A CASE STUDY OF ACCOUNTABILITY FOR SPECIAL EDUCATION SERVICE DELIVERY: A MIXED MODEL ANALYSIS

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Education

In the Department of Educational Psychology and Special Education

University of Saskatchewan

Saskatoon, SK

By

AUSTIN DEGENHARDT

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate Degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

The Head of the Department of Educational Psychology and Special Education
University of Saskatchewan
Saskatoon SK S7N 0X1

ABSTRACT

There were five main purposes for the current thesis: (1) to address the need for more quantitative studies to evaluate student academic success within the inclusive classroom setting; (2) to apply a recently released program assessment rubric for special education services to determine the level of special education service delivery in the specified location; (3) to evaluate the reliability of the results of the rubric mandated by the Saskatchewan Ministry of Education (2008); (4) to compare the results of the standardized student achievement tests with the rubric results in relation to program effectiveness; and (5) to investigate potential confounding factors related to the current study design. The goal of this thesis was to provide information to the Living Sky School Division and to the Saskatchewan Ministry of Education on the implementation and success of the inclusion model in a rural Saskatchewan setting. In addition, results were intended to provide information on assessment instruments employed in the measurement of program effectiveness.

The analysis was conducted as a mixed-methods case study that included two parts: (1) the first assessment indicated that students with learning difficulties scored significantly higher on standardized academic achievement measures while in an inclusive setting as opposed to scores while in a pullout setting; and (2) the second assessment determined that special education service delivery was *emerging/developing* to *evident*. The correlation coefficient of rubric results was calculated at $\alpha = .69$. A variety of general measurement issues, including small sample size and use of historical data, in relation to the current study design, were discussed.

ACKNOWLEDGEMENTS

I express my sincere appreciation to the following people, in no particular order, for their support, guidance, and inspiration: Dr. Ivan Kelly, Dr. Brian Noonan, Dr. Laurie Hellsten, Dr. Laureen McIntyre, Living Sky School Division central office administrators, division administrators, and teaching staff, Hollis Lai, Shelanne Hepp, and external examiner, Dr. Jay Wilson.

I also wish to express my appreciation to the members of the Educational Leave Committee in my school division for affording me this opportunity. I also thank my family, notably my wife Gail, for her support and encouragement.

DEDICATION

Dedicated to the countless students, parents, teachers and administrators who choose to persevere daily to encourage one another to overcome life's obstacles and improve our society through education.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

1.1 Introduction

The inclusion model for education, in North America, has evolved out of legislation recognizing the rights of disabled people. In the United States, Public Law 94-142, the Education for All Handicapped Children Act, was passed in 1975 (Woloschuk, 2003). The Act signaled change in the treatment of students with disabilities through the provision of free and appropriate education services. The services included special education program provision designed to meet the unique educational needs of students with disabilities (Seymour and Seymour, 1979). In Canada, in 1982, the Charter of Rights and Freedoms entrenched equality rights for people with disabilities to receive equal benefits through any programming aimed at improving their condition. (Woloschuk, 2003). In 1983, with the passing of Bill C-141, the Canadian Human Rights Act enshrined the rights of citizens with handicaps by legislating equal opportunity without hindrance of discrimination (Rolland de Denus, 1995). The Charter of Rights and newly enshrined human rights forced education systems to provide effective, non-discriminatory, and equitable programming for students with special needs.

The *Report on the Committee on Integration of Students with Special Needs in the Classroom* (Hopson et al., 1999) explained the current legal responsibilities of school divisions in educating students with special needs:

> The 1997 Supreme Court of Canada's decision in *Eaton V. Brant County Board of Education* held that the *Canadian Charter of Rights and Freedoms* does not give rise to the legal presumption of a right to be integrated into a regular classroom. The Court determined that children are not burdened or disadvantaged by such placement decisions when: (1) the best placement of the child is considered; (2) the child's best interests and

special needs are taken into account; (3) an ongoing assessment of the child's best interest is provided so that changes to the child's needs may be reflected in the placement; and, (4) the decision is made from a subjective, child-centered perspective, one that attempts to make equality meaningful from the child's point of view, rather than from the point of view of the adults in the child's life. (p. 8)

The Supreme Court of Canada ruling in *Eaton V. Brant County Board of Education* suggested that the implementation of inclusive philosophy was a decision to be made by each school division, provided that the decision was child-centered. Various delivery structures for special education services were possible and were implemented.

In the late 1990's, the inclusion movement was officially recommended and endorsed in the majority of Canada primarily through two sources (Saskatchewan Ministry of Education, 2000). The first source was a study by Bunch, Lupart, and Brown (1997), which found that educators had concerns about the inclusion movement in relation to increased teacher workload and preparation. However, the study also revealed that there was general agreement among educators that inclusion was beneficial for both regular and students who were included. The second source endorsing inclusion was a document entitled *In Unison* (Ministers, 1998). *In Unison* was an agreement and commitment from Canadian Ministers of social services from all regions to implement programs and initiatives that promoted equitable treatment and inclusion of persons with disabilities.

The Final Report of the Saskatchewan Special Education Review Committee (Saskatchewan Ministry of Education, 2000) included a major focus on two areas: (1) the widespread implementation of inclusion; and (2) accountability within the inclusive special education service delivery model. The report acknowledged the use of standardized testing as a

measure of accountability but also encouraged less reliance on standardized tests and the development and use of other measures for student progress. However, no suggestions for alternate forms of assessment were offered in the report.

A more recent government document entitled *Assessment for Learning Program: Supporting Data-Guided Decision-Making To Improve Student Learning* (Saskatchewan Ministry of Education, 2007) indicated that assessment, including the use of diverse assessment instruments is necessary to "provide educators with the information they need to inform planning, instructional practice and continuous program improvement" (p. 2). Sackney (2008) indicated that accountability is an integral part of systemic change. Sackney (2008) agreed with Hopkins (2001) who stated, "unless school improvement strategies impact directly on learning and achievement then we are surely wasting our time" (p. 8).

The Living Sky School Division, the setting of the current study, has endorsed inclusive education in its policies. The philosophy of inclusionary practice is supported, as indicated in the guiding principles of the school division, which include: care, integrity, trust, honesty, mutual respect, courage, commitment, inclusion, and innovation (Living Sky School Division, 2006). The principal of inclusionary practice is consistent with the goals of the Saskatchewan Ministry of Education, as indicated in the mandated rubric (Saskatchewan Ministry of Education, 2008).

In addition to the guiding principles, the school division mission statement indicates that the foundational beliefs include: (1) Student learning is priority number one; (2) Students learn and staff work best in caring, respectful environments; (3) Relevant, responsive, results oriented curriculum; (4) That collaborative, authentic partnerships build strength; (5) Our organization is accountable to students, parents, and community; 6) In prevention and early intervention as most effective practice; 7) Our organization strengthens through shared leadership (Living Sky School

Division, 2006). As is evident in the list, the principle of accountability is a foundational belief of the educational process within the school division. The principle of accountability is also consistent with the goals of the Saskatchewan Ministry of Education, as is mandated in the rubric for special education service delivery (Saskatchewan Ministry of Education, 2008).

The relative benefits and drawbacks of full-scale inclusion have been studied quite extensively in the United States. As Manset and Semmel (1997) indicated, a wide variety of approaches have been attempted in establishing an inclusive environment in schools. Widespread inclusion has been occurring in the United States since the late 1970's due to the passing of Federal legislation protecting the rights of persons who are disabled (Manset & Semmel, 1997). An area of concern for students with learning difficulties was their academic achievement in a regular classroom setting as results related to academic achievement were inconclusive (Manset & Semmel, 1997).

Seethaler and Fuchs (2005) reviewed five major special education journals over a five-year period from 1999 to 2004, which resulted in the identification of 806 articles. Of identified articles, only 5.46% tested either a reading or math intervention. Only 4.22% of the articles involved testing with random assignment. Results of Seethaler et al's (2005) review indicated very little quantitative investigation of educational practices in the area of inclusion.

Begeny and Martin (2007) indicated that inclusion has been widely used in Italy since the 1970's. However, the authors also stated that, despite the relatively long period of practice, there has been a lack of quantitative studies regarding the success of inclusion in Italy. Further, they stated that numerous Italian studies using surveys provided favourable results. The relative few quantitative studies did not support the favourable survey endorsements. In discussion, Begeny and Martin (2007) encouraged empirical study of academic and social outcomes of inclusion.

The *Report on the Committee on Integration of Students with Special Needs in the Classroom, January 1999* (Hopson et al., 1999) indicated that since the 1990's in Saskatchewan, there has been an attempt to integrate children at all levels of learning into the regular classroom. However, there has been no systematic quantitative investigation of academic achievement to follow up the implementation of inclusion philosophy.

The Final Report of the Saskatchewan Special Education Review Committee (Saskatchewan Ministry of Education, 2000) called for the implementation of inclusive education and for accountability measures to monitor the success of implementation. However, it has not been until the recent release of the document, *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice* (Saskatchewan Ministry of Education, 2008), that there has been a common assessment instrument for measuring the effectiveness of special education service delivery in the province of Saskatchewan.

An extensive literature review carried out by Lindsay (2007) in order to assess the degree of quantitative assessment of inclusion success. Lindsay (2007) concluded that the inclusion philosophy is based upon concern for the rights of children. Further, positive benefits of inclusion are not well established given the lack of comparison studies providing quantitative evidence of academic success. Lindsay (2007) encouraged more quantitative assessment of academic achievement for students with learning difficulties that had experienced non-inclusive and subsequent inclusive educational settings. Though Lindsay (2007) indicated a lack of quantitative assessment of academic achievement for students with learning difficulties, other jurisdictions and researchers were encouraging the use of multiple methods of program

assessment (Alberta Learning, 2004; Gall, Gall, & Borg, 2007; Oliver-Hoyo & Allen, 2006; Saskatchewan Ministry of Education, 2000; and Stake, 2004).

The document titled *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice* (Saskatchewan Ministry of Education, 2008), provided a rubric for assessing special education program delivery. The rubric contained descriptions of various levels of attainment of criteria for meeting requirements of an effective inclusion model in special education. The rubric provided an accountability measure, advocating more than the standardized testing of students, for the delivery of special education services in the province of Saskatchewan. The implementation of the rubric was consistent with recommendations made by the Saskatchewan Special Education Review Committee (2000) and with Sackney (2008) in his call for the integration of accountability measures into systemic reform.

1.2 Statement of Purpose

There were five main purposes for the current thesis. The first was to address the need, as indicated by Manset and Semmel (1997), Seethaler and Fuchs (2005), Lindsay (2007), and Begeny and Martin (2007), for more quantitative studies to evaluate student academic success within the inclusive classroom setting. The current thesis was, in particular, a response to a lack of studies comparing achievement of students that had been exposed to both inclusive and noninclusive settings (Lindsay, 2007). It was evident in the reviewed articles that the proportion of quantitative studies of academic achievement in special education students remains very low (Lindsay, 2007). The second purpose was to apply a recently released program assessment rubric for special education services to determine the level of special education service delivery in the specified location. This was established in response to the call for alternate program

assessment measures in special education settings (Saskatchewan Ministry of Education, 2000). The third purpose was to evaluate the reliability of the results of the rubric mandated by the Saskatchewan Ministry of Education (2008). This was established with the intent of providing a statistical description of the rubric results. The fourth purpose was to compare the results of the standardized student achievement tests with the rubric results. This arose from a desire to investigate the congruency or incongruency between the results of the two types of assessment. The fifth purpose of the current thesis was to investigate potential confounding factors related to the current study design. The goal of this thesis was to provide information, to the Living Sky School Division and to the Saskatchewan Ministry of Education, on the implementation and success of the inclusion model in a rural Saskatchewan setting and to provide information on assessment instruments employed in the study.

The research questions relating to the intended purposes were:

1. Do standardized test results indicate greater achievement for a group of students with learning difficulties in an inclusive setting as compared to results for the same group in a noninclusive setting?
2. Are special education service delivery goals being achieved in the particular setting according to conditions specified on the Saskatchewan Ministry of Education rubric (2008)?
3. Is the mandated rubric a reliable measure of program effectiveness?
4. Are the results of the standardized achievement tests congruent with the rubric results?
5. What are potential confounding factors related to the current study design?

1.3 Definitions

1.3.1 Special learner

"The term *special learner* denotes students who may, indeed, have special needs historically known in the province as "designated" disabilities and now as students requiring intensive supports or with intensive needs (i.e.: Intensive Supports funding recognition). However, it also includes those students who are at risk, have mild disabilities or who have needs arising from environmental effects (i.e.: Diversity funding recognition)" (Saskatchewan Ministry of Education, 2008, p. 3).

The Government of Saskatchewan Ministry of Education website stated that funding recognition is provided to school divisions for specific students as identified through a Provincial Impact Process. It further stated that funding pertains to students who have learning needs that impact on numerous areas of performance and require intense and frequent supports as documented in their Personal Program Plan (Saskatchewan Ministry of Education, 2008).

The Saskatchewan Ministry of Education Funding Manual for 2007/2008 reiterated the funding requirements (Saskatchewan Ministry of Education, 2008). It also indicated that funding is provided to a school division based on school division statistics of need. The school division makes the decision to fund specific students or programs based upon practical necessity. The target group of students with learning difficulties in the current study did not contain students with difficulties identified as sufficient to warrant special government funding. By the previous definition, the target group of students for the current case study would be considered as students with learning difficulties by virtue of being *at risk*. It is important to note that the target group of students participated in a regular adapted curricular program. Other than some adaptations, as need arose, the program was consistent with that received by their cohort and the general school

population in each chronological grade. Specific procedures applied to the target group are presented in chapter 3. Special learners are henceforth referred to as students with learning difficulties or as students with LD.

1.3.2 At-risk

In the present study, target group students were identified as *at-risk* because their initial scores on the CTBS fell below the 30th percentile. The 30th percentile was established for identification of students who were at-risk at the time that the first standardized test results were available for the target group.

