subtest is primarily one in which calculation skills are measured, unsurprisingly, there are no examples of opportunities to demonstrate understanding in non-traditional formats. Test items are presented in a traditional paper-and-pencil format for solving computation problems. No manipulatives or prompts to solve a problem in an alternative manner are provided, however, the student is permitted to use scratch paper at any time to work a problem. Students do not receive points for correct reasoning on the scratch paper. Advantages contained within the test include that test items are printed in a font-size that is easy to read and sufficient white space is provided by which to work out problems. Overall, the Numerical Operations subtest earns a rating of Needs Improvement on the Knowledge Representation variable (provides limited or no opportunity for students to represent their knowledge in non-traditional formats).

The fourth factor is Time Constraints, or the extent to which students are provided with sufficient time to solve the given problems. This variable is of interest in that time constraints may add undue pressure to the student, thereby hindering performance unnecessarily. Although some tests seek to measure the fluency with which problems are completed, ideally, students would be provided with as much time as they require in order to solve problems. At the opposite end of the continuum, all tests would be timed. Intermediately, some time restraints would be imposed. Items on the Numerical Operations subtest are presented in a non-timed fashion. The student is instructed to complete as many items as he or she is able and to alert the examiner upon completion. As such, a rating of Exceeds Expectation is awarded on this strand (students are provided with ample time to complete test items).

The fifth and final factor relates to Implications for Learners, or the extent to which the test facilitates the development of effective interventions based on the student's specific strengths and weaknesses. To earn a rating of Exceeds Expectation, the test would incorporate a method for testing the effectiveness of interventions during the assessment. To earn an intermediate rating of Meets Expectation, the test would have to suggest areas for remediation based on the student's performance on general strands; however, strategies would not be tested for effectiveness during the assessment. Finally, a rating of Needs Improvement would be incurred if test results do not link easily to intervention with no strategies tested for effectiveness during the assessment. This variable is of particular interest because as previously stated; a key purpose of assessment is to develop effective interventions (Dawson & Guare, 2004). Due to the small number of items measuring each construct, it can be difficult to get an idea of a student's specific strengths and weaknesses although a general pattern can usually be detected. Strategies are not tested for effectiveness during assessment; however, strategies based on the student's profile are available through supplementary report writing software through the test's publisher. As such, a rating of Meets Expectation is earned for this factor (performance on general strands suggests areas for remediation but strategies are not tested for effectiveness during assessment).

Turning now to the Math Reasoning subtest and examining the first factor, Content, which was assessed using the same criteria as described above for the Numerical Operations subtest. The test developer describes the Math Reasoning subtest as "a series of problems with both verbal and visual prompts that assess the ability to reason mathematically. The examinee counts, identifies geometric shapes, and solves single and multi-step word problems, including items related to time, money, and measurement. The examinee solves problems with whole numbers, fractions or decimals, interprets graphs, identifies mathematical patterns, and solves problems related to statistics and probability" (Wechsler, 2005, p. 68). It appears that this subtest, moreso than the Numerical Operations subtest, is designed to assess mathematical thinking and understanding. Unlike the Numerical Operations subtest, a greater number of items exist to measure a student's skill in a particular field. However, a smaller number of skills are measured on the Math Reasoning subtest than the

Numerical Operations subtest. Indeed, 38 subskills are listed in the Numerical Operations subtest compared to 14 subskills for the Math Reasoning subtest. On either test, the same subskill may be assessed at multiple points with increasing levels of difficulty, given that the tests are intended for use with people ranging in age from 4 to 85 (Wechsler, 2005). A list of the skills measured by the Math Reasoning subtest, as described in the WIAT-II protocol, appear in Table 7. It is unclear as to why relatively fewer subskills are measured on the Math Reasoning subtest compared to the Numerical Operations subtest. Previous research on mathematics assessment tools (e.g. Parmar et al., 1996) have criticized tests for focusing too heavily on unimportant mathematics, namely, a disproportionate emphasis on calculation. As expected, many of the items on the Math Reasoning subtest require the student to use calculations in addition to "math reasoning" in order to receive credit. For example, a student would have to possess some *understanding* of the following problem in order to determine how to go about solving it, and would additionally require the skills to *compute* the necessary calculation:

> Find the average of Andrew's four spelling test scores: 100, 85, 90, and 80.

Given that the Math Reasoning subtest contains a relatively small number of skills being measured and that some of its items are somewhat impure, in that they involve supplementary skills that are not necessarily the targeted skill being measured, it earns a rating of Meets Expectation (measures basic mathematical processes and concepts).

Table 7.
Skills Measured by WIAT-II Math Reasoning Subtest

Specific Skill	Number of Items Measuring Skill
Use whole numbers to describe quantities	5
Use geometric and spatial reasoning to solve problems	3
Use grids or graphs to make comparisons, draw conclusions, or answer questions	5
Create and solve addition and subtraction problems using whole numbers	3
Use non-standard and standard units to measure	3
Use patterns to solve problems	4
Solve problems using money	6
Tell time and use time to compare and order events	4
Use quantities less than a whole	8
Use theoretical and experimental probability to draw conclusions, answer questions, and make predictions	3
Create and solve subtraction problems using money	2
Create and solve multiplication and division problems using whole numbers	4
Create and solve addition and division problems using whole numbers	1
Create and solve multiplication and division problems using money	1

Moving on to examine Process, the Math Reasoning subtest shares some similar features with the Numerical Operations subtest. First and foremost, a qualitative observations checklist is provided in the protocol by which examiners are encouraged to note the following considerations in Table 8.

Table 8.
WIAT-II Math Reasoning Qualitative Observations Checklist

Uses paper and pencil to complete problem
Organizes work on scratch paper to facilitate problem-solving
Uses concrete aids (e.g. fingers) for computation
Breaks multi-step problem into smaller units to obtain solution
Disregards component(s) of word problem that is not required for solution
Uses correct operation(s) to compute solution
Uses repeated addition as a substitute for multiplication when problem-solving
Uses repeated subtraction as a substitute for division when problem-solving
Employs use of an effective strategy (e.g. working backwards, drawing pictures, systematic guessing, or making a table) to solve a problem

As with the Numerical Operations subtest, students are permitted to use scratch paper at any time to assist in solving a given problem. The difficulty being that since questions generally call for a verbal response, much of the student's process is not visible to the examiner. As such, even if correct reasoning exists, the examiner may remain unaware of it and the student does not receive credit for it. Also, the problem with specificity of common error-types is present in the Math Reasoning subtest, as it was in the Numerical Operations subtest. Meaning, that the qualitative observations checklist does little to enhance the examiner's awareness of the student's process by linking certain observations to specific items, common pitfalls, and suggestions for remediation. In addition, students are awarded points only if the final response is correct and as such, the Math Reasoning subtest receives a rating of Needs Improvement on the Process variable (awards points only for correctness of final answer).

The next factor is Knowledge Representation. Recall that with this factor, I sought to evaluate the extent to which students are provided with multiple opportunities to demonstrate their knowledge and understanding of a given problem.

With earlier items intended for younger students, responses involve counting where the student is shown a picture and told that touching the pictures while counting is permitted, if the student desires. By testing the child's knowledge in this manner, the child is provided with the opportunity to use the pictures as an aid, or to simply respond if the picture is deemed unnecessary for solving the problem in the child's opinion. As the items progress in difficulty, word problems are presented to the child orally and are also printed on the test manual such that the child is able to follow along. Presumably, the problems are reproduced such that the task becomes as much a mathematics task, and as little a working memory task as possible. Although the picture cues may provide students with the opportunity to demonstrate knowledge that they otherwise would be unable to show, there are little if any creative, or non-traditional attempts at drawing knowledge from a student, even if that knowledge is incomplete. Creativity may be of underestimated importance if it were able to assist examiners in drawing out knowledge that otherwise students were unable to produce. One possible reason for the importance of creativity in knowledge representation is the anxiety and discomfort that can come about when one is asked question after question for which answers are unknown. The discontinue rules of many standardized tests are such that a child must fail up to seven consecutive items before moving on to the next subtest. While some children are unaware of these failures, others are highly cognizant of them and show signs of embarrassment. Certainly, there are things that examiners may do to alleviate such negative feelings, however, it remains plausible that without creative attempts to draw out a student's knowledge, he or she may feel unwilling to risk the possibility of producing an incorrect response. Conversely, if the student believes that he or she could confidently answer at least a portion of the problem, perhaps this risk would be more readily taken. These effects may be even more likely when working with children who suffer from mathematics anxiety. Due to the limited opportunities built into the test for obtaining the maximum of a student's mathematical knowledge, a judgement of Needs Improvement is given for the Knowledge Representation variable (provides limited or no opportunity for students to represent their knowledge in non-traditional formats).

The next factor, Time Constraints, considers the extent to which students are provided with ample time by which to solve given problems. For the Math Reasoning subtest, the examiner's manual of the WIAT-II suggests that examiners should proceed to the next item if students are not actively engaged in solving a problem following one minute after presentation of the item (Wechsler, 2005). In this way, students are provided with sufficient time to solve the problems but are not faced with undue pressure to solve problems for which they do not have the skills to do so. As such, a rating of Exceeds Expectation is assigned (students are provided with ample time to complete test items).

Finally, the Math Reasoning subtest was examined based on its Implications for Learners. Similar to the Numerical Operations subtest, general strengths and weaknesses can be derived by examining the student's pattern of passed and failed items. This information would then be combined with other sources of information, such as parent and teacher reports and classroom samples in order to form a more coherent picture of the child's skills and areas of need. However, strategies are not tested for effectiveness during the assessment and as such, their effectiveness with the student cannot be guaranteed. As such, a rating of Meets Expectation is most appropriate (performance on general strands suggests areas for remediation but strategies are not tested for effectiveness during assessment).