1.3.3 Inclusion

Saskatchewan Ministry of Education (2000) indicated "inclusion may be … characterized as a philosophy of education and a set of related practices that have implications for the location of a child's instruction" (p. 14). However, no further specifics were offered.

Seven principles that are expected to be included in an inclusive environment are: inclusionary practice, differentiated instruction, parental involvement, assessment, team/collaboration, fostering independence, and assistive technology (Saskatchewan Ministry of Education, 2008).

For the purposes of the current study, *inclusion* was defined as a particular philosophy of education that is supported by the fore-mentioned principles. In practical terms, it means being included in a regular classroom setting with the various supports necessary to maximize a successful outcome for each individual student, irrespective of disability.

1.3.4 Pullout

The term *pullout* was defined as an educational philosophy in which students leave the regular classroom and travel to a smaller room to receive specialized instruction (Brandts, 1999).

It is expected that special education services be delivered by a qualified special education professional. The pullout program in the current study consisted largely of service delivery by educational associates, which falls into the lowest rubric category for fostering independence (Saskatchewan Ministry of Education, 2008).

1.3.5 Qualified special education professional

A *qualified special education professional* is a professional who has: (1) a Master's Degree in Special Education from any university; or (2) a Saskatchewan Professional A Teaching Certificate along with having successfully completed a minimum of eighteen credits of specified courses in special education. This includes courses from each of the following areas: Speech and Language; Individual Assessment of Students with Exceptional Needs; Programming for Students with Exceptional Needs; and additional credits in approved special education courses (Saskatchewan Ministry of Education, 2007).

1.3.6 Norm-referenced Test

"Norm-referenced interpretation is a score interpretation based on comparison of a test-taker's performance to the performance of other people in a specified reference population" (Frisbie, 2005, p. 24).

1.3.7 Rubric

"At the most basic level, a rubric lists criteria and levels of quality" (Andrade, 2005, p. 27).

1.3.8 Responsive evaluation

The definition of *responsive evaluation* was adopted from Stake (2004) and is described as an attitude within the evaluation process. Responsive evaluation involves the collection of accounts of personal experiences and interpretations of stakeholders within a defined

environment. Stakeholders can potentially include anyone directly or indirectly affected by the program being examined. In this thesis, the stakeholders were limited to a Superintendent, two principals, two special education teachers, and a regular classroom teacher.

1.3.9 Resilience

Resilience is the ability to respond to adversity with better than anticipated outcomes (Masten, 2001).

1.4 Significance of the Study

The current study was motivated by the recommendation by the Saskatchewan Ministry of Education (2000) that: (1) the inclusion model be comprehensively implemented within Saskatchewan; (2) accountability measures rely less on results of standardized testing; and (3) that alternate accountability measures be implemented. Benchmarks for various levels of special education service delivery assessed in the current thesis were established by the Saskatchewan Ministry of Education (2008).

This study assessed the delivery of special education services within the inclusion model in a rural Saskatchewan school by applying two different accountability measures: (1) standardized test results; and (2) rubric-guided responsive evaluation. Results of this study were important for several reasons as follows: (1) they provided a response to previous researchers who have encouraged more quantitative assessment of student achievement within inclusive settings (Begeny & Martin, 2007; Lindsay, 2007; Manset & Semmel, 1997; and Seethaler & Fuchs, 2005); (2) the current study further provided the unique opportunity to compare standardized results for the same group of students that received two different treatments in a naturally occurring school setting. Such comparisons have been quite rare (Lindsay, 2007); (3) the current study provided an assessment of program effectiveness, as measured on different

dimensions using more than one method of measurement; and (4) the results of this study will potentially provide insight into program implementation status and program effectiveness as measured by student outcomes and levels of service delivery.

1.5 Chapter Organization

Chapter 2 reviews literature related to the inclusion of students with learning difficulties, assessment of program effectiveness through the measurement of academic achievement, assessment of program effectiveness through the use of rubrics in program evaluation, and related measurement issues in the current study. A description of the research methods and procedures employed are presented in Chapter 3. An analysis of the data is presented in Chapter 4. The final chapter, Chapter 5, summarizes the findings of the study, conclusions, implications for practice, and directions for further research.

CHAPTER 2

2.1 Literature Review

This literature review was focused on two issues: (1) measurement of academic achievement within the inclusion model; and (2) complementary methods of measuring accountability including the use of compliance rubrics. Supporting topics include: classical test theory and standardized testing; test score reliability; responsive evaluation; and application of accountability measures. The review of classical test theory provides theoretical background and rationale for the use of standardized test results. The presentation of responsive evaluation methods of program assessment explains the need to include both standardized test results and rubric assessments to gain a more informed interpretation of the program being assessed (Stake, 2004). A discussion of accountability measures provides the rationale for the specific procedures and measures employed.

2.1.1 Inclusionary Practice and Academic Achievement

Manset and Semmel (1997), in a review of studies investigating academic achievement in inclusive settings, indicated that students with learning disabilities that are integrated into regular classroom settings have yielded mixed academic results. Though relatively recent studies indicated negative or insignificant results for students with learning difficulties in inclusive settings that contain no extra support or accommodation for needs (Holahan and Costenbader 2000; Schulte et al. 1990; Zigmond et al., 1995), other studies indicate that academic benefits to inclusion can be realized in supportive settings (Borman, 2007; Jenkins et al., 1994; Wang & Birch, 1984; Wang, Peverly, & Rudolph, 1984; Zigmond et al., 1995). Introducing special education procedures such as individualized instruction, grouping, use of graphic organizers, and cooperative learning into regular classrooms also increased achievement in normal-functioning

students as compared to classrooms without the special procedures (Manset and Semmel, 1997). Achievement was based on "quantified data on student academic performance…reports were made on nonacademic outcomes such as retention, special education referral rate, attendance, behavior, and self-esteem" (Manset & Semmel, 1997, p. 165). "Researchers of three model programs (Jenkins et al., 1994; Schulte et al., 1990; and SFA, 1993) reported significantly larger academic gains for students with mild disabilities or low achievement than for the controls in traditional pullout programs on measures of reading or composite score of reading, writing, and math." (Manset & Semmel, 1997, p. 172).

One of the earliest successful programs was that of Wang and Birch (1984) and their Adaptive Learning Environments Model (ALEM). Wang and Birch (1984) found greater reading and math mean score gains for ALEM students with mild disabilities than for controls. The regular group also consisted of students with learning difficulties but had received traditional pullout treatment. Wang (1984) described the effects of ALEM applied in 26 inclusive classrooms. The study indicated that properly structured inclusion, including appropriate classroom supports and individualized strategies, was effective for moderately handicapped students and that segregated treatment for specific students be considered only after having attempted full inclusion. In Wang, Peverly, and Randolph's (1984) study, average gains, based on standardized test results, for ALEM students were approximately one year in both math and reading.

The presentation of Wang, Peverly, and Randolph's (1984) research brings attention to the issue of inclusive programming and to the relatively small gains, in terms of program implementation, that have been made in a quarter of a century. Specific inclusive strategy and delivery models (Jenkins et al., 1994; Wang & Birch, 1984; Zigmond et al., 1995), shown to be

effective, have existed for a reasonably long period of time. Further research of subsequent programs indicated similar positive gains for students with learning difficulties in an inclusive setting (Borman, 2007; McLeskey & Waldron, 2000; Peterson & Hittie, 2003; Rea, McLaughlin, & Walther-Thomas, 2002).

Schulte et al. (1990) found that, as groups, students with LD all made significant gains in reading, written language, and math, regardless of whether the student received consultation but no direct special services or was in a resource room setting. Consultation would have involved specific classroom instructional strategies undertaken by the regular classroom teacher after consultation with a special education professional. The subjects of the study were children with learning disabilities assigned to one of four conditions: one period of resource room instruction per day; two periods of resource room instruction per day; consultative services combined with in-class instruction; and consultative services to classroom teachers. There were no apparent differences in student achievement based on the particular learning environment. Schulte et al. (1990) found no academic advantage to being placed in a resource room setting.

Significantly greater gains in vocabulary, reading, and language for their students, in general, in experimental versus control schools were found by Jenkins et al. (1994). Some instructional strategies used by Jenkins (1994) included: teacher-led basal-related activities; partner reading; comprehension questions; story-related writing; words out loud; word meaning; story retelling; spelling; partner checking; direct instruction in reading comprehension; home reading; integrated language arts and writing; tests; and cooperative learning procedures (Jenkins et al., 1994). Students were categorized into three different groups: (1) regular education students; (2) remedial education students; and (3) special education students. Researchers used five different standardized tests to assess achievement. A pre-test/post-test

design was used, which demonstrated that regular education students significantly outscored both remedial and special education students. Remedial education students significantly outscored special education students. On post-test measures, students in the experimental school demonstrated significantly superior gains on several Metropolitan Achievement Test (MAT) scales, including reading vocabulary, total reading, and language, with a marginally significant effect on reading comprehension. The positive effects were spread across all student types including regular, remedial and special education in the experimental schools (Jenkins et al., 1994).

Zigmond et al. (1995) described three research projects designed to compare effectiveness of inclusion with special education students. The research projects were undertaken at the University of Pittsburgh, the University of Washington, and Vanderbilt University. The explicit goal for the University of Pittsburgh and University of Washington models was to eliminate pullout remedial and special education services. Consequently, both schools terminated all forms of pullout service when the project began. Zigmond et al. (1995) compared the achievement gain of each student with learning disabilities with the average gain of the relevant grade-level peer group.

In the University of Pittsburgh and University of Washington models, traditional special education services were almost entirely discontinued. Remedial and special education staff members provided support to general education teachers. In the Pennsylvania schools, the special education teachers assumed the roles of co-teacher and co-planner of the instruction to be delivered in the general education classroom. The Vanderbilt University model encouraged several special education reforms. First, special education instruction was intensified in an attempt to raise the number of students who had skills that would permit reintegration. Second,

as students were reintegrated the special educators began to serve as consultants to general educators. Special educators and general educators worked together to terminate pullout services for many children already identified as having special needs. The Vanderbilt model developed and implemented caseloads for teachers. Caseloads were adjusted to reduce the general educator's class size and to increase that of the special educator, as well as to distribute students with disabilities and other low-achieving students equitably. Results revealed differences across the three projects, with 53% (Pittsburgh), 38% (Washington), and 62% (Vanderbilt) of students with learning disabilities gaining ground on their peers. Overall, 61% of the special education students had moved up in relative standing.

Borman et al. (2007) assessed both the potential cumulative effects of the program on school-level achievement outcomes and the longitudinal outcomes of students who attended Success for All (SFA) and control schools across 3 years. The Success for All Foundation originated in 1987 at John Hopkins University and was initiated to address lower levels of achievement in schools by students in poverty and in ethnic minorities. SFA is a commercially available program that has been developed over many years. The SFA program did not employ traditional approaches such as pullouts (Success for All, 2008). All students were treated on an inclusive school-wide basis. Borman et al. (2005a) carried out analysis of the 1st-year achievement data for the kindergarten and first-grade sample. Statistically significant positive effects on the Woodcock Word Attack scale but no effects on three other reading measures were found. The second-year analyses, reported by Borman et al. (2005b), focused on the literacy outcomes for two distinct student samples. One set of analyses was for a two-year sample of students who remained enrolled at the treatment and control schools over the full two years of the study. Analyses revealed statistically significant school-level effects on three of the four

literacy outcomes measured, with effects exceeding half of a school year gain on the Word Attack outcome. Results indicated a time effect that had not been apparent in other literature. Borman et al. (2005b) indicated an effect for time of exposure. The results indicated by Borman et al. (2007) are relevant to the current thesis because the current thesis target group was educated in a noninclusive setting for two consecutive years followed by three consecutive years in an inclusive setting.

Lindsay (2007) expressed concern about the lack of quantitative studies on the effectiveness of inclusive programming. Accordingly, Lindsay directed an extensive literature review in order to assess the degree of quantitative assessment of inclusion success. The study assessed articles from eight major journals, which included *The Journal of Special Education*, *Exceptional Children*, *Learning Disabilities Research and Practice*, *Journal of Learning Disabilities*, *Remedial and Special Education*, *The British Journal of Special Education*, the *European Journal of Special Needs Education*, and the *International Journal of Inclusive Education*. The study reviewed articles from a six-year span from 2000 to 2005. One thousand three hundred and seventy three papers on the effectiveness of inclusive education were reviewed. Of the 1373 papers, only nine, about 0.5 %, compared performance results for special needs students, including academic achievement across inclusive and noninclusive settings. Only five articles, including two reviews, compared special needs students with regular students. Lindsay (2007) concluded that the inclusion philosophy is based upon concern for the rights of children. Further, positive benefits of inclusion are not well established given the lack of comparison studies providing quantitative academic evidence of success. Lindsay (2007) encouraged more quantitative assessment of academic achievement for students with learning difficulties that had experienced non-inclusive and subsequent inclusive educational settings.

2.1.2 Inclusionary Practice and Benefits beyond Academic Achievement Results

Consistent with Schulte (1990) and Zigmond (1995), Willrodt (1995) found no significant difference in passing rates on a measure in math and reading between the study groups. However, the study concluded that, since significant social and emotional gains had emerged, the decision as to whether an inclusive program is more beneficial for special needs students cannot be made based solely on expected academic improvements. Additionally, in preschool children aged three to five, Holahan and Costenbader (2000) indicated no difference for lower functioning students and found no academic advantage to being placed in a resource room setting. Students with learning difficulties in the resource room setting made significant gains. Similarly, students with learning difficulties in the inclusive setting also made significant gains. Neither setting demonstrated an academic advantage over the other. However, there were other advantages to inclusion (Blum & Libbey, 2004; Bost & Riccomini, 2006; Holahan & Costenbader, 2000; Klem & Connell, 2004;).

Holahan and Costenbader (2000) found statistically significant gains in social and emotional development, as measured on the Brigance Diagnostic Inventory of Early Development, between inclusive and self-contained classrooms. There were no statistically significant differences in the areas of self-help skills and general knowledge and comprehension. Inclusion made a positive difference for the relatively higher functioning students, those with relatively better developed social and emotional skills. There was no difference for relatively lower functioning students, those with relatively less developed social and emotional skills.