The purpose of the third research question was to examine critically an existing measure of mathematics achievement so as to provide an avenue through which revisions may be suggested. The outcome was that both mathematics subtests of the WIAT-II earned a rating of Exceeds Expectation on the Time Constraints factor, a rating of Meets Expectation on the Content and Implications for Learners factors, and a rating of Needs Improvement on the Process and Knowledge Representation factors. A summary of the results for the third research question is presented in Table 9.

Table 9.
Summary of Results for Research Question 3

Factor	Numerical Operations Rating	Math Reasoning Rating
Content	Meets Expectation	Meets Expectation
Process	Needs Improvement	Needs Improvement
Knowledge Representation	Needs Improvement	Needs Improvement
Time Constraints	Exceeds Expectation	Exceeds Expectation
Implications for Learners	Meets Expectation	Meets Expectation

Research Question 4

In an effort to remain as impartial as possible, the newly developed Pilot Items were subjected to the same scrutiny, as was the case with the more established item set, the WIAT-II. As such, the Pilot Items were examined using the Evaluation Rubric with the aim of determining the extent to which the Pilot Items aligned with the factors outlined by the NCTM to be of importance for assessing mathematics (i.e. Content, Process, Knowledge Representation, Time Constraints, and Implications for Learners).

The first factor, Content, proved to be one in which the Pilot Items did not fare as well as in other areas. However, this result is unsurprising given that the Pilot Items were developed with the purpose of testing the feasibility of an alternative item set aimed at measuring mathematical thinking and understanding. As such, the Pilot Items were not intended to be an exhaustive list, but rather, examples of items that might do a better job of tapping into students' mathematical knowledge bases. The Pilot Items measure skills that were identified in the literature review based on historical assessments of mathematics in combination with more modern constructs identified by prominent researchers within the field (e.g. David Geary, Nancy Jordan). A list of the skills measured by each of the Pilot Items appears in Table 10. Given that the pilot items represent only examples of possible test questions, a broader range of items would need to be developed if these types of test items were to be incorporated into standardized assessment tools. As such, the Pilot Items warrant a ranking of Needs Improvement in the area of Content (measures a limited range of mathematical processes and concepts).

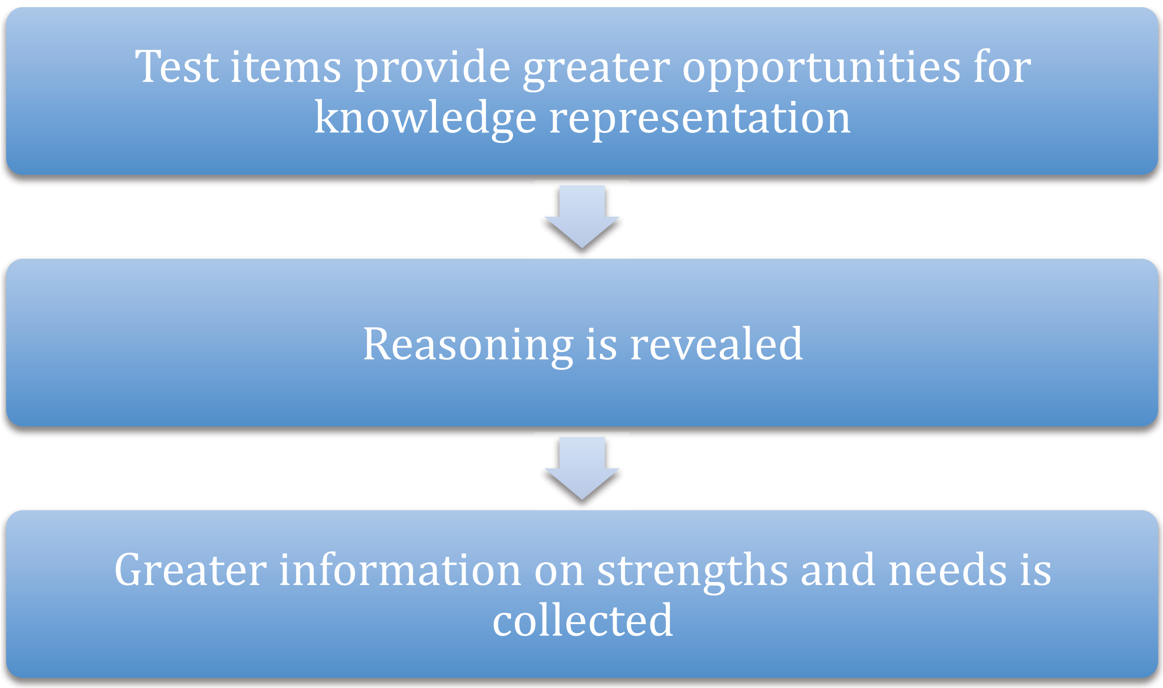
Table 10.
Skills Measured by the Pilot Items

Skill	**Item #**
Ability to demonstrate understanding of a problem beyond simple computation	1.1
Ability to estimate the reasonableness of one's response	1.2
Ability to identify the appropriate operation to solve a given problem	2.1
Ability to demonstrate understanding of a problem beyond simple computation	2.2
Ability to generate a word problem from given algorithm	2.3
Ability to inhibit unessential pieces of information	3.1

Ability to represent a given problem using correct numbers/operation, or by drawing a picture	4.1

The second factor, Process, is in many ways the crux of this research project. One of the main objections to the currently used mathematics assessment tools is the all-or-nothing classification of knowledge. Alternatively, the Pilot Items deconstruct traditional mathematics problems such that students are provided with greater opportunities to represent their knowledge, even if it is incomplete. The Process variable is not interested in measuring students' ability to memorize procedures, but rather with test questions that capture students' mathematical thinking and reasoning processes. Certainly, there is some overlap between the Process and Knowledge Representation variables. By providing students with more non-traditional formats to represent their understanding of problems, reasoning is revealed. The point of difference for the purposes of this evaluation is that non-traditional formats are the foci on the Knowledge Representation variable whereas assessments of mathematical reasoning are the foci on the Process variable. Figure 7 shows a pictorial representation of the relationship between the Process and Knowledge Representation variables.

Figure 7.
Relationship Between the Process and Knowledge Representation Variables



It is believed that the Pilot Items do more to capture students' thinking and reasoning processes than traditional assessment tools. With the Pilot Items, reasoning is accounted for by deconstructing problems into component parts, rather than focusing solely on the correctness of the final answer. In doing so, students are provided with opportunities and credited for understanding aspects of problems, even if the traditional representation of knowledge is incorrect. For example, one participant was unable to respond correctly to any of the selected items from the WIAT-II in which students are awarded points only when the correct response is produced. However, this same participant was able to demonstrate some understanding when tested with the

Pilot Items. Indeed, the nature of the scoring criteria for the pilot items was such that for most questions, students had the opportunity to receive partial credit for items, unlike the WIAT-II test questions. This example demonstrates the grey area of learning. Although the participant had not mastered the ability to recall addition, subtraction, multiplication, and division facts; knowledge existed below the surface and was only uncovered when a different type of question was asked. As such, the Pilot Items merit a rating of Exceeds Expectation on the Process factor (measures and awards points for correct reasoning).

The third factor, Knowledge Representation, provides an opportunity for the Pilot Items to contribute to the existing gap in both research and practice. Recall that on this variable, the WIAT-II mathematics subtests received a rating of Needs Improvement, given that the test questions provided limited or no opportunity for students to represent their knowledge in non-traditional formats. In many ways, the very nature of the Pilot Items represents a somewhat non-traditional approach to measuring mathematical knowledge. Traditionally, students are asked to solve a computation problem or word problem and the correct answer is taken to indicate mastery, which presumably would include understanding. However, as described above, with regard to the pilot items, test questions are deconstructed such that the items no longer measures one's ability to carry out a mathematical procedure, but rather one's mathematical thinking and understanding. To do so, different approaches, such as drawing a picture or creating a word problem from a given algorithm, are included as acceptable forms by which students may convey their understanding of a given problem. If the content of the test were expanded beyond simply a pilot study, this would be an important variable to build upon. Due to the availability for students to demonstrate their knowledge beyond traditional approaches, a rating of Exceeds Expectation is awarded to the Knowledge Representation variable (provides many opportunities for students to represent their knowledge in non-traditional formats).

The fourth factor, Time Constraints, did not differ among any of the tests reviewed for the purposes of this study. Similar to the Math Reasoning subtest of the WIAT-II, the Pilot Items were presented in a manner in which students were provided with as much time as they desired to pursue any given problem. However, if it was clear that a student was uncomfortable with a problem (e.g. broken eye contact, fidgeting), then the examiner proceeded to the next problem while providing some encouragement, such as saying “That one was tricky. This next one may be easier”. As such, students were provided with ample time, yet were not pressured to solve problems beyond their skill levels. As such, the Pilot Items receive a rating of Exceeds Expectation on the Time Constraints variable (students are provided with ample time to complete test items).

Finally, the Implications for Learners factor will be considered with regard to the Pilot Items. Similar to the Content variable, this factor was negatively impacted due to the small sample size of a pilot study being one in which initial ideas are tested for significance prior to undertaking a study of more consequence. Ideally, common error types would be developed, linked with suggestions for remediation, and tested for effectiveness during assessment. Due to the small number of Pilot Items, it is difficult to get a sense of a student’s pattern of strengths and weaknesses. As such, the Pilot Items obtain a rating of Needs Improvement on the Implications for Learners variable (test results do not link easily to interventions and strategies are not tested for effectiveness during intervention).

Overall, the Pilot Items were shown to contribute to the field of assessment of mathematics achievement, albeit imperfectly. Indeed, the areas in which the Pilot Items were poorly ranked were those in which the existing assessment measure received greater ratings and conversely, the areas in which the Pilot Items were highly ranked were those in which the existing measure lacked. Both measures received an excellent rating on the Time Constraints variable. A summary of the results for the fourth research question appears in Table 11 and a summary of ratings for both item sets appears in Table 12.