Dirling (1999) made a case for inclusion and outlined procedures that could be adopted in a regular classroom. Suggested procedures included engaging in positive interactions with students, building upon students' areas of strength, and exhibiting and modeling positive

behaviours. The establishment of Dirling's (1999) recommended classroom structure and procedure built resilience in students by encouraging them to solve problems in a positive way. Dirling's (1999) finding was based upon interpretation of research done by Benard (1991), which indicated that resilient children usually have four attributes: social competence; problem-solving skills; autonomy; and a sense of purpose and future. The identification of these four areas provided direction for later research studies.

Consistent with Dirling's (1999) findings, Klem and Connell (2004) stated, "Studies show students with caring and supportive interpersonal relationships in school report more positive academic attitudes and values, and more satisfaction with school. These students also are more engaged academically" (p. 262). Klem and Connell (2004) found that children with learning difficulties, in grades three to eight, aged seven to fifteen, who were engaged, demonstrated psychological investment, interest, and emotional involvement in their school, have more positive academic attitudes and values and were more satisfied with school. Higher levels of engagement have been linked with improved performance (Klem & Connell, 2004). Engagement was measured using a battery of 6 separate tests known collectively as the Research Assessment Package for Schools (RAPS). The RAPS contained surveys that were completed by parents, students, and teachers. The Student Performance and Commitment Index (SPCI) was used to assess academic achievement and behaviour. SPCI thresholds used to identify students as *at-risk* fell below the 35th percentile on reading for elementary students and below the 25th percentile for secondary students. By linking improved performance to school engagement, Klem and Connell (2004) suggested that increased engagement, as fostered and promoted by an inclusive setting, would lead to increased academic achievement. The individual assessment

component of the current thesis was based upon Klem and Connell's (2004) suggested link between inclusion and academic achievement.

Blum and Libbey (2004) presented a summary of a comprehensive literature review in relation to the topic of school engagement, which they called connectedness. Their summary included the following points. Student success could be improved through stronger bonds with the school. In order to feel connected, students should experience high expectations for academic success, feel supported, and feel safe. School connectedness could positively impact negative behavior. School connectedness was linked to higher achievement. School connectedness was related to a reduction in socially inappropriate or self-destructive behaviors. Connectedness could be fostered through appropriate curriculum and teaching strategies. The findings of Blum and Libbey (2004) demonstrated a wide variety of positive student outcomes including increased academic achievement but also increased achievement in non-academic areas. Therefore, measures beyond academic achievement assessment are desirable to provide a broader spectrum of program evaluation. The rubric for program evaluation that is mandated by the Saskatchewan Ministry of Education (2008) provided such an alternative method of assessment.

Subsequent support for inclusion was provided by Wiener and Tardiff (2004) in a Canadian study of children in grades 4-8. Students were divided into four distinct groups based upon the level of assessed learning difficulties and the type of treatment that the respective groups received. Children experiencing mild to moderate learning difficulties were placed into one of two situations. The first placement involved attendance in a regular classroom with a special education teacher attending the classroom for between 60 and 90 minutes per day to address specific needs. The second placement involved a traditional resource room approach

where students left the classroom to get assistance in a resource room environment for up to 90 minutes per day. In addition, children with serious learning disabilities were placed in either an inclusive classroom or in a self-contained special education classroom. Children in the inclusive classroom spent the entire day in the general classroom and received services from two teachers that were team teaching. Students in the self-contained special education classroom spent at least half of the day segregated with some integration with the greater student population. The focus of the study was to assess the social and emotional impact of inclusive treatment (Wiener & Tardiff, 2004). The authors reported more positive outcomes on social measures for an inclusive approach as opposed to segregation. The authors stated that whenever differences were apparent between groups, it was "always the children in the more inclusive settings who fared better" (Wiener & Tardiff, 2004, p. 27). In general, children in more inclusive settings were better accepted by their peers, had higher self-perceptions of mathematics competence, and experienced fewer teacher-rated problem behaviours. However, academic achievement was not measured.

Bost and Riccomini (2006) identified specific traits of students with learning disabilities as typically having trouble with national measures of academic performance, needing effective interventions, and needing intensive and systematic explicit instruction. The finding that students with learning disabilities typically have trouble with national measures of academic performance might provide an indicator for the inconsistent results reported by Manset and Semmel (1997).

There are aspects other than academic achievement, to consider when making program decisions (Fraturra & Capper, 2006). They concluded that segregated treatment of special education students is: cost ineffective; produces no more achievement success than inclusive

environments; is based on failure, marginalization and blame; and is ethically questionable. The authors cited supporting evidence (McLeskey & Waldron, 2000; Peterson & Hittie, 2003; Rea, McLaughlin, & Walther-Thomas, 2002) of increased academic achievement and more positive social outcomes for all students in an inclusive setting. Frattura and Capper (2006) argued that a segregation model has three underlying weaknesses that are an impediment, rather than an aid, to student success. First, the authors reported that students are placed into programs for assistance only after having an established record of failure across various dimensions. They iterated that the traditional treatment for underachieving students was to attempt to get the student to be flexible to available programming. They suggested that programs should be flexible and not impose upon the student to make the required adjustments. Second, removing students with difficulties from the regular classroom resulted in a diminished capacity in the opportunity for teacher development because the regular classroom teacher was deprived of related professional growth. Third, segregated programs were not typically individualized. Instead, students were required to fit the program. The weakness of that approach is that the many different specific learning styles and needs could not be addressed within a rigidly structured and prescribed program. Frattura and Capper (2006) discouraged the compartmentalization of services and encouraged a more global approach to professional preparation in terms of special education training. Even as recent as the publication of Frattura and Capper's (2006) article, the debate about the most appropriate model for implementing inclusion persisted.

The inclusion philosophy has been officially encouraged in Saskatchewan since the release of the Final Report of the Saskatchewan Special Education Review Committee: Directions for Diversity: Enhancing Supports to Children and Youth with Diverse Needs (Saskatchewan Ministry of Education, 2000). In addition to recommending the adoption of an

inclusive philosophy for providing special education services to students with learning difficulties, the Final Report also recommended accountability measures in order to provide focus and guidance for program development. Respective school divisions developed their own strategies for addressing the needs of students with LD.

2.1.3 Classical Test Theory and Standardized Testing

Van der Linden (2005) indicated that Classical Test Theory grew out of Alfred Binet's development of standardized psychological tests near the beginning of the twentieth century. Standardized testing has existed, at least, since its use in ancient China as a civil servants exam (Van der Linden, 2005). Standardized test use evolved throughout the twentieth century and became widely used in educational and psychological applications. Van der Linden (2005) postulated that part of the standardized test appeal was the relative ease of scoring and score analysis because of the objective format. The economics of continuous test development and redesign were also factors in perpetuating standardized test use. Van der Linden (2005) further asserted that most standardized tests have followed the psychometric tradition and have undergone a relatively long period of development including revision, validation, and reliability measures. In addition, Van der Linden (2005) indicated that standardization is a core notion of Classical Test Theory.

Classical Test Theory, according to DeVellis (2003), is based upon the assumptions that each score derived from the administration of a particular item contains a component of true ability and a component of error. There is also an assumption that true scores and error scores are uncorrelated. That is to say that the portion of an observed score that is representative of a subject's true ability is independent of any error also affecting the observed score (Kline, 2005). Further, the mean error score over the repeated administration of items will equal zero because,

statistically, the errors resulting in an artificially inflated score will eventually balance the errors resulting in an artificially deflated score. This assumption is important because the standard error of measurement for a test can be obtained from a single test administration to a very large group, such as CTBS (Nelson Publishing, 1997) procedures, as opposed to multiple administrations to a single subject (Kline, 2005).

DeVellis (2003) stated that the use of Classical Test Theory methods is appropriate when the assessment instrument contains items that are roughly parallel in importance in relation to the measurement of an underlying latent variable. Classical Test Theory methods have also demonstrated surprising efficacy when applied to hierarchical psychological variables (DeVellis, 2003). Standards-based evaluation instruments generally contain multiple items that are roughly parallel in importance for measuring ability in a specific domain.

Stake (2004) encouraged the use of standards-based evaluation and indicated that researchers need to be explicit about measurement criteria, standards, and other factors of evaluation. Standards-based evaluation could be particularly useful in evaluating program effectiveness. Some of the factors to be considered in standards-based evaluation are: recipient needs, program goals, evaluation criteria, and evaluation standards. Stake (2004) stated that each factor must be assessed in order to get a representative indication of program effectiveness.

Consistent with Van der Linden (2005), Stake (2004), and DeVellis (2003), the Alliance for Excellent Education (2008) stated that, despite their drawbacks, "standardized tests are currently the best objective and quantifiable measure of student learning available" (p. 3).

Townsend (1993) related that standardized testing has been a traditional measure to assess program effectiveness worldwide. Almost ten years later, Miller (2004) stated that standardized tests have been used extensively in the United States to measure program

effectiveness. Miller (2004) also indicated that, though standardized tests cannot measure all types of learning, they have been widely adopted for monitoring standards of student achievement. However, there has also been recognition that other types of assessment are necessary (Saskatchewan Ministry of Education, 2000). Zimmerman (2008) indicated that the education systems in the United States are in an era of increased testing for accountability. With recent recommendations and requirements, the Saskatchewan emphasis on accountability appears to parallel developments in the United States. The use of standardized tests for measuring program effectiveness persists (Zimmerman, 2008).

One standardized test instrument that has been used to assess student academic achievement is the CTBS (Nelson Publishing, 1997). The Alberta Initiative for School Improvement (AISI): Improving Student Learning (Alberta Learning, 2004), Provincial Report for Cycle 1 (2000-2003) indicated that the CTBS was one of several standardized tests that have been utilized in the province of Alberta for measuring program effectiveness. The CTBS (Nelson Publishing, 1997) was originally adapted from the Iowa Test of Basic Skills (ITBS) (Hoover, Dunbar, & Frisbie, 2007). The ITBS has been in use since 1935 and was adapted for use in Canada in the early 1960's (Anderson, Bernier, and Hebert, 2004). The Canadian version was modified to reflect the educational emphasis of Canadian curricula. The CTBS (Nelson Publishing, 1997) assessed general basic academic skills as opposed to specific curriculum content. The CTBS (Nelson Publishing, 1997) was used to measure student academic achievement in the current study.

2.1.4 Corroboration

In addition to standardized test results for student outcomes, The AISI Provincial Report (Alberta Learning, 2004) for Cycle 1 (2000-2003) reported the use of "Descriptions of Quality"

in 62.5% of the improvement projects, comprising the greatest proportion of effectiveness measures. Consistent with the Alberta procedures, a Saskatchewan-generated rubric has been provided in the living document titled *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice* (Saskatchewan Ministry of Education, 2008). Stake (2004) encouraged a mixed model approach to program evaluation, specifically the use of responsive evaluation procedures in the form of interviews, checklists, or surveys, as well as cross-referencing findings through traditional statistical methods. Stake (2004) also encouraged triangulation of findings, which can include cross-reference of results derived from different methods, different groups, or different levels of observation or judgment. In a literature review by Oliver-Hoyo and Allen (2006), the authors strongly encouraged the use of varied methods when investigating concepts. The authors also suggested that triangulation of results from varying methods provides a more valid inference as results converge toward the same conclusion.

2.1.5 Responsive Model

Stake (2004) encouraged the adaptation of specific accountability models to suit the specific situation. The author was reluctant to classify his approach but indicated that it has been referred to as the Countenance Model or the Responsive Model. Stake (2004) was not prescriptive in the methods used but rather in the intentions behind the measures. Responsive evaluation is concerned with assessing program quality through information gathered from stakeholders and participants. Stake (2004) also encouraged the marriage of standards-based evaluation with what might be considered as more qualitative methods.

Stake (2004) defined responsive evaluation as "a general perspective in the search for quality and the representation of quality in a program" (p. 86). Responsive evaluation is useful

for monitoring program effectiveness in its formative stages. Stake (2004) also asserted that responsive evaluation is often concerned with investigating issues.

Stake (2004) stated that "for it to be a good responsive evaluation, methods need to accommodate the here and now, serving the evaluation needs of the stakeholders at hand" (p. 94). It is uncommon for responsive evaluations to concentrate on standardized testing (Stake, 2004). The author recognized experience as an important component of judgment and that subjective perceptions have value. Stake (2004) emphasized the importance of triangulation of observations and the necessity to collect both qualitative and quantitative data. A mixed methods approach is consistent with Stake's (2004) recommendations that research findings be cross-referenced through varying forms and levels of data collection. Blending the student achievement results with systemic program evaluation results required the application of a specific evaluation model, the responsive model.

2.1.6 Measure of Reliability

It is desirable to employ reliable measurement scales in order to increase statistical power (DeVellis, 2003). DeVellis (2003) indicated that coefficient alpha is a widely used method of assessing reliability. Coefficient alpha is a modified version of the Kuder-Richardson 20 (KR20) that allows for analysis of continuous response formats. Other potential reliability measures such as the KR20 and KR21 are applicable only if the items contain dichotomous responses. Clark and Watson (1995) indicated that a minimum of 200-300 respondents is required to elicit meaningful results using factor analysis. Clark and Watson (1995) also indicated that using coefficient alpha with a scale containing more than 40 items would be inappropriate. The rubric mandated by the Saskatchewan Ministry of Education (2008) is available in one version containing 12 items. Kline (2005) warned that it is not in the researcher's best interest to modify

a predesigned test because changing the test could potentially change the test properties. In the current thesis the mandated rubric (Saskatchewan Ministry of Education, 2008) was applied in its original form.