Table 11.
Summary of Results for Research Question 4

Factor	**Outcome**
Content	Needs Improvement
Process	Exceeds Expectation
Knowledge Representation	Exceeds Expectation
Time Constraints	Exceeds Expectation
Implications for Learners	Needs Improvement

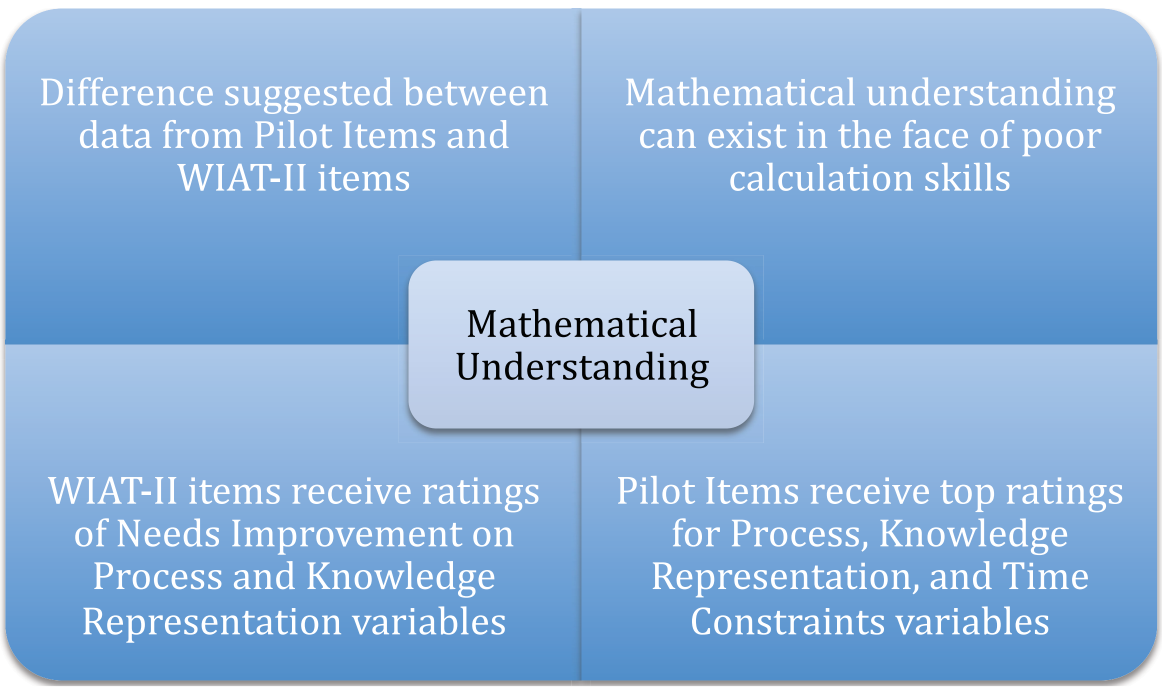
Table 12.
Summary of Ratings for WIAT-II items and Pilot Items

Factor	Numerical Operations Rating	Math Reasoning Rating	Pilot Items Rating
Content	Meets Expectation	Meets Expectation	Needs Improvement
Process	Needs Improvement	Needs Improvement	Exceeds Expectation
Knowledge Representation	Needs Improvement	Needs Improvement	Exceeds Expectation
Time Constraints	Exceeds Expectation	Exceeds Expectation	Exceeds Expectation
Implications for Learners	Meets Expectation	Meets Expectation	Needs Improvement

Conclusion

The results of the present study impart many interesting findings. Unfortunately, the first two research questions, which sought to shed light on the relationship between mathematical thinking and understanding and computation skills, proved to bring forward statistically insignificant results. The small number of participants who took part in the study undoubtedly influenced this result. However, qualitative analysis suggested that there is indeed a difference between the types of information collected by the assessment tools. Also, qualitative analysis of the second research question suggested that mathematical understanding can exist in the face of poor calculation skills. Future research will need to confirm these findings. The findings of the third and fourth research question are also of a qualitative nature. By comparing the item sets to the Evaluation Rubric, it was discovered that there are areas of strength and weakness in each item set. While the WIAT-II items were superior in content and in linking assessment findings to intervention, the Pilot items were more adept at assessing students' processes and in providing greater opportunities for knowledge representation. One item set makes up for what the other lacks and as such, it would seem that an assessment tool that combined the most favourable aspects of each item set would be exemplar. The results of these four research questions suggest ideas for future research and add to the existing knowledge base in the field of assessment of mathematics achievement. Figure 8 presents a summary of the research findings with the central theme of mathematical understanding at the heart.

Figure 8.
Summary of Research Findings



Chapter 6: Conclusion

Recapitulation of the Context

Mathematics often gets a bad reputation; it frequently becomes less about understanding and more about selecting the appropriate procedure or recalling a particular fact (Hiebert, 1984). One of the reasons for this is that teachers and researchers alike have tended to focus on the mechanical aspects of mathematics and have failed to provide a real-world context for students to make sense of mathematics. Despite an increased awareness of the need to incorporate mathematical understanding into curriculum, mathematics has continued to be seen by many as limited to numbers, algorithms, and the mechanical application of procedures that relates little to everyday life (Carpenter & Leher, 1999). Teachers may feel unprepared to teach mathematics, even at an early grade level (Chapin & O'Connor, 2007). As such, students may pick up on teacher feelings of anxiety and in turn develop their own feelings of anxiety toward the subject. Also, when teachers teach mathematics as they may have been taught (i.e. without truly understanding the subject), they may use confusing language that makes the task of making meaning out of mathematical procedures all the more difficult (Carpenter & Leher, 1999). This reality is truly a shame, given that mathematics is a fascinating subject that has the power to aid students not only in their future mathematical pursuits, but also in other areas.

Within the domain of psychoeducational assessment of mathematics difficulties and disabilities, much of the research has focused on the progression of counting strategies to inform researchers of the nature of mathematics difficulties. This focus is too narrow given the broad scope of mathematics. Traditionally, assessment of learning difficulties or disabilities has required the use of standardized aptitude and achievement measures. There has been some movement away from this discrepancy model to include other frameworks, such as response to intervention, curriculum based assessment, and dynamic assessment. However, many districts continue to use the discrepancy model as the primary source of data to inform stakeholders of the nature and severity of a student's learning problems.

Within today's reality of scarce resources and diverse special needs, it is paramount that researchers discover new ways of assessing students' knowledge. Widely used achievement measures assessing mathematical knowledge have been questioned within the context of the present study, due to their over-reliance on calculation problems to enlighten educational professionals of students' areas of strength and need (Parmar et al., 1996). These frequently used assessment tools have also been criticized for their lack of ability in aiding understanding of students' problems, as well as their deficiencies in linking assessment results to meaningful and individualized interventions (Parmar et al., 1996). The aim of the present study was to assess critically the mathematics subtests of a commonly used assessment tool (WIAT-II) and to develop and test seven alternative test items that aim to measure mathematical understanding.

It is hoped that this research project will be of interest to educational professionals as it calls to attention several important issues in the domain of psychoeducational assessment. The information gathered from assessment tools are used to make decisions that are important in the lives of children and their families, including educational programming, classroom supports and accommodations, and determination of learning disability status. As such, continuous efforts must be made to ensure that assessment tools are of a consistently high quality. Currently used assessment tools of mathematics achievement may be flawed and the present research project seeks to address some of these blemishes. Past research has suggested that studies in the area of MD assessment have typically focused on calculation and counting skills (Fletcher et al., 2002; Parmar et al., 1996). Additionally, there has been a focus on the development of calculation skills within the elementary school classroom (Romberg & Kaput, 1999). The results of this study suggest that measuring calculation skills disproportionately to the other branches of mathematics may not provide a good estimate of a student's mathematical knowledge, skill, or ability. The awareness that important decisions are tied to the results of assessment tools combined with awareness of the flaws inherent in some popular assessment batteries better arms educational professionals to meet the needs of the students with whom they work.

Contributions to the Field

The present research project contributes to the existing literature base in a number of valuable ways. First, a new measure was developed to assess mathematical thinking and understanding. This tool was created based on the work of several prominent researchers within the field, including the work of Ruch and colleagues (1925), Nancy Jordan (2006), David Geary (1999, 2001, 2005), Lambdin and Walcott (2007), Jitendra and DiPipi (2003), Carpenter and Leher (1999), and the National Council of Teachers of Mathematics (1995, 2000, 2006). Since there was not an existing measure of mathematical thinking and understanding by which to compare it to, the validity of the new measure was established through critique and revision according to the expertise of two specialists within the field. Although statistical analysis was unavailable as a result of the small number of participants in the present study, future research may lend further support to the pattern of mathematical understanding existing in the face of poor calculation skills, which were observed qualitatively in the present study. Second, a contribution is made to the field in the form of an Evaluation Rubric, developed based on the work of the NCTM (1995, 2006). This tool may be used in future studies by which other assessment tools may be analyzed and refined. Last, this research project contributes to the existing knowledge base in terms of the systematic analysis, which was conducted on a widely used academic achievement measure in the area of mathematics (WIAT-II). The mathematics subtests of the WIAT-II were subjected to an evaluation based on adherence to the assessment standards of the NCTM (1995) and aspects of the NCTM's Large Scale Assessment Tool (2006). From this assessment, knowledge is gained in terms of areas of strength and those in need of improvement. It is hoped that popular assessment tools such as the WIAT-II will be refined further, especially in the areas identified in the Evaluation Rubric as in need of improvement, specifically Process and Knowledge Representation. It is hoped that professionals within the field of school psychology and related disciplines will make use of the contributions put forth by this project.

Links to Variables of Importance

Points must be made in relation to the results of the present assessment in light of the variables introduced earlier in the study that are known to impact the learning, application, and assessment of mathematical knowledge. Most pertinently, sections to be revisited include a reanalysis of the concept of *mathematics misunderstood* in the context of the two item sets tested in this study, the impact that anxiety, gender, culture, stereotypes, and language may have had on the results, and finally, an exploration of the place where the Pilot Items fit in with the assessment frameworks previously discussed. Suggestions for future research will be discussed throughout, according to each sub-topic. It is my hope that by discussing the results of the present study in the context of elements presented in the literature review that the reader will be presented with a coherent picture of the manner in which the results tie in with the many factors that are involved in mathematics important to both students and professionals.

The WIAT-II items represent the concept of *mathematics misunderstood* in the context of the present study. The items are presented in a traditional format with little room for creativity in either question or answer. Responses are scored only for the correctness of their final solution and do little to highlight the examiner's awareness of the student's thinking process. Advantages of the WIAT-II items include a reasonably broad range of content assessed and the fact that students are provided with sufficient time by which to complete the problems. Results link reasonably well to interventions, although more could be done to make the process more straightforward. Certainly, more could be expected from such a prominent assessment tool.

Conversely, the Pilot Items were developed to offer an alternative to traditional tests of mathematical ability. In the areas where the WIAT-II items are weakest, the Pilot Items are strongest; specifically, in the areas of providing opportunities for students to demonstrate their knowledge in non-traditional formats, and in facilitating the examiner's awareness of the students' mathematical thought processes. In this manner, the Pilot Items do a better job of accessing students' authentic mathematical knowledge and test for understanding, rather than simply testing calculation proficiency. Although in their present form, the Pilot Items are not yet comprehensive enough to constitute a full mathematics achievement battery, the foundation has been laid for future study to develop these items into an assessment tool of greater significance.

The influence of gender, culture, and stereotypes appeared to have little bearing on the results of the study. No observations were made that would cause the researcher to believe that either of the participants' performance suffered or was bolstered as a result of gender, culture, or stereotypes. However, these types of influences can, in some cases, have an effect without being overt (Inzlicht & Ben-Zeev, 2000). Perhaps future studies might include a self-report measure on the personal meaning of gender, culture, and stereotypes for each participant in order to gauge better the potential effects on performance and a larger sample.

The influence of anxiety seemed to be more pronounced than that of gender, culture, and stereotypes, although did not appear to have severely detrimental effects on performance. The researcher was able to calm some fears by building rapport before, during, and after the assessment. Participants were praised for their hard work and the researcher maintained an open and friendly body language. Additionally, the parent of Participant B accompanied him during the assessment, which likely put him at ease. In Participant A's case, testing was done in her home, which would also suggest that she might be more relaxed than in another environment. However, it was observed that both participants appeared more confident in solving most of the Pilot Items, compared to the WIAT-II items. Perhaps this self-assured behaviour occurred as a result of the novelty of the Pilot Items. In future studies, it would be interesting to administer a measure of state anxiety following each item set in order to determine any effects that anxiety might have on the results of the testing. Also, it would be interesting to ask participants how they felt at the completion of each subtest.

The Pilot Items were developed with special attention to the language used during assessment. Care was taken to ensure that, to the greatest possible extent, the items were measures of mathematics knowledge, understanding, and skill, rather than of language. Certainly, the word problems may have caused more of an issue due to the added language processing demands in addition to the mathematical requirements. As previously reported, students with LD tend to have greater difficulty solving word problems in general (Jitendra & DiPipi, 2003). Unsurprisingly, then, both participants performed better on the items that called for less of a language demand. Jitendra and DiPipi (2003) identified three main areas that students with LD tend to struggle with when solving word problems: problem representation, identifying relevant information, and identifying the necessary operation(s). As such, the Pilot Items present problems to students that test each of these skills individually, rather than requiring coordination of all skills in a single question. As expected, participants performed better on the Pilot Items than on the WIAT-II items; however, it would be interesting to see if this result would be repeated in a study with a greater number of participants.

Due to the academic diversity present in every classroom (Fuchs & Fuchs, 2005), we cannot expect children’s skills to progress at precisely the same time. Rather, we can expect and prepare for students to be constantly at different levels. Teachers may be better able to support their students' mathematical development by teaching component skills required for solving word problems and teaching students to integrate these skills based on the students' knowledge and performance level, rather than requiring all concepts to be mastered at the completion of any given chapter in the curriculum outcomes. The use of an assessment tool containing items such as those found in the Pilot Items would assist teachers in supporting students' development in stages by identifying skills already possessed and subsequently systematically introducing elements still requiring mastery. It is here that future research would need to build on the development of such intervention plans based on particular stages of knowledge.

Upon expansion and further studies of reliability and validity, the Pilot Items would be suitable for use within the Discrepancy model, RTI, CBA, or Dynamic Assessment frameworks. Much work would be required in the areas of broadening the content of the test and forming a standardization sample by which comparisons to measures of IQ could be made for those working from the Discrepancy model. However, it is expected that given the harsh criticism the Discrepancy model has received in recent years, it will likely become less of a standard practice in years to come. As such, it is suggested that future efforts to expand the Pilot Items into a usable mathematical achievement assessment battery would focus on the development of a broader range of content, a set of error types that could be easily linked to intervention strategies, and a standardized procedure by which to test the effectiveness of interventions during assessment.

Implications for Practice

One theme presented in this paper has been the critique of mathematics assessments that measure calculation skills disproportionately to other branches of mathematics. The outcome of doing so can be thought of as having two potentially negative consequences, a sort of *dual dilemma*. The dual dilemma refers to the two negative outcomes that can result in testing students' mathematical prowess with a test that has an excessive amount of calculation questions. In the first scenario, the student is highly proficient in performing computations but has little or no understanding of the underlying concepts. Within the context of the present study, there was not occasion to work with such a student. However, it is suspected that larger studies in the future might encounter this phenomenon. The second scenario arises in the case where the student has little skill in doing calculations, which masks conceptual knowledge and understanding. This happening occurred in both participants tested in the present study. While both students demonstrated relatively poor calculation skills, a greater wealth of mathematical understanding was uncovered when participants were tested with the Pilot Items. Although the Pilot Items were originally intended to replace traditional mathematics test items, given that each item set appears to measure a different construct, it is now believed that the best course of action would be to integrate items measuring mathematical understanding with existing measures in order to create the most comprehensive assessment of students' mathematical knowledge.

As previously stated, it is difficult if not impossible to generalize based on the results of only two participants. However, given that both participants demonstrated mathematical knowledge that was not revealed by the WIAT-II items, a cautionary note may be made. The implication is to be aware that students who score poorly may possess some mathematical knowledge that is not captured using this assessment tool. Carefulness is necessary in making decisions that are important in the lives of students. The results of the present study suggest that the WIAT-II mathematics subtest results should be interpreted cautiously. The stakes are high, and thus, professionals are urged to take care when using the WIAT-II alone to measure mathematics achievement. Additionally, the results of testing should be corroborated with parent and teacher reports, classroom samples, and additional testing if available. These additional measures will ensure that an informed decision is made. In the future, it is hoped that research will shed more light on the ability of popular assessment tools to measure students' mathematical thinking and understanding, as distinct from their calculation skills.

Limitations

The present study was limited by a few factors. First, the small sample size prevented any meaningful statistical analyses from being computed. Second, only one rater evaluated students' responses to the Pilot Items and the WIAT-II items. Also, only one rater analyzed the Pilot Items and WIAT-II items according to the Evaluation Rubric. Future research would need to have at least two raters in order to establish inter-rater reliability. Last, the orders of the subtests were given in exactly the same sequence. It is suggested that future studies administer the subtests in different succession so as to control for order effects.

Areas for Future Research

It would be of interest to gain further clarity to the first two research questions via a study with a greater number of participants. A larger sample size would provide more information regarding the difference between the data collected between the two assessment tools and clarify the nature of the relationship between calculation skills and mathematical thinking and understanding. Also, it would be interesting to refine the Pilot Items such that a lesser language demand is required of students and the items are a more pure measure of mathematical thinking and understanding. A system for linking assessment results to evidence-based interventions would also be useful outcome from future endeavours.

Conclusion

It is hoped that this research project will provide meaningful insights to educational professionals and encourage further study in how best to assess students' mathematical knowledge. Although only preliminary conclusions may be drawn, the present study demonstrates that mathematical understanding can exist despite under-developed calculation skills. This finding is a cue to parents, teachers, school psychologists, and others that children may possess valuable mathematical understanding even when they are struggling with computation. This result lends support to the idea that calculation skills are not a good overall measure of mathematical knowledge. Also, this research is a signal to school psychologists and others who use standardized assessment tools to measure mathematics achievement to be cautious of tests that are strongly reliant on computation problems to provide an overall picture of mathematical knowledge, skill, and understanding. The Evaluation

Rubric developed for the purposes of this study can assist in evaluations of other assessment tools not reviewed in this project. Future researchers are urged to build on the existing Pilot Items to create a wider range of content and to link this assessment tool with evidence-based interventions. It is hoped that the reader has gained a greater appreciation for the many variables that influence learning and performance in mathematics and a deeper understanding of assessment of mathematics, beyond computation.

References

Ackerman, P.T., & Dykman, R.A. (1995). Reading-disabled students with and without comorbid arithmetic disability. *Developmental Neuropsychology*, 11, 351-371.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11, 181-185.

Ashcraft, M.H., & Kirk, E.P. (2001). The Relationships Among Working Memory, Math Anxiety, and Performance. *Journal of Experimental Psychology*, 130 (2), 224-237.

Ashcraft, M.H., Krause, J.A., & Hopko, D.R. (2007). Is Math Anxiety a Mathematical Learning Disability? In D.B. Berch, & M.M. Mazzocco (Eds)., *Why is Math So Hard for Some Children?* (pp. 329-348). Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Barwell, R. (2005). Language in the Mathematics Classroom. *Language & Education: An International Journal*, 19(2), 97-102.