2.2 Summary

The current literature review has focused upon the evaluation of inclusionary practices in schools. Standardized tests were a traditionally employed and relatively widespread method for gathering student academic achievement results (Van der Linden, 2005) as part of measuring program effectiveness. However, standardized test results for students with LD showed inconsistent academic achievement progress within inclusive settings (Manset & Semmel, 1997; Schulte, 1990; Willrodt, 1995; Zigmond, 1995). In response to the inconclusive results, researchers developed models focusing on specific instructional strategies and on structured programs (Borman, 2007; Jenkins et al., 1994; Wang & Birch, 1984; Wang Peverly, & Rudolph, 1984; Zigmond et al., 1995). Results were more promising for specifically structured programs but remained inconclusive. The Inclusion model endured because of the social and emotional benefits demonstrated by students with LD (Dirling, 1999; Klem & Connell, 2004). As models became refined, it became apparent that some modified classroom procedures were successful (Borman, 2007). It was also apparent that some students, regardless of environment, do not realize significant academic gains (Holahan & Costenbader, 2000).

Perhaps due to the inconsistencies shown in standardized test results for students with LD in inclusive settings, research tended not to focus on academic achievement results but more on social and emotional outcomes (Lindsay, 2007). However, Borman (2007) reported that the amount of time spent in an inclusive setting was an important factor. Borman (2007) suggested that students with LD demonstrated a significant but delayed response to the inclusive

environment.

Inclusionary practice has been recommended in Saskatchewan since the year 2000 (Saskatchewan Ministry of Education, 2008). In addition to the implementation of inclusionary practice and accountability, using a variety of measures was also a key recommendation. Historically, standardized testing has been employed in order to assess program effectiveness possibly because of its long history of use, availability, and ease of administration and analysis (Van der Linden, 2005). However, research has identified reliance on standardized measures alone as inadequate (Stake, 2004). In recent decades, other measures have been employed (Alberta Learning, 2004). Researchers and jurisdictions encouraged the use of complementary and varied methods of assessment (Alberta Learning, 2004; Gall, Gall, & Borg, 2007; Oliver-Hoyo & Allen, 2006; Saskatchewan Ministry of Education, 2000; Stake, 2004). The Responsive Model (Stake, 2004) for program assessment is conducive to the movement toward increased accountability in education service delivery because it recommends a blend of accountability measures rather than the simple administration of standardized tests. Blum and Libbey (2004) demonstrated a wide variety of positive student outcomes in an inclusive setting, including increased academic achievement and increases in non-academic areas. Wiener and Tardiff (2004) made a positive case for inclusion based on social and emotional gains but did not assess academic achievement. Bost and Riccomini (2006) identified specific traits of learning disabled students as typically having trouble with national measures of academic performance, suggesting that additional or alternative forms of assessment might be appropriate.

The Saskatchewan Ministry of Education (2008) has mandated a rubric, which does not take account of student academic achievement, for assessment of special education program delivery, offering an evaluation that is not based upon student achievement results.

From a Classical Test Theory point of view, regardless of the assessment method chosen for the task, the issue of reliability of test results is central to assessment considerations (DeVellis, 2003). DeVellis (2003) indicated that coefficient alpha is a widely used method of assessing reliability. Kline (2005) warned against modifying a predesigned test because of potentially compromised reliability.

To conclude, the literature has indicated a need for quantitative analysis of student achievement results for students with LD, particularly those that have been exposed to both non-inclusive and inclusive educational settings (Lindsay, 2007). There is relatively widespread encouragement for complementary and alternative methods of program assessment in addition to evaluation of standardized test results (Alberta Learning, 2004; Gall, Gall, & Borg, 2007; Oliver-Hoyo & Allen, 2006; Saskatchewan Ministry of Education, 2000; Stake, 2004). The Responsive Evaluation Model (Stake, 2004) provided a philosophical framework for blending methods of assessment in order to gain a more complete evaluation of program effectiveness. The Saskatchewan Ministry of Education (2008) has constructed a rubric for assessing the effectiveness of special education service delivery. Regardless of chosen evaluation methods, test score reliability, calculated through a correlation coefficient is a major consideration from a Classical Test Theory point of view (DeVellis, 2003).

CHAPTER 3

3.1 Overview

In consideration of assumptions, limitations, and recommendations apparent in the literature review, Classical Test Theory methods have been selected in the current study. The rubric mandated for use by the Saskatchewan Ministry of Education represents roughly equivalent items for the assessment of a single underlying latent variable, inclusive practice. The rubric provided the basis for description on a continuum and did not represent degrees of correctness. Reliability through internal consistency measures was investigated.

In assessing reliability, the CTBS (Nelson Publishing, 1997) employed in the current study, was administered to more than 40,000 subjects (Nelson Publishing, 1998). Comparisons were made within the parameters established for the cohort of subjects. The universality of the test parameters was important to the current thesis because students with LD were subject to the same standards established by a sample that was made up of students that did not, for the most part, have LD. Traditional measurements, such as the CTBS, have been generally accepted as standard indicators of academic achievement within school populations. However, students with LD were not considered typical subjects for assessment with the CTBS (Nelson Publishing, 1997). The current thesis employed the results of the CTBS scores as a consequence of the school division's preference of use of the CTBS for measuring academic achievement regardless of student designation.

Program effectiveness was assessed with the rubric (Saskatchewan Ministry of Education, 2008) and also with the CTBS (Nelson Publishing, 1997). Reliability was assessed through a measure of internal consistency by calculating Cronbach's coefficient alpha for the rubric and by referring to reliability and validity information provided by the publishers of the

CTBS. Evaluation standards were assessed on the CTBS (Nelson Publishing, 1997) through norming procedures carried out by the test developers. Evaluation standards for the rubric (Saskatchewan Ministry of Education, 2008) were not prescribed by the rubric developers. Coefficient alpha was selected for the investigation based upon recommendations and limitations, evident in the literature review, with regards to: the number of items, the number of constructs being investigated, and sample size.

The rubric employed in the current study contained a Likert-style scale offering a continuous selection format. Factor analysis was deemed as inappropriate because the rubric employed in the current study was administered to only six participants and was composed of only 12 continuous items. The rubric was applied in its original unmodified form in order to provide an authentic assessment as mandated by the Saskatchewan Ministry of Education (2008). Given the parameters of the current study, and in consideration with limitations and requirements outlined in the literature review, coefficient alpha was selected as being most appropriate for assessing reliability.

In regards to the validation and reliability measures employed for the CTBS (Nelson Publishing, 1997), the current study relied on validity and reliability evidence provided by the test developer and publisher. Test validation and reliability coefficients received from the publisher are presented in chapter 4.

The current study incorporated both qualitative and quantitative measures, as encouraged by Stake (2004) and Gall, Gall, and Borg (2007). One facet of the current study presented an objective interpretation of program effectiveness through analysis of standardized test results. A second facet of the current study involved gathering subjective perceptions of participants within the setting being studied. A third facet involved an objective measure of internal consistency for

the rubric. A fourth facet of the current study involved investigating potential confounding variables arising from the current study design.

3.2 Nature of the Study

The purpose of this study was to apply and assess accepted accountability measures for the delivery of special education programming by: (1) performing a quantitative analysis of academic achievement of a group of rural Saskatchewan elementary students over a period of five years using historical CTBS (Nelson Publishing, 1997) results; (2) performing a program evaluation of special education service delivery by applying a recently mandated Ministry of Education rubric (Saskatchewan Ministry of Education, 2008) designed to assess program implementation compliance; (3) assessing the reliability of rubric results; (4) comparing the results of the two selected measures; and (5) investigating potential confounding factors in the current study design. In the current study, accountability measures were recognized and recommended in the two Saskatchewan government documents: *Directions for Diversity: Enhancing Supports to Children and Youth with Diverse Needs* (Saskatchewan Ministry of Education, 2000) and *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice* (Saskatchewan Ministry of Education, 2008).

The research questions relating to the intended purposes were:

1. Do standardized test results indicate greater achievement for a group of students with learning difficulties in an inclusive setting as compared to results for the same group in a noninclusive setting?

2. Are special education service delivery goals being achieved in the particular setting according to conditions specified on the Saskatchewan Ministry of Education rubric (2008)?
3. Is the mandated rubric a reliable measure of program effectiveness?
4. Are the results of the standardized achievement tests congruent with the rubric results?
5. What are potential confounding factors related to the current study design?

3.3 Participants

3.3.1 General Description of the School and School Division

The Living Sky School Division, which is located in west central Saskatchewan, provided data for this study. It provides educational services through 32 schools located in 19 communities. The division serves over 5500 students and employs over 900 staff (Living Sky School Division, 2008).

The site from which the data was gathered is a rural elementary school providing instruction for Kindergarten through Grade 6. The school provides educational programming to children from the immediate community and First Nations in the near vicinity. The percentage of First Nations children is generally at about 20 to 30% but fluctuates annually. The First Nations communities also offer elementary education programming. Consequently, some students move between the programs offered on the First Nation and the programs offered in the rural public school. The community is centered, primarily, around an agricultural economy. The data collection site serves approximately 110 to 120 students annually (Living Sky School Division, 2008). At any given time, approximately 16% of the school population would be identified as being *at-risk* using the same cutoff criteria, scoring below the 30th percentile, used

with CTBS (Nelson Publishing, 1997) scores to identify the target group of the current study (Living Sky School Division, 2008).

3.3.2 CTBS Participants

All information pertaining to participants and the treatment that they received was obtained from a written description provided by the regular classroom teacher that had been hired to create an inclusive classroom for the target group.

3.3.3 Target Group

The target group of the study was a group of students who were experiencing learning difficulties upon entering Grade 1 in 2002/2003. They were collectively identified as requiring remedial treatment. Consequently, the target group students were assigned to an educational assistant (EA), usually individually but occasionally in pairs. The students worked with their respective educational assistants in small rooms designated throughout the school. The routine occurred daily, for the duration of the school day, throughout the year. The students were essentially segregated from their classmates with at least one student never experiencing a regular classroom environment for two years. The target group continued to follow the same routine through grades 1 and 2.

In Grade 3 (2003/04), the target group students were placed in a large classroom with their peers. Halfway through the year, the classroom teacher requested assistance because the size of the class and the disparity in ability levels became difficult to manage. A regular classroom teacher was hired to teach a newly created class. The class was split, with 9 previously segregated students and one student from the original cohort forming one group of 10. The high achieving regular student specifically requested inclusion because of a friendship with a previously segregated student. The class size was set at ten based upon a school decision of the

maximum size that would allow for effective instruction.

The target group worked within the class of 10 with the newly hired regular classroom teacher for the morning. This smaller class received Language Arts, Math, Social Studies, and Science during that time. Some of the students had academic needs and some displayed behavioural issues. The target group continued to follow the regular Grade Three program with adaptations such as the following: (1) visuals were widely used, both for content and for expected behaviour; (2) each student had a help card as used in the Dawn Reithaug Orchestrating Success in Reading Program (Reithaug, 2002), which they used when they needed help from the classroom teacher or the Educational Assistant (E.A.); (3) assignments were often shortened or adapted. Students were given alternative ways to present their material, for example orally rather than written; (4) activity breaks were given, as needed; (5) students were offered alternate places to do their work rather than their desk. For example, one student recognized when his frustration level was rising and would choose to leave the classroom and work in the hallway outside of the room; (6) hands-on activities were included; (7) instructions were given orally and graphically, and the students would be asked to rephrase instructions back to the teacher; (8) handouts were used instead of note-taking; (9) Students were not placed under time constraints to read at their reading level and do accelerated reading quizzes; (10) some scribing by an educational assistant was used for specific students who experienced writing difficulties; (11) an educational assistant was present for much of the day to assist students in the classroom; and (12) day treatment procedures were used in the classroom to have students comply with behaviour expectations. Day treatment procedures in schools typically consist of teaching social skills and learning strategies. The programs usually follow a behavioural philosophy and focus on managing disruptive behaviours (Tse, 2006). The Reithaug Program (Reithaug, 2002) was a strategy for

assisting struggling readers and is based upon five components of reading: phonemic awareness; phonics; fluency; vocabulary; and strategies for comprehending text.

In the afternoon, the target group was integrated back to a larger class made up of themselves and their cohort peers. The total combined class size was twenty-three students. The morning split and afternoon integration continued the following year when the target group students were in grade 4 (2005/2006). The listed adaptations continued as students progressed into higher grades.

When the class went into Grade Five (2006/2007), the two groups were mixed differently. An effort was made to divide the groups evenly based on numbers of boys and girls and numbers of First Nations students in each group with less emphasis on academics, CTBS scores, and behaviour. E.A. assistance was provided for part of the core subject teaching time. Both classes had the same regular classroom teacher for Language Arts and the same teacher for Math and Science. For Grade Six, the students were split into two groups by a random draw.

The target group of students was identified as requiring remediation in grade 1 in the form of pullout with an educational assistant. Although no records are available from the school division to indicate the particular methods used to identify the target group as requiring remediation, examination of available CTBS scores indicated that all members of the target group had scored below the 30th percentile. In fact, the mean score of the target group was at the sixteenth percentile in comparison to the cohort group mean at the 82nd percentile. Percentile placement was based on national norms (Nelson Publishing, 1997). Consistent with the definition provided in chapter 1, the target group was considered to be *at-risk*. The progress of the target group students was investigated in comparison to their cohort group and a regular classroom group. The gains in achievement of each group, on its own, were investigated over the

years of the study.

3.3.4 Cohort Group

The cohort group was the group of age-equivalent classmates to the target group. Together with the target group, they formed the complete grade 1 class in the 2002/2003 school year. The cohort group was made up of a core of 10 continuously attending students. There were a few other students that joined and left the group over the time of the study. They were not included as part of the cohort group. The cohort group had the unique characteristic of having the lowest achievers removed from the group. In that respect, they were not completely representative of the greater typical elementary school population.

3.3.5 Regular Group

The regular group sample was also used in the comparison because, as previously stated, the cohort group was not academically representative of a regular classroom. The regular group consisted of one group of 10 students that were in continuous attendance for the entire period of the study. They were chosen from the grade 2 class of the 2002/2003 school year. They were chosen for their continuous participation on test measures over the course of the study. The regular group was included in the analysis because they represented a completely integrated population. No students were removed nor received treatment different from that received by other group members as a consequence of testing results. All groups, target, cohort, and regular were following a regular Saskatchewan Learning curriculum. Individual adaptations were used, as deemed appropriate by regular classroom teachers, across all groups.

3.4 Rubric Participants

Rubric participants consisted of the following: one Superintendent of Student Services; two principals, one from a grade 7 to 12 high school and one from a K to 6 elementary school;

two special education teachers, one from a grade 7 to 12 high school and one from a K to 6 elementary school; and one regular classroom teacher, who was the teacher of the target group of students for their first two years of inclusion. All rubric participants were employees of the same school division and were assessing the levels of special education service delivery within two different schools. The Superintendent of Student Services based the assessment upon service delivery within the entire school division. Only the two special education teachers and the Superintendent of Student Services could be defined as qualified special education professionals according to criteria provided in the definitions in chapter 1.

3.5 Instrumentation

3.5.1 Canadian Test of Basic Skills

The first instrument applied in the current study was the Canadian Tests of Basic Skills (CTBS) Form K Levels 6-8 and Form L Levels 10-12 (Nelson Publishing, 1997). The Canadian Tests of Basic Skills (CTBS) employed in the current study was a norm-referenced achievement test battery to assess achievement in several different academic areas such as: science; mathematics; language; and computer skills; from grades Kindergarten to 12. The 1997 CTBS Form L, levels 9-13 (Nelson Publishing, 1997) was available for grades 3 to 8, and Form K, levels 6 -8 are intended for use in elementary schools from Kindergarten up to and including approximately grade 7.

The CTBS (Nelson Publishing, 1997) assessed multiple knowledge areas. For example, Form K Level 5 included the areas of vocabulary, word analysis, listening, language, language total, mathematics, and a composite score. The Form K levels 7 and 8 included the following areas: vocabulary, word analysis, reading, listening, language, language total, mathematics concepts, mathematics problems, mathematics computations, math total, science, sources of

information, and a composite score. Depending upon the level being assessed and the desired information, the CTBS could provide relatively comprehensive information on student achievement.

The CTBS provided six scores, which were a raw score, developmental standard score, grade equivalent, national percentile rank, local percentile rank, and national stanine for each subject category as well as a composite category (Nelson Publishing, 1998). The raw score was simply the number of questions that were answered correctly. The developmental standard score was the median score that occurred at the various grade levels. For example, if the median raw score for grade 2 was 168 and a student in grade 2 scores 155, that student was functioning at a level that was lower than what would be expected from a student in grade 2. The raw median score provided a developmental comparison relative to other scores achieved by grade 2 students (Nelson Publishing, 1997).

The grade equivalent score was provided as two numbers representing the level of achievement in years and months. The grade equivalent score was also a relative comparison to the achievement levels of other students based on their expected achievements at monthly time intervals within each grade. Scores could be compared to national, division-wide, or local/ school percentiles.

The national percentile rank score indicated a student's ranking in comparison to all other Canadian students that took part in the CTBS at the same time of year. Possible scores ranged from 1 to 99 in order to account for potential movement of upper and lower limits as new cohorts were continually being tested. The percentile ranking indicated the percentage of Canadian students that scored below the given number. If student XYZ achieved a national percentile ranking of 99, that meant that 99% of all Canadian students who participated in the same testing

cohort scored less than student XYZ (Nelson Publishing, 1997). The current thesis employed the national percentile for comparison.

The local percentile rank that indicated a student's relative level of achievement in comparison to all other students in the same grade within the same school system. The Form K CTBS Norms Booklet (Nelson Publishing, 1998) stated that because some schools tested more or fewer students than the sampling specifications required, it was necessary to assign a weight. Because the individual school percentile rank was somewhat sensitive to small test differences, and could possibly be subject to assigned weights for ranking purposes, the national percentile rank appeared to be the best choice as a comparison measure of achievement over time and was recommended for that purpose (Nelson Publishing, 1998).

The national stanine score was a normalized standard score that fell on a scale of 1 to 9. It roughly represented the normal distribution of scores achieved by students with most scores occurring in the middle of the scale at number five, and fewer occurring as each end of the scale was approached at either 1 or 9 (Nelson Publishing, 1998).

The data available for the current study was obtained in various forms. Division scores were not available because schools within the division did not uniformly use the CTBS (Nelson Publishing, 1997) to measure achievement. Consequently, division-wide scores were not calculated. Within-school percentiles were not used because the school was not using the data to compare the performance of classes to each other. The two consistently available scores were national grade equivalency and national percentile ranking. The easier of the two to conceptualize, in terms of making comparisons for students from low elementary grades, was the national percentile ranking. For example, if a student's score fell below kindergarten level or, contrarily, fell above a grade 12 level, the score lost practical meaning. Contrarily, extreme

scores could be accounted for within the limits of national percentile ranking (Nelson Publishing, 1998).

The raw test score could not be used because scores for all subjects increased over time as achievement increased, with no established reference point on which to make a comparison year-over-year (Nelson Publishing, 1998). Grade level scores could not be used for the same reason (Nelson Publishing, 1998). Percentile ranking was used for the comparison because it was norm-referenced and indicated a student's relative placement according to national samples (Nelson Publishing, 1998). It also showed any movement up or down the ranks at any given time, not just relative to national norms, but also compared to the norm placement of peers, other local groups, and to oneself (Nelson Publishing, 1998).

The norm referencing for a regular class was included because the same-class peers, the cohort group, were not typical due to the removal of the lowest scores. Consequently, percentile scores for the regular group were used to establish the typical rate of achievement. There was disparity in sample sizes between other groups and the target (n = 7). The original group was made up of nine students, but the sporadic and discontinuous attendance of two students disqualified them from inclusion in the target group. Their class cohort sample included more subjects (n = 10). The regular group sample was the same size as the cohort group (n = 10). The disparate sizes were unlikely to affect the power of the analysis, provided that the correct statistical procedures were followed (Tanguma & Speed, 2000).

The Form K CTBS Norms Booklet (Nelson Publishing, 1998) recommended that, for time comparisons, the student percentile or school percentile score be used. Accordingly, the current study employed national percentile ranking scores as the unit of measurement to assess academic achievement. In the current study, classroom teachers, guided by provided instructions,

scored grades K-2 tests. Scoring for grades 3-6 tests was purchased from the test provider, so tests for those grades were not hand scored. A relatively comprehensive printout was provided giving basic statistical information on individual and group performance. Test samples are norm-referenced on a continual basis. The inferences made from the tests must be considered within the intended context (Nelson Publishing, 1998). The results provided a measure of student achievement in comparison to the national cohort. They also indicated individual achievement relative to previous levels of achievement (Nelson Publishing, 1998).

3.5.2 Validity and Reliability for Quantitative Measures

Messick (1989) emphasized that there are consequences to the interpretation and application of test results. Messick (1995) stated, "The challenge in test validation is to link these inferences to convergent evidence supporting them as well as to discriminant evidence discounting plausible rival inferences" (p. 8). Messick (1995) identified six distinguishable validity aspects which were "delineated emphasizing content, substantive, structural, generalizability, external, and consequential aspects" (p. 5). Messick also indicated that all forms of validity are aspects of construct validity. Construct validity encompasses all other types, but each type is important to the degree that specific inferences are made (Messick, 1995).

Kane (2001), and similarly Frisbie (2005), stated that it is the interpretations, inferences, and decisions that have validity, not the test itself. Kane (2001) indicated five purposes related to the interpretation and use of test results, which were: evaluation; generalization; extrapolation; explanation; and decision-making. Kane (2001) explained that there is always a balance to be considered. The level of evaluation might influence the level of explanation. The precision desired for a very refined evaluation might limit generalization or extrapolation. An instrument

designed for generalization might not have the precision required for individual evaluation (Kane, 2001).

The CTBS (Nelson Publishing, 1997) forms are updated continually through collaboration with scholars at the University of Iowa including Drs. Hoover, Hieronymous, Frisbie, and Dunbar. Canadian curriculum consultants include Ethel Shaw-King, Professor Emeritus from the University of Calgary. Dr. Shaw-King was the originator of the CTBS project in Canada in the 1960's and has continually worked toward improving content validity within the Canadian context (Anderson, Bernier, & Hebert, 2004).

CTBS (Nelson Publishing, 1997) content is continually updated with consideration for evolving curriculum guides and teaching procedures. Item selection for the CTBS (Nelson Publishing, 1997) forms used in this thesis was done through consultation with educators across Canada (Anderson, Bernier, and Hebert, 2004). Educator comments, resulting from the item review, were used as a guide to item modification or deletion. A sample in excess of 3000 students completed the items in a field trial in the fall of 1993. Based on data analysis of field trial results, final items were selected (Nelson Publishing, 1998).

Information from the Form K CTBS Norms Booklet (Nelson Publishing, 1998) indicated specific norming procedures. As part of an effort to obtain a national representative sample, the test norms were based upon results from over 40,000 students from various geographic and demographic areas including all types of school systems across Canada. Demographic participation was relatively universal. However, students were excluded in cases where their level of English was insufficient to enable understanding of test items. Students with significant mental challenges were also excluded from the sample (Nelson Publishing, 1998).

Construct validity is rooted in the original ITBS (Hoover, Dunbar, & Frisbie, 2007),

which underwent approximately 25 years of administration and modification before being adapted for Canadian use. Content validity is addressed through ongoing expert panel review (Anderson, Bernier, and Hebert, 2004). The test makers indicated that they have attempted to establish convergent validity by correlating the CTBS (Nelson Publishing, 1997) with the Canadian Cognitive Abilities Test (CCAT) (Nelson Publishing, 1997). The Form K CTBS Norms Booklet (Nelson Publishing, 1998) explains the specific sampling instructions to be followed in the co-administration of the CTBS (Nelson Publishing, 1997) and the CCAT (Nelson Publishing, 1997). In an attempt to preserve convergent validity of the CTBS (Nelson Publishing, 1997) and the CCAT (Nelson Publishing, 1997), any student that received modified test administration procedures was not included in the normative results (Nelson Publishing, 1998).

Test developers have continued to address concerns raised through reviews and studies of the CTBS (Nelson Publishing, 1997). Contact with the Nelson Publishing representative for western Canada resulted in data from the 2002 norming procedure. Correlations were provided for CTBS (Nelson Publishing, 2002) and CCAT (Nelson Publishing, 2002) samples that were co-normed. The data indicated correlations of .61 to .79 on quantitative mathematics measures, with correlations increasing with grade level. Non-verbal measures of spelling were the lowest correlated measures ranging from .40 to .44. Most correlations on other areas of assessment were in the .5, .6, and .7 ranges (Nelson Publishing, 2002).

Information provided by Nelson Publishing indicated that the test provider has assessed the reliability of differences between the various levels of the CTBS (Nelson Publishing, 2002) test forms as well as the differences between subject areas. The correlations for the reliabilities of differences ranged from lows around .45 to highs around .80 with most being in the .60 to .70

range (Nelson Publishing, 2002). The reliability of differences offered some evidence of discriminant and concurrent validity in that the various test levels were measuring evolving concepts that were distinguishable through varying group performances (Nelson Publishing, 2002).

The following Table 3.1 presents reliabilities of differences data from the 2002 norming procedure using levels 11 and 12 for a number of learning areas including vocabulary, reading, various language tasks, various mathematics skills, science, reference materials, and maps and diagrams. Permission for publishing was obtained from Nelson Publishing (see Appendix A).

Table 3.1. *Reliabilities of differences between Levels 11 & 12 of the CTBS.*

Level 11, Grade 5 - above diagonal												
	V	R	L1	L2	L3	L4	M1	M2	M3	SC	RM	M&D
V	---	0.50	0.74	0.67	0.66	0.55	0.63	0.61	0.79	0.53	0.57	0.62
R	0.54	---	0.77	0.68	0.66	0.54	0.67	0.63	0.80	0.52	0.59	0.64
L1	0.73	0.77	---	0.65	0.61	0.69	0.73	0.73	0.79	0.78	0.70	0.76
L2	0.66	0.66	0.58	---	0.38	0.58	0.60	0.61	0.66	0.67	0.61	0.65
L3	0.70	0.69	0.63	0.38	---	0.54	0.61	0.60	0.68	0.65	0.56	0.62
L4	0.57	0.57	0.65	0.54	0.54	---	0.63	0.58	0.76	0.57	0.53	0.62
M1	0.67	0.69	0.74	0.61	0.64	0.65	---	0.37	0.62	0.62	0.59	0.53
M2	0.61	0.63	0.74	0.62	0.64	0.60	0.40	---	0.65	0.56	0.55	0.48
M3	0.80	0.80	0.78	0.65	0.69	0.75	0.65	0.68	---	0.77	0.74	0.72
SC	0.55	0.58	0.77	0.67	0.70	0.64	0.61	0.57	0.77	---	0.55	0.55
RM	0.64	0.65	0.70	0.58	0.60	0.57	0.61	0.55	0.72	0.64	---	0.55
M&D	0.63	0.63	0.75	0.63	0.65	0.64	0.53	0.45	0.71	0.52	0.56	---
Level 12, Grade 6 - below diagonal												

The CTBS (Nelson Publishing, 1997) provides an index of general standing in comparison with the concurrent test cohort. The CTBS (Nelson Publishing, 1997) is not an indicator of specific curricular achievement but does give insight into relative academic progress (Nelson Publishing, 1998). Efforts have been made to continually improve the validity and reliability of the test forms. The CTBS (Nelson Publishing, 1997) is adapted from the ITBS (Hoover, Dunbar, & Frisbie, 2007), which now has a 75-year history of administration and development through the University of Iowa and notable faculty in the area of measurement and evaluation (Anderson, Bernier, and Hebert, 2004). The test battery is focused on general skills in various subject areas. As a tool to gauge general skill development and progress, the CTBS (Nelson Publishing, 1997) appears to be a useful instrument (Anderson, Bernier, & Hebert, 2004). Cautions regarding administration, use, and interpretation are provided (Nelson Publishing, 1998).

3.5.3 Rubric for Program assessment

Graf (2004) stated that rubrics are becoming an increasingly popular form of assessment and that rubrics do not just assess. They set the framework for the subject's preparation and the examiner's evaluation of a specific concept. By assisting the subject's preparation, the rubric becomes operative in the instructional process. The rubric is operative by being part of the preparation for assessment and also serving as the final assessment. Graf (2004) recognized that rubrics are useful on a number of levels ranging from facilitation of instruction, to communication with students and parents, to guiding instructional procedures.