Ben-Zeev, T., Duncan, S., & Forbes, C. (2005). Stereotypes and Math Performance. In J. Campbell (Ed), *Handbook of Mathematical Cognition* (pp. 235-249). New York, NY: Psychology Press.

Berch, D.B. (2005). Making Sense of Number Sense: Implications for Children with Mathematical Disabilities. *Journal of Learning Disabilities*, 38, (4), 333-339.

Berch, D.B., & Mazzocco, M.M. (2007). *Why Is Math So Hard for Some Children?* Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Bonnot, V., &Croizet, J.C. (2007). Stereotype Internalization, Math Perceptions, and Occupational Choices of Women with Counter-Stereotypical University Majors. *Swiss Journal of Psychology,* 66 (3), 169-178.

Borasi, R. (1994). Capitalizing on Errors as “Springboards for Inquiry”: A Teaching Experiment. *Journal for Research in Mathematics Education*, 25 (2), 166-208.

Bryant, B., & Rivera, D. (1997). Educational assessment of mathematics skills and abilities. *Journal of Learning Disabilities*, 30(1), 57-68.

Boulet, G. (2007). How Does Language Impact the Learning of Mathematics? Let Me Count the Ways. *Journal of Teaching and Learning, 5* (1), 1-12.

Brady, P., & Bowd, A. (2005). Mathematics anxiety, prior experience, and confidence to teach mathematics among pre-service education students. *Teachers and teaching: theory and practice*, 11 (1), 37-46.

Brown, R., & Josephs, R. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, 76(2), 246-257.

Burns, M. K., & Senesac, B. J. (2005). Comparison of dual discrepancy criteria for diagnosis of unresponsiveness to intervention. *Journal of School Psychology*, 43, 393–406.

Butterworth, B. (1999). A Head for Figures. *Science*, 284 (5416), 928.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language*, 16, (1), 37-55.

Carpenter, T.P., & Leher, R. (1999). Teaching and Learning with Understanding. In E. Fennema, & T.A. Romberg (Eds.)., *Mathematics Classrooms that Promote Understanding* (pp. 19-32). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Case, R., Griffin, S., & Kelly, W. (1999). Socioeconomic gradients in mathematical ability and their responsiveness to intervention during early childhood. *Developmental health and the wealth of nations: Social, biological, and educational dynamics* (pp. 125-149). New York, NY, US: Guilford Press.

Chafouleas, S.M., Riley-Tillman, T.C., & Eckert, T.L. (2004). A Comparison of School Psychologists' Acceptability, Training, and Use of Norm-Referenced, Curriculum-Based, and Brief Experimental Analysis Methods to Assess Reading. *School Psychology Review*, 32, (2), 272-281.

Chapin, S.H., & O'Connor, C. (2007). Academically Productive Talk: Supporting Students' Learning in Mathematics. In W.G. Martin, M.E. Strutchens, & P. C. Elliott (Eds.).*The Learning of Mathematics: Sixty-ninth Yearbook,* (pp. 113-128). Reston, VA: The National Council of Teachers of Mathematics.

Chiappe, P. (2005). How Reading Research Can Inform Mathematics Difficulties: The Search for the Core Deficit. *Journal of Learning Disabilities*, 38 (4), 313-317.

Clements, D.H., & Sarama, J. (2008). Focal Points – Pre-K to Kindergarten. *Teaching Children Mathematics*, 14 (6), 361-365.

Cramer, K., & Wyberg, T. (2007). When Getting the Right Answer Is Not Always Enough: Connecting How Students Order Fractions and Estimate Sums and Differences. In W.G. Martin, M.E. Strutchens, & P. C. Elliott (Eds.).*The Learning of Mathematics: Sixty-ninth Yearbook,* (pp. 205-220). Reston, VA: The National Council of Teachers of Mathematics.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Harcourt Brace Jovanovich College Publishers.

Dawson, P., & Guare, R. (2004). *Executive Skills in Children and Adolescents: A Practical Guide to Assessment and Intervention*. New York: The Guilford Press.

Deno, S.L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education*, 37, (3), 184-192.

Devlin, K. (2000). *The Math Gene*. Great Britain: Basic Books.

Donovan, M.S., & Cross, C.T. (2002). *Minority students in special and gifted education.* Washington, DC: National Academy Press, http://www.nap.edu/catalog/ 10128.html

Eysenck, M., & Calvo, M. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, 6(6), 409-434.

Fennema, E. (1989). The study of affect and mathematics: A proposed generic model for research. In D.B. McLeod & V.M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 205-219). New York: Springer-Verlag.

Fennema, E., Franke, M., Carpenter, T.P., & Carey, D. (1993). Using Children's Mathematical Knowledge in Instruction. *American Educational Research Journal*, 30 (3), 555-583.

Fletcher, J.M., Coulter, W.A., Reschly, D.J., Vaughn, S. (2004). Alternative Approaches to the Definition and Identification of Learning Disabilities: Some Questions and Answers. *Annals of Dyslexia*, 54 (2), 304-331.

Fletcher, J.M., Foorman, B.R., Boudousquie, A., Barnes, M.A., Schatschneider, C., & Francis, D.J. (2002). Assessment of reading and learning disabilities: A research-based intervention-oriented approach. *Journal of School Psychology*, 49, (1), 27-63.

Fletcher, J.M., Lyon, G. R., Fuchs, L.S., & Barnes, M.A. (2007). Mathematics Disabilities. In *Learning Disabilities From Identification to Intervention* (pp. 208-235). New York: The Guilford Press.

Freeman, L., & Miler, A. (2001). Norm-referenced, Criterion-referenced, and Dynamic Assessment: what exactly is the point? *Educational Psychology in Practice*, 17 (1), 3-16.

Fuchs, D., & Fuchs, L.S. (2005). Peer-Assisted Learning Strategies: Promoting word recognition, fluency, and comprehension in young children. *Journal of Special Education,* 39, 34-44.

Fullilove, R.E., Treisman, P.U. (1990). Mathematics Achievement among African American Undergraduates at the University of California, Berkeley: An Evaluation of the Mathematics Workshop Program. *Journal of Negro Education*, 59, (3), 463-478.

Geary, D.C. (2004). Mathematics and Learning Disabilities. *Journal of Learning Disabilities*, 37, (1), 4-15.

Geary, D.C. (2005). Role of Cognitive Theory in the Study of Learning Disability in Mathematics. *Journal of Learning Disabilities*, 38 (4), 305-307.

Geary, D., & Brown, S. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology*, 27(3), 398-406.

Geary, D.C., & Hoard, M.K. (2001). Numerical and arithmetical deficits in learning-disabled children: Relation to dyscalculia and dyslexia. *Aphasiology*, 15 (7), 635-647.

Geary, D.C., Hoard, M.K., & Hamson, C.O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, 74, 213-239.

Gelman, R. & Butterworth, B. (2005). Number and language: how are they related? *TRENDS in Cognitive Sciences*, 19, (1), 6-10.

Gersten, R., Jordan, N., & Flojo, J. (2005). Early Identification and Intervention for Students with Mathematics Difficulties. *Journal of Learning Disabilities*, 38 (4), 293-304.

Hale, J.B., Flanagan, D.P., & Naglieri, J.A. (2008). Alternative Research-Based Methods for IDEA 2004 Identification of Children With Specific Learning Disabilities. *Communiqué*, 36, (8), 1, 14-17.

Hammil, D. D., & Bryant, B.R. (1991). Standardized assessment and academic intervention. In H.L. Swanson (Ed.), *Handbook on the assessment of learning disabilities: Theory, research, and practice* (pp. 373-406). Austin, TX: PRO-ED.

Hiebert, J. (1984). Children's Mathematics Learning: The Struggle to Link Form and Understanding. *The Elementary School Journal*, 84 (5), 497-513.

Hiebert, J., & Carpenter, T.P. (1992). Learning and Teaching with Understanding. In D.A. Grouws (1), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-97). New York: MacMillan Publishing Company.

Hintze, J.M., Christ, T.J., & Methe, S.A. (2006). Curriculum-Based Assessment. *Psychology in the Schools*, 43 (1), 45-56.

Hosp, J. L., & Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children*, 70, 185–199.

Hyde, J., Fennema, E., Ryan, M., & Frost, L. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, 14(3), 299-324.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11(5), 365-371.

Jitendra, A., & DiPipi, C. M. (2003). *An Exploratory Study of Schema-Based Word-Problem-Solving Instruction for Middle School Students with Learning Disabilities: An Emphasis on Conceptual and Procedural Understanding.* Accessed Feb 21 2008 http://www.ldonline.org/article/5678.

Jordan, N., Kaplan, D., Oláh, L., & Locuniak, M. (2006). Number Sense Growth in Kindergarten: A Longitudinal Investigation of Children at Risk for Mathematics Difficulties. *Child Development*, 77(1), 153-175.

Jussim, L., & Eccles, J.S. (1992). Teacher Expectations II: Construction and Reflection of Student Achievement. *Journal of Personality and Social Psychology*, 63 (6), 947-961.

Kaput, J.J. (1988). Linking representations in the symbol systems of algebra. In S. Wagner & C. Kieran (Eds.). *Research issues in the learning and teaching of algebra*, Vol. 4, (pp.167-194). Reston, VA: National Council of Teachers of Mathematics.

Keller, J. (2007). Stereotype Threat in Classroom Settings: The Interactive Effect of Domain Identification, Task Difficulty, and Stereotype Threat on Female Students' Maths Performance. *British Journal of Educational Psychology*, 77, 323-338.

Kersaint, G. (2007). The Learning Environment: Its Influence on What Is Learned. In W.G. Martin, M.E. Strutchens, & P. C. Elliott (Eds.). *The Learning of Mathematics: Sixty-ninth Yearbook,* (pp. 3-26). Reston, VA: The National Council of Teachers of Mathematics.