The second assessment instrument (see Appendix F) employed in the current study is the rubric published in the living document titled *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice*

(Saskatchewan Ministry of Education, 2008). The rubric, in its original form, was adopted for use in this study. Twelve forced choice items based on a grid containing four possible responses, for each of the twelve principles presented, were used to investigate teachers' and administrators' perceptions of the degree of successful program implementation (i.e., not evident, emerging/developing, evident, and exemplary). For example, three of the scored item categories were *inclusionary practice, fostering independence,* and *service coordination.* The wording in the survey instrument was prescribed and remained unchanged as the instrument was mandated for use in its original form until updated by the Saskatchewan Ministry of Education.

3.5.4 Validity and Reliability for Qualitative Measures

Moskal (2003) presented rubrics and instruction as complementary concepts and asserted that performance assessment could be efficiently carried out using rubrics. At the same time, rubrics could facilitate instruction by providing a guide for performance and a focus for instruction. Moskal (2003) indicated that reliable scoring should be possible if the rubric was well designed and if care was taken to address specific characteristics of the performance. Achieving reliability is not a major issue with subjective scoring (Moskal, 2003). The current study, through the use of a rubric, assessed the degree of implementation of a specific model for program delivery. Though it was not concerned with individual performance and instruction, it was consistent with Moskal's (2003) assertion that assessment can guide, focus, and facilitate program implementation.

Other researchers also addressed the subjectivity of rubric scoring (Frey & Schmidt, 2007). It was argued that any assessment that used a subjective rubric to assess tasks that might be part of what would be performed in the real world was an authentic assessment. Validity evidence is unnecessary because the test is authentic and, therefore, is automatically valid (Frey

& Schmidt, 2007). However, the automatic validity of authentic assessment was questioned in the current thesis.

The mandated rubric (Saskatchewan Ministry of Education, 2008) employed in the current study was a relatively recent document with no accompanying documentation in relation to validity and reliability measures. However, because methods are available to investigate construct validity and reliability, one of the tasks of the current study was to calculate a measure of internal consistency of the rubric results.

3.6 Data Collection

3.6.1 Source of Data

The Director of the school division was contacted during the planning of the study, was informed of the study's parameters and intent, and subsequently agreed to support the research (see Appendix B). Student application for Approval of a Research Protocol was submitted to the Office of Research Services at the University of Saskatchewan in May 2008, and exempted on June 19, 2008 (see Appendices C and D). A participant agreement, briefly describing the research project and the expectations of participants, was sent by email to the rubric evaluators requesting their participation (see Appendix E). The school division was contacted to obtain a de-identified, anonymous, historical data set to be used for secondary analysis. Due to the nature of the analysis and the data being de-identified, parental consent was not required. Rubric evaluators, upon agreeing to participate, were sent a copy of the rubric and the instructions for completion (see Appendix F). The de-identified data file was received at the end of June 2008.

3.6.2 Confidentiality

Participant confidentiality of rubric evaluators was ensured by the assignment of a lettered code system. Data will be locked, secured, and stored at the University of Saskatchewan

for a minimum of five years. In consideration of risk to students, the study was deemed to be below minimal risk through the use of de-identification of data at the source. Evaluators were assured that protection of their privacy was obligatory, and that they were free to withdraw from the study at any time. An email confirmation was required from the evaluators who volunteered. Participating evaluators and stakeholders were invited for data debriefing, upon contact with the primary researcher, when the thesis is completed.

3.7 Data Analysis

Data was entered and analyzed using the Statistical Package for the Social Sciences (SPSS). A quality check of the data verified that all entered data were not outside the expected range of scores. A data quality check was performed to monitor for missing data according to the numeric codes received in the data set. Missing scores were not included in the statistical analysis. If a subject was missing a score at any point over the duration of the testing period, that subject's score was disqualified from use in the comparisons. If scores were missing for a particular period of time, that interval was excluded from comparisons. Independent student variables that included grade attended, year of attendance, coded group assignment, and coded identification numbers were collected. Student achievement scores from the Composite category of the CTBS (Nelson Publishing, 1997) were also collected. The research questions guided the analyses employed for this thesis.

3.7.1 CTBS Analysis

Data analysis included a comparison over time for treatment, both within groups and between groups, on one dependent variable. The groups considered were a target group, a cohort group, and a regular group. A comparison of the target group results was performed for the year-over-year data. Also, there was a parallel comparison to a cohort group as well as to a regular

class group that was used as a representative of the regular school population. Specific statistical methods included a Mixed Model Factorial ANOVA, a one-way ANOVA, and t-tests.

3.7.2 Assumptions

According to Tabachnick and Fidell (2007), ANOVA has three main assumptions that must be met in order for results to be generalizable. Independent samples are assumed. Normal distribution is assumed. Homogeneity of variance is assumed. The independent variable or variables must be categorical. The dependent variables must be continuous and at least at the interval level. Results of tests of assumptions are discussed in the following paragraphs.

3.7.3 Level of Significance

The level of statistical significance for the current study is set at $p \leq .05$. The current study is unique in the specific school division. So, in that sense, the current study was exploratory. A significance level of $p \leq .05$ is consistent with that employed in other quantitative studies of student achievement (Huber, 2001; McDonnell et al., 2003; Rea, McLaughlin, & Walther-Thomas, 2002; Wiener & Tardiff, 2004). A quasi-experimental design was employed in the current study.

3.8 Independent Samples

3.8.1 Establishment of Representativeness

In order to preserve as much statistical uniformity as possible within groups and across the time period of the study, a group of continuously attending students was selected to represent regular class achievement. The regular group was necessary for giving context to the performance of the target group. The target group was selected based upon the identification of their learning difficulties. Therefore, the target group scores were the lowest of those collected for the Grade 1 class of 2002/2003. The cohort group was, essentially, the remainder of the

students from the Grade 1 class of 2002/2003. The cohort group could be compared to the target group but could not be considered as representative of regular class achievement because the lowest achievement scores had been removed from the group. Therefore, in order to make a meaningful comparison, it was necessary to establish a group that could be considered as displaying regular levels of achievement when compared to other regular classes. The logical choice was the Grade 2 class of 2002/2003. The regular group was in continuous attendance over the time period of the study. However, it was necessary to establish that the regular group was in fact representative of a typical class in the school. Representativeness was established by comparing mean national percentile rankings (Nelson Publishing, 1997) for all classes in the 2002/2003 school year.

A one-way ANOVA was performed for all classes, (*n = 6),* in attendance in the 2002/2003 school year. For the purposes of the analysis, the Grade 1 class achievement was measured before the split into the target and cohort groups. The rationale for the combined score was to preserve the complete spectrum of scores before any outside intervention had occurred. All other classes were treated the same way, as one complete unit with no scores removed.

3.8.2 Rubric Analysis

Rubric analysis commenced with the assignment of numerical values to the rubric categories with 1 being assigned to the description depicting the lowest level of program implementation (not evident), 2 being assigned to the next highest level of description (emerging/developing), 3 being assigned to the next highest level of description (evident), and 4 being assigned to the description depicting the highest level of program implementation (exemplary) (Saskatchewan Ministry of Education, 2008). Subsequently, calculations of score means and standard deviations were performed. Rubric-based data was analyzed according to

specific procedures, including a measure of internal consistency, recommended by Stake (2004) for the responsive evaluation method of assessing program quality.

3.9 Research Questions

3.9.1 Research Question 1

The first research question posed was: Do standardized test results indicate greater achievement for a group of students with learning difficulties in an inclusive setting as compared to results for the same group in a noninclusive setting?

The first research question involved achievement of the target group of students. However, the performance results would be meaningless without having some sort of context within which to be compared. For example, an increase or decrease in performance might be an artefact of general school climate or procedures. Without a comparison to a group that was representative of the general school population, results might be misinterpreted, leading to incorrect or unsubstantiated inferences. Inclusion of a group that was representative of the general school population would provide an indicator of performance that could be expected within the regular school routine. Without comparison to the cohort group, the same misinterpretations might be made. For example, change might be due to characteristics inherent across the entire original grade 1 class that includes both the target and cohort groups. Inclusion of the cohort group would provide an indicator of performance for the rest of the original grade 1 class. Parallel performances would suggest influential factors beyond the treatment received by the target group.

The comparison to the cohort and regular groups served to provide contextual indicators of the types of performances that could be expected without treatment. The target group represented the lowest achievement levels of the original grade 1 class. The cohort group

represented the highest achievement levels. Inclusion of both the cohort and regular groups in comparisons might offer some insight into whether potential treatment effects were a result of the regression to the mean phenomenon. The cohort group and regular group were necessary for between-subject comparisons, addressing the issue of universal effects of programming not related to the treatment of the target group. The target group served as its own control when addressing the research question regarding the results of the two types of treatment received.

An initial test using a one-way ANOVA was implemented to establish the representativeness of the regular group. The regular group scores on the Composite Score from the CTBS (Nelson Publishing, 1997) for 2003 were compared to all other 2003 classes including the original combined grade 1 class. The one-way ANOVA results were used to establish that there were no significant differences between the 2003 classes and the regular class in national percentile rankings for the Composite score of the CTBS (Nelson Publishing, 1997). If no significant differences existed, the regular group would be used as an indicator of expected achievement, in terms of national percentile ranking, of a regular class within the elementary school.

Tabachnick and Fidell (2007) recommended a univariate analysis be used for the primary data analysis based upon the current study design. The characteristics of the current study prompted the use of a Mixed Model Factorial ANOVA for initial data analysis (Arkkelin, 2007; and Stevens, 2007). The design was relatively uncomplicated with two discrete independent variables. The first independent variable had three levels being target, cohort, and regular groups. The second independent variable consisted of four distinct time intervals. There was one continuous dependent variable (CTBS Composite Score) measured at four distinct time intervals. The dependent variable scores fell within a range of 1 to 99. The current research

questions involved measuring within and between-group differences over time. There were no covariates in the analysis.

If an interaction effect were evident in the Factorial analysis, a series of paired t-tests would be used to compare achievement results within groups. The rationale for using t-tests as opposed to a Repeated Measures ANOVA was that the data set was incomplete for 2006. With the missing data, a continuous linear effect could not be assumed. Therefore, the data was analyzed using paired comparisons. Adjustments for inflated chance of Type I error due to simultaneous multiple paired comparisons were made by default within the SPSS program (Tabachnick & Fidell, 2007).

The current study had an unbalanced design, which predicated the use of specific sums of squares equations in the ANOVA (Tanguma & Speed, 2000). Though Lewsey, Gardiner, and Gettinby (2001) and Langsrud (2003) made a case for using Type II sums of squares for an unbalanced design, the current study used the Statistical Package for the Social Sciences (SPSS) default, which is Type III sums of squares (Wielkiewicz, 2000). The rationale for the use of Type III sums of squares in the current study was based on simulation results presented in the article by Tanguma and Speed (2000), which stipulated specific measures for specific study designs. Notably that for unequal \sample size with no missing data, Type III is appropriate (Tanguma & Speed, 2000). Tanguma and Speed's results were consistent with the SPSS default.

3.9.2 Research Question 2

Are special education service delivery goals being achieved in the particular setting according to conditions specified on the Saskatchewan Ministry of Education rubric (2008)?

Numerical values were assigned to the rubric categories with 1 being assigned to the description depicting the lowest level of program implementation and 4 being assigned to the

description depicting the highest level of program implementation. Rubric results were analyzed using the mean and standard deviation of scores in order to assess the current level of program delivery.

3.9.3 Research Question 3

The third research question posed was: Is the mandated rubric a reliable measure of program effectiveness?

Results of the survey were used as part of an analysis to test for internal consistency by calculating Cronbach's Coefficient Alpha in order to establish the reliability and construct validity of the rubric results.

3.9.4 Research Question 4

The fourth research question posed was: Are the results of the standardized achievement tests congruent with the rubric results?

This question was addressed by comparing the standardized test results for student academic achievement to the rubric results on program effectiveness.

3.9.5 Research Question 5

The fifth research question posed was: What are potential confounding factors related to the current study design?

This research question was included on order to investigate confounding elements present in the current study as a result of the current study design.

CHAPTER 4: RESULTS

4.1 Overview

The current study had five purposes: (1) to address the need for more quantitative studies to evaluate special education student success within the inclusive classroom setting as compared to a noninclusive setting; (2) to apply a program assessment rubric for special education services to determine the level of special education service delivery in the specified location; (3) to evaluate the reliability of the results of the rubric mandated by the Saskatchewan Ministry of Education (2008); (4) to compare the results of the standardized student achievement tests with the rubric results, (5) and to discuss potential confounding factors related to the current study design.

The current study performed a secondary analysis of historic data collected from standardized test results collected in a rural western Canadian school division from 2002 to 2007. Student group performances over the time period of the study were compared at annual intervals.

The study also gathered data through a survey of a recently mandated rubric released by the Saskatchewan Ministry of Education (2008). Information gathered indicated the status of inclusionary practice implementation within the subject school division. A measure of internal consistency of the rubric results was performed in order to gain statistical insight into the properties of the rubric. Investigation into measurement issues pertaining to the use of standardized test results and rubric application within the current context was carried out.

4.2 Establishment of Representativeness

Results of the one-way ANOVA indicated that the overall mean achievement for all classes was (M=58.5) on a national percentile ranking for the Composite Score on the CTBS (Nelson Publishing, 1997). The greatest achievement discrepancy was between Grade 2

(*M*=69.5), the highest, and Grade 3 (*M*=47.09), the lowest national percentile rankings for the 2002/2003 school year. However, the difference was not significant at $p = .26$. Therefore, the regular group scores appeared to be representative of scores of a regular class. The following sections explain the results of the study in the order of the research questions posed.

4.3 Research Question *1*

Do standardized test results indicate greater achievement for a group of students with learning difficulties in an inclusive setting as compared to results for the same group in a noninclusive setting?

In order to put the results into a meaningful context, comparisons were made between the target group, the cohort group, and the regular group. A primary analysis of the data was implemented in order to determine whether or not meaningful differences existed.

4.4 Primary Analysis

A 3 X 4 Factorial ANOVA with repeated measures on one factor was employed to analyze group differences in achievement across the four time intervals for the study. ANOVA output indicated no significant effects for *time* as F (2.554) = 1.264, $p > 0.05$. However, a significant *time*group* effect was indicated at F (5.107) = 3.216, $p < 0.01$. Between-subjects output indicated a significant effect with F (2) = 39.376, $p < .001$. A post-hoc one-way ANOVA was utilized to compare group achievement results at each time interval. Results indicated that the target group results were significantly different from the cohort group and the regular group ($p < 0.01$). There was no significant difference between scores for the cohort group and regular group ($p > 0.05$).

There was no significant difference between the achievement levels of the cohort and regular groups, and the achievement patterns were almost parallel. The cohort group scores were

expected to be higher than the regular group scores because the lowest scores had been removed from the cohort group. Despite the removal of the lowest scores when the target group was established, cohort group achievement was not significantly greater and the achievement pattern mirrored the regular group. On the other hand, the target group performance pattern was quite different from the other two groups (See *Figure 4.4*). The target group performance was in the same direction but more extreme than the other two groups for only one of the comparisons. On the other two comparisons, the target group performed in the opposite direction of both the cohort and regular groups.

4.5 Examination of Effects

Paired t-tests were employed to investigate effects indicated in the Factorial ANOVA results. Paired t-tests were selected over a one-way ANOVA design because of missing data records for the 2005/2006 school year. The missing records, limiting the number of comparisons, compromised the linearity of the comparisons. Therefore, SPSS was used to run a series of t-tests.

4.6 Regular Group

The regular group consisted of students that were one grade ahead of the target and cohort groups. Results from paired t-tests for the regular group over the four time intervals indicated no significant differences in achievement performance as measured by national percentile ranking for Composite scores on the CTBS (Nelson Publishing, 1997). Results of the regular group t-tests are reported in Table 4.1. The regular group achievement pattern graph is presented in *Figure 4.1*.

Table 4.1. *Results of paired samples t-tests for the regular group.*

Years	Mean	SD	Std. Error	t	df	Sig. 2-tailed
2003-2004	9.7	21.432	6.777	1.431	9	.186
2003-2005	5.7	14.637	4.629	1.231	9	.249
2003-2007	4.7	19.351	6.119	.768	9	.462
2004-2005	-4.0	31.330	9.907	-.404	9	.696
2004-2007	-5.0	22.624	7.123	-.702	9	.500
2005-2007	-1.0	21.161	6.692	-.149	9	.885

alpha = .05

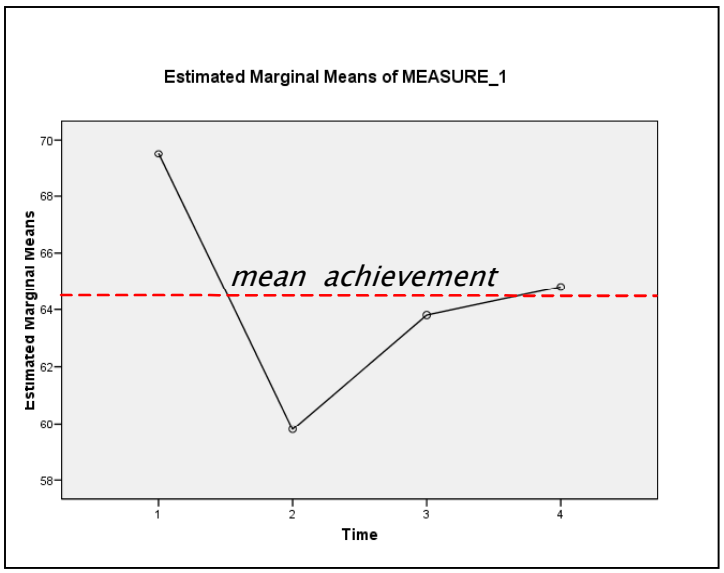

Figure 4.1 Diagram showing the regular group achievement pattern graph.

4.7 Cohort Group

The cohort group consisted of the age-equivalent peers of the target group. The cohort group represented Results from paired t-tests for the cohort group over the four time intervals

indicated significant differences in achievement performance as measured by national percentile ranking for Composite scores on the CTBS. There was a significant decrease in achievement from *time 1* to *time 2* (t (9) = 3.617, p = .006). The mean achievement score at *time 2* (M = 69.6) was significantly lower than the mean at *time 1* (M = 82.2). Though there was an increase for the next time interval, the difference remained significant between *time1* and *time 3* (t (9) = 2.694, p = .025). The mean achievement score at *time 3* (M = 74.0) was significantly lower than the mean at *time 1* (M = 82.2). A further increase in achievement resulted in no significant difference between *time1* and *time 4*. There were no other significant differences. Results of the cohort group t-tests are reported in Table 4.2. The cohort group achievement pattern graph is presented in *Figure 4.2*.

Table 4.2. *Results of paired samples t-tests for the cohort group.*

Years	Mean	SD	Std. Error	t	df	Sig. 2-tailed
2003-2004	12.6	11.017	3.484	3.617	9	.006
2003-2005	8.20	9.624	3.043	2.694	9	.025
2003-2007	6.40	15.233	4.817	1.329	9	.217
2004-2005	-4.40	11.909	3.766	-1.168	9	.273
2004-2007	-6.2	12.874	4.071	-1.523	9	.162
2005-2007	-1.8	17.171	5.430	-.331	9	.748

alpha = .05

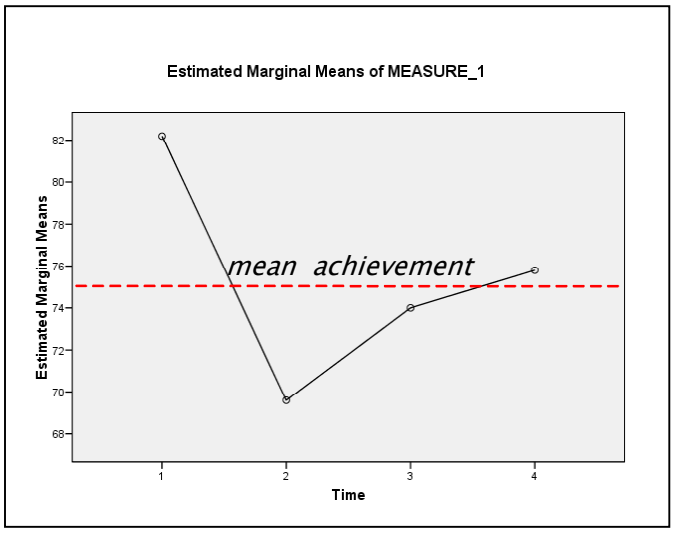

Figure 4.2 Diagram showing the cohort group achievement pattern graph.

4.9 Target group

The target group was composed of students with learning difficulties identified as *at-risk*, entering grade one, in the 2002/2003 school year. Results from paired t-tests for the target group over the four time intervals indicated significant differences in achievement performance as measured by national percentile ranking for Composite scores on the CTBS. There was an increase from *time 1* to *time 2*, but the increase was not significant ($p > .05$). There was a significant increase in achievement from *time 2* to *time 3* (t (6) = - 2.772, p = .03). Results indicated that the mean achievement score at *time 3* (M = 39.86) was significantly higher than the mean at *time 2* (M = 30.57). Consequently, there was also a significant difference between *time 1* and *time 3* (t (6) = - 3.542, p = .012). The mean achievement at *time 1* was (M = 16.43) as compared to *time 3* (M = 39.86). Though there was a subsequent decrease for the last time interval, the difference remained significant between *time1* and *time 4* (t (6) = -3.612, p = .011). The mean achievement at *time 1* was (M = 16.43) as compared to *time 4* (M = 35.14). There were no other significant differences. Results of the target group t-tests are reported in Table

4.3. The target group achievement pattern graph is presented in *Figure 4.3*.

Table 4.3. *Results of paired samples t-tests for the target group.*

Years	Mean	SD	Std. Error	t	df	Sig. 2-tailed
2003-2004	-14.143	19.274	7.285	-1.941	6	.100
2003-2005	-23.429	17.501	6.615	-3.542	6	.012
2003-2007	-18.174	13.708	5.181	-3.612	6	.011
2004-2005	-9.286	8.864	3.350	-2.772	6	.032
2004-2007	-4.571	10.998	4.157	-1.100	6	.314
2005-2007	4.714	9.569	3.617	1.303	6	.240

alpha = .05

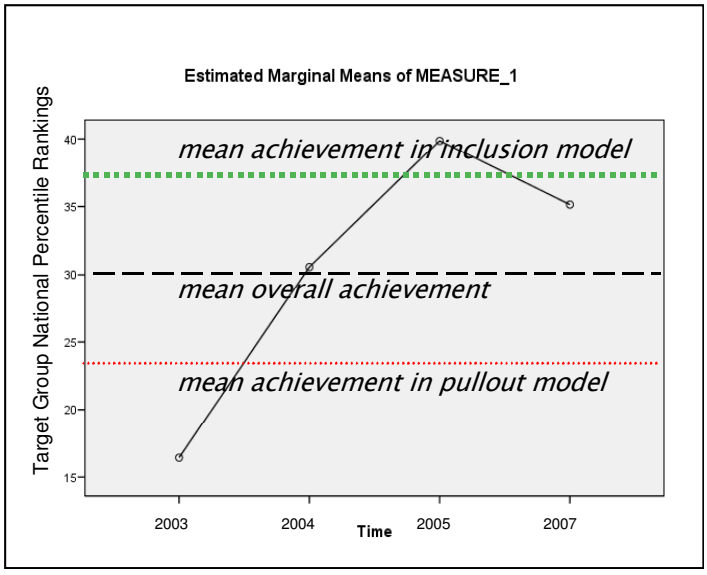

Figure 4.3 Diagram showing the target group achievement pattern graph.

In relation to research question one, there were significant differences in achievement results ($p < .05$). Students in the target group demonstrated significantly lower academic achievement results than both the cohort group ($p < .05$) and the regular group ($p < .05$) over the

course of the comparisons. There were no significant differences between the regular group and the cohort group ($p > .05$). Patterns of achievement indicated that the target group results reflected an independent pattern while both the cohort and regular group patterns appeared to parallel one another, as shown in *Figure 4.4.*

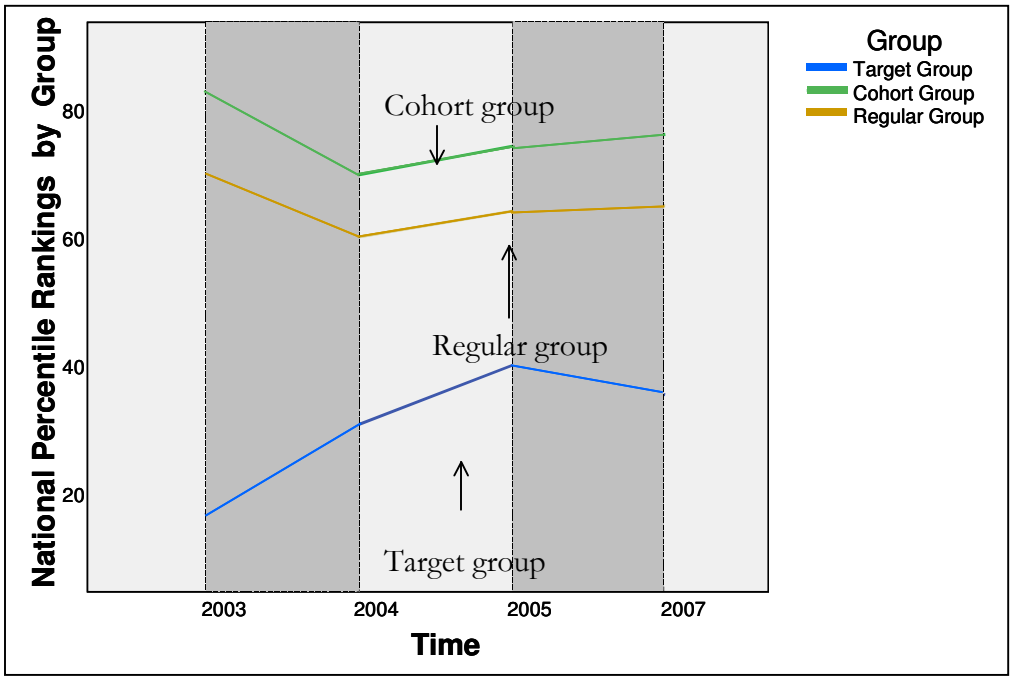

Figure 4.4 Diagram showing achievement pattern graph for all groups.

4.10 Comparison of Pullout Model to Inclusion Model

The pullout model refers to removal from regular classroom activities and receiving service delivery by educational associates. The inclusion model refers to being included in a regular classroom setting with the various supports necessary to maximize a successful outcome for each individual student, irrespective of disability. A paired t-test was employed to compare the mean achievement for the target group while being instructed within the pullout model to achievement while being instructed within the inclusion model. The comparison enlisted the

mean national percentile ranking for the Composite score on the CTBS under each of the two instructional models experienced by the target group. Results indicated a significant difference in achievement between target group rankings while in the pullout model as compared to the inclusion model (t (13)= - 3.061, p = .009). The mean achievement for the inclusion model was (M = 37.50) as compared to the pullout model (M = 23.50) (See *Figure 4.5*). In relation to research question number one, there were significant differences in achievement results between the pullout and inclusion models, as shown in *Figure 4.5*. Students in the target group demonstrated significantly higher academic achievement results in the inclusion model as measured on the national percentile ranking on the Composite score of the CTBS.