Kilpatrick, J., Swafford, J., Findell, B., & National Academy of Sciences – National Research Council. (2001). Adding It Up: Helping Children Learn Mathematics. (ERIC Document Reproduction Service No. ED468038) Retrieved August 21, 2007, from ERIC database.

Kozey, M., & Siegel, L.S. (2008). Definitions of Learning Disabilities in Canadian Provinces and Territories. *Canadian Psychology*, 49, (2), 162-171.

Kulak, A. G. (1993). Parallels between Math and Reading Disability: Common Issues and Approaches. *Journal of Learning Disabilities*, 26, (10), 66-73.

Lambdin, D.V., & Walcott, C. (2007). Changes Through the Years: Connections between Psychological Learning Theories and the School Mathematics Curriculum. In W.G. Martin, M.E. Strutchens, & P. C. Elliott (Eds.). *The Learning of Mathematics: Sixty-ninth Yearbook,* (pp. 3-26). Reston, VA: The National Council of Teachers of Mathematics.

Learning Disabilities Association of Nova Scotia (n.d.). *General Information.* Accessed August 25 2007. http://ldans.nsnet.org/information.html

Learning Disabilities Roundtable. (2002). *Specific learning disabilities: Finding common ground.* Washington DC: U.S. Department of Education, Office of Special Education Programs, Office of Innovation and Development.

Linn, R.L. (2001). A Century of Standardized Testing: Controversies and Pendulum Swings. *Educational Assessment, 7* (1), 29-38.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Schulte, A., & Olson, R. (2001). Rethinking learning disabilities. In C. E. Finn, Jr., R. A. J. Rotherham, & C. R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Lyon, G. R., Fletcher, J. M., & Barnes, M. C. (2002). Learning disabilities. In E. Mash & R. Barkley (Eds.), *Child psychopathology* (vol. II), (pp. 520-586). New York: Guildford.

McCloskey, M. (2007). Quantitative Literacy and Developmental Dyscalculia. In D.B. Berch, & M.M. Mazzocco (Eds)., *Why is Math So Hard for Some Children?* (pp. 415-429). Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

McKean, E. (Ed.). (2005). The New Oxford American Dictionary (2nd ed.). Oxford University Press.

Meacham, J., McClellan, M. Pearse, T., & Greene, R. (2003). Student Diversity in Classes and Educational Outcomes: Student Perspectives. *College Student Journal*, 37 (4), 627-642.

Moss, J., & Case, R. (1999). Developing Children's Understanding of the Rational Numbers: A New Model and an Experimental Curriculum. *Journal for Research in Mathematics Education*, 30 (2), 122-147.

National Association of School Psychologists. (2003). *NASP recommendations: LD eligibility and identification for IDEA reauthorization*. Bethesda, MD: Author.

National Council of Teachers of Mathematics. (1995). *Assessment Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2000). *Principals and Standards for School Mathematics*, Reston, VA: Author.

National Council of Teachers of Mathematics (2006). *Large Scale Assessment Tool*. Reston, VA: Author.

Nova Scotia Ministry of Education. (1996). *Special Education Policy Manual*. Halifax, Nova Scotia: Author.

Parmar, R.S., Frazita, R., & Cawley, J.F. (1996). Mathematics assessment for students with mild disabilities: An exploration of content validity. *Learning Disability Quarterly*, 19 (2), 127-136.

Pedrotty, D. (1997). Mathematics Education and Students with Learning Disabilities: Introduction to the Special Series. *Journal of Learning Disabilities*, 30 (1), 2-19.

Pedrotty Bryant, D. (2005a). *Math Disability in Children: An Overview*. Accessed Dec 3 2007 http://www.schwablearning.org/articles.aspx?r=1001.

Pedrotty Bryant, D. (2005b). *Working With Your Child's Teacher to Identify and Address Math Disabilities* Accessed Dec 3 2007 http://www.schwablearning.org/articles.aspx?r=1055.

Pedrotty Bryant, D., & Pedrotty B. R. (2008). Introduction to the Special Series: Mathematics and Learning Disabilities. *Learning Disability Quarterly*, 31 (1), 3-8.

Pica, P., Lemer, C., Izard, V., Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306, 499-503.

President's Commission on Excellence in Special Education. (2002). A new era: Revitalizing special education for children and their families, http:// www.ed.gov/inits/commissionsboards.

Proctor, B. E., Floyd, R. G., Shaver, R. B. (2005). Cattell-Horn-Carroll Broad Cognitive Ability Profiles of Low Math Achievers. *Psychology in the Schools*, 42, (1), 1-12.

Rathus, S.A. (1988). *Understanding Child Development.* New York: Holt, Rinehart, and Winston, Inc.

Reisman, F. K., & Hutchinson, T. (1985). *Sequential Assessment of Mathematics Inventory*. Columbus, OH: Charles E. Merrill.

Rivera, D. (1997). Mathematics education and students with learning disabilities: Introduction to the special series. *Journal of Learning Disabilities*, 30(1), 2.

Romberg, T. A., & Kaput, J. (1999) Mathematics Worth Teaching, Mathematics Worth Understanding. In E. Fennema, & T.A. Romberg (1), *Mathematics Classrooms that Promote Understanding* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Royer, J. M., & Walles, R. (2007). Influences of Gender, Ethnicity, and Motivation on Mathematical Performance. In D. B. Berch & M.M. Mazzocco (Eds.), *Why is Math So Hard for Some Children?* Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Ruch, G. M., Knight, F. B., Greene, H. A., & Studebaker, J. W. (1925). *Compass diagnostic tests in arithmetic.* Chicago: Scott, Foresman.

Salvia, J., & Yesseldyke, J.E. (1995). *Assessment* (6th ed.) Boston: Houghton Mifflin.

Sanez, T.I., & Huer, M.B. (2003). Testing Strategies Involving Least Biased Language Assessment of Bilingual Children. *Communication Disorders Quarterly*, 24, (4), 184-193.

Sattler, J.M., & Hoge, R.D. (2006). *Assessment of Children: Behavioural, Social, and Clinical Foundations* (5th ed.) San Diego: Jerome M. Sattler, Publisher, Inc.

Savage, R.S., Frederickson, N., Goodwin, R., Patni, U., Smith, N., & Tuersley, L. (2005). Relationships Among Rapid Digit Naming, Phonological Processing, Motor Automaticity, and Speech Perception in Poor, Average, and Good Readers and Spellers. *Journal of Learning Disabilities*, 38, (1), 12-28.

Shalev, R.S., Manor, O., & Gross-Tsur, V. (2005). Developmental Dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology*, 47, (2), 121-125.

Shapiro, E.S., Angello, L.M., & Eckert, T.L. (2004). Has Curriculum-Based Assessment Become a Staple of School Psychology Practice? An Update and Extension of Knowledge, Use, and Attitudes From 1990 to 2000. *School Psychology Review*, 33, (2), 249-257.

Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, 25, 618-629.

Small, M.S. (1990). Do You Speak Math? *The Arithmetic Teacher*, 37, (5), 26-29.

Smith, R. A., & Davis, S. F. (2003). *The Psychologist as Detective: An Introduction to Conducting Research in Psychology, Third Edition*. Upper Saddle River: NJ. Pearson, Prentice Hall.

Sroufe, L.A. (1997). Psychopathology as an outcome of development. *Development and Psychopathology*, 9, 251-268.

Stanovich, K.E. (2005). The future of a mistake: Will discrepancy measurement continue to make the learning disabilities field a pseudoscience? *Learning Disabilities Quarterly, 28* (2), 103-106.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43,* 93-126.

Stearns, P. N. (1998). Why Study History? *American Historical Association*. Retrieved April 10 2008. http://www.historians.org/pubs/free/WhyStudyHistory.htm

Steele, C.M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist*, 52 (6), 613-629.

Steele, C. (1998). Stereotyping and its threat are real. *American Psychologist*, 53(6), 680-681.

Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34. pp. 379-440), San Diego. CA: Academic Press.

Stuebing, K., Fletcher, J.M., LeDoux, J.M., Lyon, G.R., Shaywitz, S.E., &Shaywitz, B.A. (2002). Validity of IQ-discrepancy classification of reading difficulties: A meta-analysis. *American Education Research Journal*, 39, 469-518.

Taylor, R.L., Tindal, G., Fuchs, L., & Bryant, B.R. (1993). Assessment in the nineties: A possible glance into the future. *Diagnostique*, 18 (2), 113-122.

Tobias, S. (1994). *Overcoming Math Anxiety*. New York: W.W. Norton & Company.

Torgesen, J.K. (2001). Intensive Remedial Instruction for Children with Severe Reading Disabilities: Intermediate and Long-Term Outcomes from Two Instructional Approaches. *Journal of Learning Disabilities*, 34, (1), 33-58.

Torgesen, J.K. (2002). The Prevention of Reading Difficulties. *Journal of School Psychology*, 40, (1), 7-26.

Trujillo, K.M., & Hadfield, O.D. (1999). Tracing the Roots of Mathematics Anxiety Through In-Depth Interviews with Pre-Service Elementary Teachers. *College Student Journal*, 33 (2), 219-232.

Vaidya, S.R. (2004). Understanding Dyscalculia for Teaching. *Education*, 124 (4), 717-720.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-Achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, 33, 223-238.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children – Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2005). *Wechsler Individual Achievement Test – Second Edition*. San Antonio, TX: The Psychological Corporation.

Widmer, C., & Chavez, A. (1982). Math Anxiety and Elementary School Teachers. *Education*, 102 (3), 272-276.

Witzel, B.S., Ferguson, C.J., & Brown, D.S. (2007). *Developing Early Number Sense for Students with Disabilities*. Accessed Feb 22 2008 http://www.ldonline.org/article/14618.

Woodcock, R., & Johnson, M. (1990). *Woodcock-Johnson psycho-educational battery: Test of achievement, Third Edition.* Chicago: Riverside.

Wynn, K. (1992). Addition and Subtraction by Human Infants. *Nature*, 358, 749-750.

Appendix A: Pilot Items

Pilot Items

(Based on Ruch et al., 1925)

I.

```
     47
   + 64
```

1.1 *Describe in your own words what this problem is asking.*

Scoring Criteria:
2 pts

- It's asking you to combine the two numbers to find the total
- You find the sum of the two numbers
- You take 47 and add it to 64 and see what you get

1pt

- Describes procedure (e.g. first you add 7 + 4….)
- You add 47 and 64

0 pts

- Describes an operation other than addition

1.2 Which one is a good estimate to the problem we just talked about? (Show problem again)

a) 1011
b) 110
c) 556

Scoring Criteria:
2 pts:
110

0 pts:
1011 or 556
(Do not give partial credit)

II.

```
    110
   - 17
```

2.1 Which operation should you use to solve this problem?
a) Addition
b) Subtraction
c) Multiplication

d) Division

2.2 Tell in your own words what _________ (insert stated operation) means.

Scoring Criteria:
2 pts

- General: Provides an acceptable response describing subtraction
- You subtract 17 from 110 and see what's left over
- You take 17 away from 110 and find the difference
- Subtract the bottom number from the top number

1 pt

- General: Provides an accurate yet incomplete description of subtraction
- You take the smaller number away from the bigger number

0 pts

- Incorrectly describes subtraction
- Take 110 away from 17 and find the difference
- Describes one of the other 3 operations (addition, multiplication, or division)

2.3 Make up your own word problem using the numbers above (i.e. 110-17).

Scoring Criteria:
2 pts

- Correctly represents both quantity and operation

1pts

- Correctly represents quantity OR operation

0 pts

- Does not correctly represent either number or operation

III.
Last year, Mrs. Craig's class had 26 students. The year before that, there were 27 students. This year, there are 22 students. This year, there are an equal number of boys and girls in the class. Over the last three years, what is the average number of students in Mrs. Craig's class?

3.1 What important pieces of information do you need to know in order to solve this problem?

Scoring Criteria:
2 pts

- Names all 3 important pieces of information (number of students in class each year for the past 3 years)

- Does not name unessential information (equal number of boys and girls in the class this year)

1pt

- Names 1 or 2 pieces of important information
- Names 1 or fewer pieces of unimportant information

0 pts

- Does not name any important piece of information
- Names unimportant information

IV.
Samantha has 104 Christmas cards she wants to send. The last day to drop the cards at the post office in order for them to be delivered on time is December 22nd. Pretend that today is December 15th. How many Christmas cards must Samantha prepare each day in order to have them ready to mail on the 22nd?

4.1 Express this problem using symbols OR Draw a picture to demonstrate your knowledge of this problem.

Scoring Criteria:
2 pts

- $104 \div 8 = \_\_$
- Draws a picture (e.g. of a calendar) and correctly indicates how many cards Samantha has to complete each day in order to meet the mailing deadline

1 pt

- Draws a picture (e.g. of a calendar) and indicates that Samantha has to complete between 10 and 15 cards each day in order to meet the mailing deadline
- Algorithm is partially correct (either numbers or operation)

0 pts

- Algorithm is incorrect (both numbers and operation)
- Picture does not make sense given the information provided in the problem

---

Appendix B: Evaluation Rubric

Evaluation Rubric

(Based on NCTM's Large Scale Assessment Tool, 2006)

	Exceeds Expectation	Meets Expectation	Needs Improvement
Content	Measures a broad range of mathematical processes and concepts	Measures basic mathematical processes and concepts	Measures a limited range of mathematical processes and concepts
Process	Measures and awards points for correct reasoning	Informal or qualitative assessment of reasoning	Awards points only for correctness of final answer
Knowledge Representation	Provides many opportunities for students to represent their knowledge in non-traditional formats (e.g. pictures)	Provides some opportunities for students to represent their knowledge in non-traditional formats	Provides limited or no opportunity for students to represent their knowledge in non-traditional formats
Time Constraints	Students are provided with ample time to complete test items	Some time restraints are imposed	All tests are timed
Implications for Learners	Potential strategies for remediation are tested for effectiveness during assessment	Performance on general strands suggests areas for remediation but strategies are not tested for effectiveness during assessment	Test results do not link easily to interventions and strategies are not tested for effectiveness during assessment

Appendix C: Consent Forms

Free and Informed Administrative Consent Form

Mathematics Beyond Computation: An Examination of Assessment Tools and Exploration of an Alternative Item Set
Researcher: Sarah Angelopoulos

To Whom It May Concern:

Please allow me to introduce myself. My name is Sarah Angelopoulos, and I am a graduate student in the Department of Education at Mount Saint Vincent University. As part of my Master of Arts in School Psychology program, I am conducting research under the supervision of Dr. Geneviève Boulet. My research is being conducted in the area of assessment of mathematics learning disabilities.

Given the current emphasis on mathematical understanding in both the mathematics research field and in the classroom, it is important that psycho-educational assessment tools give students the opportunity to demonstrate their mathematical skills and knowledge through a variety of formats. Tests should measure students' mathematical thinking and understanding as well as their skill in carrying out mathematical procedures. My goal is to test the feasibility of a brief item set that focuses on capturing students' mathematical thinking and understanding beyond that accounted for by calculation proficiency in comparison to traditional assessment tools. The results of the present research may assist test developers in revising psycho-educational assessment batteries to better reflect today's emphasis on mathematical understanding and the different ways in which this knowledge may be represented.

I am inviting you to participate in my study. I plan to assess 5 students who have been diagnosed with Learning Disabilities using a brief item set developed to assess mathematical thinking and understanding. The target age range includes students from grades 4, 5, and 6.

The item set consists of a total of 12 test questions and will take between 15 and 30 minutes to complete. Questions include items such as asking students to represent their understanding of a given problem by drawing a picture, selecting the appropriate operation from a list of response choices, choosing the best estimate of a given problem (e.g. 30 + 42), as well as more traditional questions such as solving a standard subtraction problem (e.g. 45-12). The students' responses to the questions will help to determine whether or not these types of questions provide information about a student's mathematical thinking and understanding beyond that captured by the ability to compute basic calculation problems. If desired, parents are welcome to accompany their child during testing for observational/supervisory purposes. Parents who wish to accompany their child must agree not to engage in any verbal discourse with the student during testing or otherwise prompt the child as such behaviours may invalidate the results of the study.

Additionally, the researcher will ask parents to provide demographic information about their child, including the child's ethnicity, age, and any diagnosed medical or mental health conditions. Past research has suggested that Japanese and Chinese children outperform their European and North American counterparts on tests of mathematical skills. One reason suggested for this phenomenon is that the counting rules of the Chinese and Japanese languages appears to facilitate the learning of place value, counting, and computation more so than the English, French, or Spanish languages. As such, it will be of interest to collect data regarding the ethnic background of each of the participants in the study. In terms of age, this information is of interest in that older students may perform better than younger students, perhaps due to the greater number of years of instruction and experience with the subject matter. Although the test items are designed to be accessible to students ranging in grade from 4 to 6, age is an important variable that may influence the results of the study and therefore is in need of consideration. Information on diagnosed medical and/or mental health conditions allows the researcher to consider the implications that diagnoses may have on the results of the study.

There is a possibility that particularly math anxious students will experience feelings of anxiety (e.g. worrying) upon learning that the proposed research project is related to mathematics. In the event that a child was math anxious, the researcher will assure all students that participation is completely voluntary and that refusal to participate will not harm the student in any way. It is hoped that such reassurance will be sufficient to ensure that no participant in the present study suffers from excessive anxiety.

All information will be kept strictly confidential, and participants will not be identifiable to the general public in the final thesis. Once the initial data collection is complete, raw data will be coded such that no identifying information (e.g. name, school) will be present on the data forms. Data will be stored in a locked filing cabinet in a locked office on Mount Saint Vincent University property for five years. Following, data will be destroyed. Upon receiving parental consent, students will also be described the nature of the research project, be provided an opportunity to ask questions, and will also decide whether or not to participate in the study. Participation will be voluntary throughout the process, and students or their parents/legal guardians will be free to withdraw from the study at any time without penalty. Students will be assured that declining to participate in the study will not harm their grades in any way. Upon completion of the study, an executive summary of the research will be prepared and distributed to all participants.

In order to protect the privacy of participants, the researcher requests that participating schools agree to do the following: select students who meet the study's target sample, mail letters of consent to the parents of the identified students, schedule times/rooms for the researcher to meet with participants, and mail the executive summaries to participants.

It is my hope that this research will help educational professionals to understand more fully the scope of students' mathematical understanding.

If you have any questions, please feel free to contact me directly at [redacted] or [redacted]. My supervisor, Dr. Geneviève Boulet, may also be contacted at 902-457-6305, or genevieve.boulet@msvu.ca. This research activity has met the ethical standards of the University Research Ethics Board at Mount Saint Vincent University. If you have any questions or concerns about this study and wish to speak with someone who is not directly involved with this study, you may contact the University Research Ethics Board, by phone at 902-457-6350 or by e-mail at research@msvu.ca.

Sincerely,

Sarah Angelopoulos, BA
Master's Student, Department of Graduate Education
Mount Saint Vincent University

I, ____, Executive Director of LDANS, give my permission for Mrs. Sarah Angelopoulos to administer a set of 12 test items assessing mathematical skill and understanding to students at this school for whom parental consent and individual student assent has been given. It is understood that this research is being conducted by Mrs. Angelopoulos as part of the requirements for her degree of Master of Arts in School Psychology, under the supervision of Dr. Geneviève Boulet.

I understand that all information shared between participants and Mrs. Angelopoulos will be strictly confidential, that each student's participation is completely voluntary, and that he/she may withdraw participation at any time during the study without penalty.

By signing this consent form, you are indicating that you have read the above information, fully understand the above information, and agree to participate in this study.

Executive Director's signature **Date**

Researcher’s signature **Date**

One signed copy to be kept by the researcher, one signed copy to the participant.

Free and Informed Parental/Legal Guardian Consent Form

Mathematics Beyond Computation: An Examination of Assessment Tools and Exploration of an Alternative Item Set
Researcher: Sarah Angelopoulos

Dear Parent(s)/Guardian(s),

Please allow me to introduce myself. My name is Sarah Angelopoulos, and I am a graduate student in the Department of Education at Mount Saint Vincent University. As part of my Master of Arts in School Psychology, I am conducting research under the supervision of Dr. Geneviève Boulet. My research is being conducted in the area of assessment of mathematics learning disabilities.

Given the current emphasis on mathematical understanding in both the mathematics research field and in the classroom, it is important that psychoeducational assessment tools give students the opportunity to demonstrate their mathematical skills and knowledge through a variety of formats. Tests should measure students' mathematical thinking and understanding as well as their skill in carrying out mathematical procedures. My goal is to test the feasibility of a brief item set that focuses on capturing students' mathematical thinking and understanding beyond that accounted for by calculation proficiency in comparison to traditional assessment tools. The results of the present research may assist test developers in revising psychoeducational assessment batteries to better reflect today's emphasis on mathematical understanding and the different ways in which this knowledge may be represented.

I am inviting your child to participate in my study. I plan to assess 5 students who have been diagnosed with Learning Disabilities using a brief item set developed to assess mathematical thinking and understanding. The target age range includes students from grades 4, 5, and 6. Although some students dislike mathematics or exhibit worrying or nervous behaviours regarding mathematics, it is helpful for researchers to have the opportunity to work with students who have such feelings in an effort to effectively address this reality for many students. Therefore, parents of "math-anxious" students are encouraged to consider allowing their child to participate in the study. However, there is a possibility that particularly math anxious students will experience feelings of nervousness (e.g. worrying) upon learning that the research project is related to mathematics. In order to alleviate some of these fears, the researcher will assure all students that participation is completely voluntary and that refusal to participate will not harm the student in any way. It is hoped that such reassurance will be sufficient to ensure that no participant in the present study suffers from excessive anxiety.

The item set consists of a total of 12 test questions and will take between 15 and 30 minutes to complete. Questions include items such as asking students to represent their understanding of a given problem by drawing a picture, selecting the appropriate operation from a list of response choices, choosing the best estimate of a given problem (e.g. 30 + 42), as well as more traditional questions such as solving a standard subtraction problem (e.g. 45-12). The students' responses to the questions will help to determine whether or not these types of questions provide information about a student's mathematical thinking and understanding beyond that captured by the ability to compute basic calculation problems. If desired, parents are welcome to accompany their child during testing for observational/supervisory purposes. Parents who wish to accompany their child must agree not to engage in any verbal discourse with the student during testing or otherwise prompt the child as such behaviours may invalidate the results of the study.

Additionally, the researcher will ask parents to provide demographic information about their child, including the child's ethnicity, age, and any diagnosed medical or mental health conditions. Past research has suggested that Japanese and Chinese children outperform their European and North American counterparts on tests of mathematical skills. One reason suggested for this phenomenon is that the counting rules of the Chinese and Japanese languages appears to facilitate the learning of place value, counting, and computation more so than the English, French, or Spanish languages. As such, it will be of interest to collect data regarding the ethnic background of each of the participants in the study. In terms of age, this information is of interest in that older students may perform better than younger students, perhaps due to the greater number of years of instruction and experience with the subject matter. Although the test items are designed to be accessible to students ranging in grade from 4 to 6, age is an important variable that may influence the results of the study and therefore is in need of consideration. Information on diagnosed medical and/or mental health conditions allows the researcher to consider the implications that diagnoses may have on the results of the study.

All information will be strictly confidential, and participants will not be identifiable to the general public in the final thesis. Once the initial data collection is complete, raw data will be coded such that no identifying information (e.g. name, school) will be present on the data forms. Data will be stored in a locked filing cabinet in a locked office on Mount Saint Vincent University property for five years. Following, data will be destroyed. Upon receiving parental consent, students will also be described the nature of the research project, be provided an opportunity to ask questions, and will decide whether or not to participate in the study. Participation will be voluntary throughout the process, and students will be free to withdraw from the study at any time without penalty. Students will be assured that declining to participate will not harm their grades in any way. Upon completion of the study, an executive summary of the research will be prepared and distributed to all participants.

It is suggested that parents/legal guardians who elect to grant permission for their child to participate in the research project discuss their decision to do so with the child. This conversation would be particularly important for children who might have worries or anxieties related to mathematics. As part of the consent process, students will also give their consent to participate or decline to participate once the researcher has presented the purpose of the project to the student.

It is my hope that this research will help educational professionals to understand more fully the scope of students' mathematical understanding.

If you have any questions, please feel free to contact me directly at [redacted], or [redacted]. My supervisor, Dr. Geneviève Boulet, may also be contacted at 902-457-6305, or genevieve.boulet@msvu.ca. This research activity has met the ethical standards of the University Research Ethics Board at Mount Saint Vincent University. If you have any questions or concerns about this study and wish to speak with someone who is not directly involved with this study, you may contact the University Research Ethics Board, by phone at 902-457-6350 or by e-mail at research@msvu.ca.

Sincerely,

Sarah Angelopoulos, BA
Master's Student, Department of Graduate Education
Mount Saint Vincent University

I, ____, Parent/Legal Guardian of ____, give my permission for Mrs. Sarah Angelopoulos to administer a set of 12 test questions assessing mathematical skill and understanding to my child.

I understand that this research is being conducted by Mrs. Angelopoulos for the purposes of her research under the supervision of Dr. Geneviève Boulet, that all information will be kept confidential, and that no identifying information will be revealed in the final project.

In addition to my consent, I understand that my child will be given the opportunity to choose to participate or decline to participate in this study and that my child may withdraw from this study at any time without penalty.

I have read the above information, have been given the opportunity to ask questions, and am satisfied that I have sufficient information to give consent for my child to participate in this study.

By signing this consent form, you are indicating that you have read the above information, fully understand the above information, and agree to participate in this study.

____________________________ _________________
Parent/Legal Guardian of above stated child Date

___________________________ _________________
Researcher's signature Date

One signed copy to be kept by the researcher, one signed copy to the participant.

Free and Informed Participant Assent Form

Mathematics Beyond Computation: An Examination of Assessment Tools and Exploration of an Alternative Item Set
Researcher: Sarah Angelopoulos

I am a graduate student in the Faculty of Education at Mount Saint Vincent University. As part of my Master of Arts in School Psychology degree, I am conducting research under the supervision of Dr. Geneviève Boulet. I am inviting you to participate in my study looking at mathematical skill and understanding. The purpose of the study is to compare a short set of questions that measure mathematical understanding to another short set of more traditional math questions (e.g. 42 + 57). By comparing students' answers to these two sets of questions, I'm hoping to find more information about the relationship between understanding math and being able to do basic math problems (i.e. 42 + 57).

This study involves a set of 12 questions designed to measure students' mathematical thinking and understanding. Questions include items such as asking students to show their understanding of a problem by drawing a picture, choosing the best operation from a list of response choices, and choosing the best estimate of a given problem (e.g. 30 + 42), as well as more traditional questions such as solving a standard subtraction problem (e.g. 45-12). It will take approximately 15-30 minutes to complete the questions. The results of the research project will help researchers to improve tests that measure students' mathematical knowledge, skill, and understanding.

Participating in this research project will help researchers to understand better students' mathematical understanding and the types of questions that may best allow students to show what they know about math. Some students, who don't like math, may feel worried or nervous about participating in a research project that is related to doing math. However, your participation is completely voluntary and deciding not to participate will not harm your grades in any way. If you choose to participate in this study, you may change your mind and decide that you don't want to participate anymore at any time without penalty.

Although the researcher will know whom the students are who participated in the study, participants will not be identified in the final research project. Once the students complete the questions, I will make a photocopy and put a code on the form so that there's no information that could identify the student (e.g. name, school). The original copy will be stored in a locked filing cabinet in a locked office at Mount Saint Vincent University for five years and then will be destroyed. Once the research project is complete, I will make a summary of the research, which will be mailed to your house, and your parent or guardian will explain it to you. I will be available at that time to answer any questions that you may have.

If you have any questions, please feel free to contact me directly at [redacted], or [redacted]. My supervisor, Dr. Geneviève Boulet, may also be contacted at 902-457-6305, or genevieve.boulet@msvu.ca. This research activity has met the ethical standards of the University Research Ethics Board at Mount Saint Vincent University. If you have any questions or concerns about this study and wish to speak with someone who is not directly involved with this study, you may contact the University Research Ethics Board, by phone at 902-457-6350 or by e-mail at research@msvu.ca.

Sincerely,

Sarah Angelopoulos, BA
Master's Student, Department of Graduate Education
Mount Saint Vincent University

I, ____, student, am volunteering to participate in the research being conducted by Sarah Angelopoulos and understand that this research is being conducted as part of the requirements for her degree of Master of Arts in School Psychology.

The purpose of this research has been explained to me by Mrs. Angelopoulos, and I understand that I may withdraw from this study at any time without penalty.

I understand that the assessments being conducted by Mrs. Angelopoulos for the purposes of her research are confidential and that my identity will not be revealed. I have read this permission form and/or it has been read to me, and I have been given the opportunity to ask questions before signing.

By signing this consent form, you are indicating that you fully understand the above information and agree to participate in this study.

_____________________ _____________

Participant's signature **Date**

_____________________ _____________

Researcher's signature **Date**

One signed copy to be kept by the researcher, one signed copy to the participant's parent(s)/guardian(s).