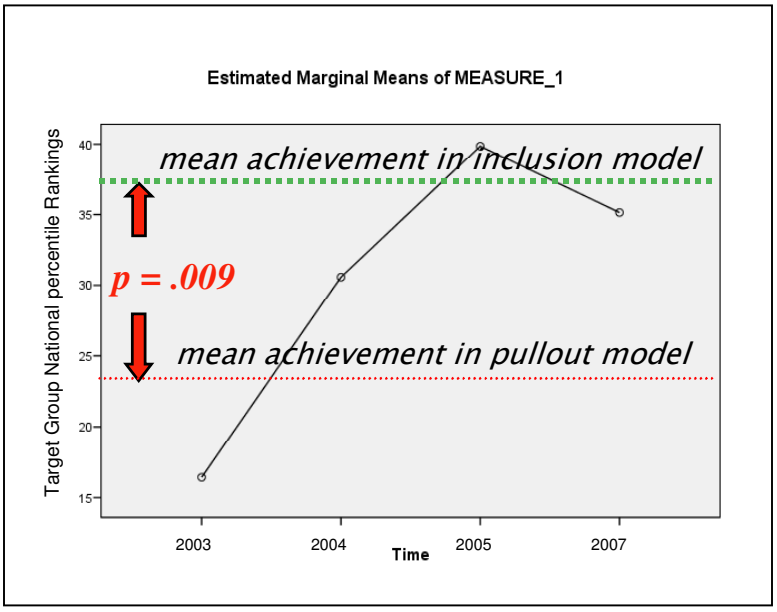

Figure 4.5 Diagram showing comparison of regular group achievement in inclusive and pullout settings.

4.11 Research Question 2

Are special education service delivery goals being achieved in the particular setting according to conditions specified on the Saskatchewan Ministry of Education rubric (2008)?

4.12 Analysis of Rubric Results

Numerical values were assigned to the rubric categories with 1 being assigned to the description depicting the lowest level of program implementation and 4 being assigned to the description depicting the highest level of program implementation. Subsequently, calculations of score means and standard deviations were performed on the ordinal scores.

All raters selected ratings in either category 2 or 3 for each item with the exception of the regular classroom teacher who selected category 4 for items 4, 8, and 12. No rater selected category 1 on any item. The range for all evaluators' scores was 1 over all items, with the exception of the regular classroom teacher, who had a score range of 2 over all items. The regular classroom teacher's range per item was 1 for all items with the exception of items 4 and 12. The mean and standard deviation was calculated for each evaluator's scores on the rubric. The means for evaluator scores on the entire list of items ranged from M = 2.42 to M = 3.08. The mean for all evaluators on the entire rubric was M = 2.67. The means for each item ranged from M = 2.33 to M = 3.17. The mean for all items on the entire rubric was M = 2.67.

The standard deviation of raters' scores ranged from SD = .39 to .67. The mean standard deviation for raters was .51. The standard deviation for item scores ranged from SD = .41 to .82. The mean standard deviation for items was .54. Results indicated that scores were relatively tightly clustered near the mean given the few categories of the scale. The three items scored with a *4* demonstrated the greatest standard deviation scores. In relation to the rubric categories, no item received a rating of 1 (not evident). All other items received a rating of either 2 (emerging/developing) or 3 (evident). Only three items, selected by the regular classroom teacher, received a rating of 4 (exemplary). They were item 4 (assessment, SD = .82), item 8 (referral process, SD = .63), and item 12 (special education teacher qualifications, SD = .63).

Results of analysis indicated that there was complete agreement that program implementation was at least at the *emerging/developing* stage or had attained the *evident* stage. Only the regular classroom teacher felt that program implementation was at the *exemplary* stage for the principles related to assessment, referral process, and special education teacher qualifications. Special education program delivery appeared to be primarily established and was assessed as being either *emerging/developing* or *evident*.

4.13 Research Question 3

Is the mandated rubric a reliable measure of program effectiveness?

4.14 Internal Consistency of Rubric Results

DeVellis (2003) indicated specific assumptions and criteria, earlier stated, for the use of Cronbach's coefficient alpha (α). The author also indicated that a reliable scale could provide increased statistical power without relying on increased sample size. As DeVellis (2003) stated, increased internal consistency reliability scores suggest increased construct validity. Consequently, in accordance with rationale presented in chapter 2, the current study employed Cronbach's co-efficient alpha in order to ascertain the level of internal consistency of the rubric results.

Coefficient alpha scores for the rubric varied according to which items were included. The overall score for the scale was $\alpha = .69$. Rubric scores obtained from subjects are presented in Table 4.4. Results of the calculation of the correlation coefficient are presented in Table 4.5.

Table 4.4. *Scores obtained from administration of the rubric.*

Item	1	2	3	4	5	6	7	8	9	10	11	12
Subject												
ABC	2	2	3	2	2	3	3	2	3	3	2	3
DEF	3	3	3	4	2	3	3	4	3	3	2	4
GHI	3	2	3	2	2	2	3	3	3	2	2	2
JKL	3	2	3	2	2	2	2	3	2	2	3	3
MNO	3	3	2	3	3	3	2	3	3	3	2	3
PQR	2	3	2	3	2	3	2	3	3	2	2	3

Table 4.5. *Results of correlation coefficient calculations for rubric responses.*

Coefficient alpha calculations	
variance of sum totals	9.87
sum of variances for items	3.67
sum items/sum totals	0.37
1- sum items/sum totals	0.63
k	12
k-1	11
k/k-1	1.09
(k/k-1)/[(1-(sum items/sum totals)]	0.69
alpha	0.69

According to DeVellis (2003) a coefficient score of .65 to .70 is minimally acceptable, with .65 being the minimum acceptable level when decisions concerning groups of individuals

are involved. Coefficients from .70 to .80 are respectable (DeVellis, 2003). However, it was also indicated that longer scales tend to elicit higher coefficients (Clark & Watson, 1995). Given the relative brevity of the current rubric, and according to the benchmark suggested by DeVellis (2003), $\alpha = .69$ is an acceptable correlation coefficient.

Following the initial analysis including all items, the rubric results were analyzed by subset. The teacher qualification subset consisted of only one item. Therefore, it was not possible to perform the coefficient alpha calculation. The inclusion subset consisted of seven items and yielded a coefficient of $\alpha = .39$. The intervention subset consisted of four items and yielded a coefficient alpha of $\alpha = .64$. Neither subset attained the acceptable minimum benchmark as an acceptable coefficient ($\alpha = .65$). However, when all items were included, the coefficient alpha surpassed the minimum acceptable benchmark ($\alpha = .65$). Results appeared to be consistent with assertions by Clark and Watson (1995) and DeVellis (2003) that more items, up to approximately 40, tend to yield higher coefficient alpha results if the items are related. In the current study, it appeared that including most of the 12 principles was necessary to establish acceptable levels of internal consistency.

After having established the rubric's correlation coefficient including all items, some items and combinations of items were removed from the scale based upon interitem correlations. Interestingly, the rubric coefficients were quite stable over the item removal process with almost all correlation coefficients falling within the .61 to .71 range. However when item 4 was removed, the coefficient dropped to .55, the lowest of all comparisons. Consequently item 4 was considered to be a central component to the underlying construct. With the removal of item 3 alone, the coefficient increased to .73. With the removal of item 11 alone, the coefficient increased to .74. With the removal of item 3 and item 11, the coefficient increased to .78.

4.15 Research Question 4

Are the results of the standardized achievement tests congruent with the rubric results?

Results from *Research question 1* and *Research question 3* indicated significantly greater scores on the CTBS (Nelson Publishing, 1997) between target group rankings while in the pullout model as compared to the inclusion model (t (13)= - 3.061, p = .009). Results from administration of the mandated rubric (Saskatchewan Ministry of Education, 2008) suggest that special education program delivery was primarily established and was assessed as being either *emerging/developing* or *evident*. Further discussion is presented in chapter 5.

4.16 Research Question 5

What are potential confounding factors related to the current study design?

Results of the investigation into the potential confounding effects of the current study design indicated numerous confounding factors. Specific confounding factors are discussed in chapter 5.

CHAPTER 5: DISCUSSION

5.1 Summary

The purpose of this chapter is to discuss: (1) the findings of the study; (2) the implications of the results; (3) the limitations of the study; and (4) suggestions for future research directions. The goal of this thesis was to provide information, to the Living Sky School Division and to the Saskatchewan Ministry of Education, on the implementation and success of the inclusion model in a rural Saskatchewan setting and to provide information on assessment instruments employed in the study.

5.2 Purpose and Procedures

The purpose of this study was to apply and assess accepted accountability measures for the delivery of special education programming by: (1) describing the academic achievement of a group of rural Saskatchewan elementary students over a period of five years using historical CTBS results; (2) describing program status by applying a recently mandated Ministry of Education rubric designed to assess program implementation compliance; (3) assessing the reliability of the rubric; (4) comparing the results of the two selected measures; and (5) investigating the potential confounding factors related to the current study design. In this thesis, accountability measures were recognized and recommended in the Saskatchewan government documents: *Directions for Diversity: Enhancing Supports to Children and Youth with Diverse Needs* (Saskatchewan Ministry of Education, 2000) and *School Division Student Services – Service Delivery Model Rubrics 2007-2008: Facilitating and Monitoring Effective Practice* (Saskatchewan Ministry of Education, 2008).

5.3 Findings

1. A small group of students with learning difficulties, identified as *at-risk*, achieved significantly higher results on a standardized academic achievement test, Canadian Test of Basic Skills (Nelson Publishing, 1997), while participating in an inclusive educational setting as compared to results while participating in a pullout setting.
2. Special education program delivery, as evaluated by the Saskatchewan Ministry of Education rubric (2008), was assessed overall as *emerging/developing* to *evident*.
3. The results of the mandated rubric (Saskatchewan Ministry of Education, 2008) achieved an acceptable correlation coefficient of $\alpha = .69$.
4. One item, pertaining to program assessment, of the mandated rubric (Saskatchewan Ministry of Education, 2008) was found to be central to maintaining an acceptable level of reliability as calculated in the correlation coefficient.
5. Two items, pertaining to outside agencies and parental participation, of the mandated rubric (Saskatchewan Ministry of Education, 2008) were found to be confounding factors in maintaining an acceptable level of reliability as calculated in the correlation coefficient.
6. Both the CTBS and Rubric results indicated positive results for program effectiveness.

5.3.1 Significant Increase in Academic Achievement Results

Historical data from CTBS tests was analyzed for three groups of students. The target group was composed of students with learning difficulties identified as *at-risk*. The cohort group consisted of the age-equivalent peers of the target group. The regular group consisted of students that were one grade ahead of the target and cohort groups. The Composite score from the CTBS (Nelson Publishing, 1997) was selected for analysis based upon the scope of the investigation and also in consideration of statistical and inferential limitations of selecting alternate subset

scores. National percentile ranking (NPR) scores, from the CTBS (Nelson Publishing, 1997), were used in the comparisons in an attempt to mediate group differences and to provide a common comparison measure across all subjects and treatments. Results of the analysis indicated that the target group achieved significantly higher NPR scores on the Composite subset of the CTBS (Nelson Publishing, 1997) while in the inclusive setting as compared to NPR scores within the pullout setting. NPR scores for the cohort group were consistently higher, though not significantly, than those of the regular group. The target group NPR scores were consistently significantly lower than both the regular and cohort group rankings. The pattern of achievement of the cohort group paralleled that of the regular group. The pattern of achievement for the target group was in the same direction as both the regular and cohort groups for one of the time intervals subsequent to the initial rankings. However, The target group achievement was in the opposite direction to both the regular and cohort groups on two of the three total time intervals. The pattern differences indicated that students with learning difficulties had unique achievement patterns when compared to other students not identified as having learning difficulties. This finding also suggested that alternate measures of academic achievement might be more appropriate for students with learning difficulties. Results also supported previous studies that have demonstrated that students with learning difficulties in an inclusive setting could make significant gains on standardized tests measuring academic achievement (McLeskey & Waldron, 2000; Peterson & Hittie, 2003; Rea, McLaughlin, & Walther-Thomas, 2002).

5.3.2 Program Assessment using Rubric Results

The current study employed the recently released and mandated rubric (Saskatchewan Ministry of Education, 2008) to assess program implementation of the inclusion model and special education service delivery within that model. The reason for employment of the rubric

was twofold. First, the rubric would provide an indicator of the level of service being provided. Second, the results generated by responses to the rubric would be either disputed or endorsed based on a measure of internal consistency.

The rubric consisted of 12 principles with 3 subsets. The first subset of 7 principles addressed the concept of inclusion. The second subset of 4 principles addressed the concept of intervention. The third subset consisted only of a single item and addressed the concept of teacher qualification. Participants included one Superintendent of Student Services, two principals, two special education teachers, and one regular classroom teacher. Numerical values were assigned to the rubric categories with 1 being assigned to the description depicting the lowest level of program implementation and 4 being assigned to the description depicting the highest level of program implementation. Subsequently, calculations of score means and standard deviations were performed on the ordinal scores.

Results of analysis indicated that there was complete agreement that program implementation was at least at the *emerging/developing* stage or had attained the *evident* stage. Only the regular classroom teacher felt that program implementation was at the *exemplary* stage for the principles related to assessment, referral process, and special education teacher qualifications. Special education program delivery appeared to be primarily established and was assessed as being either *emerging/developing* or *evident*.

Further analysis showed that each subset of the rubric, analyzed on its own, failed to meet the minimum acceptable benchmark for demonstrating internal consistency. However, the entire scale including all items elicited a coefficient alpha of $\alpha = .69$, which met the minimum acceptable benchmark of $\alpha = .65$ (DeVellis, 2003). Further analysis included the systematic removal of items from the rubric. Results suggested that *assessment* was an important item

within the scale. Results further suggested that *parental involvement* and *service coordination* detracted from the internal consistency of the scale. Interestingly, *parental involvement* and *service coordination* both represent components that fall outside of the general education setting. The removal of the two detracting items resulted in a coefficient alpha of $\alpha = .784$ for the remaining scale items. Results of the analysis, generally, suggested that the rubric could elicit reliable results when assessing the quality of special education service delivery. However, the original scale should not be further reduced to subsets of small numbers of items because calculation of correlation coefficients of subscales resulted in values below the minimal acceptable benchmark of $\alpha = .65$ for each subscale on its own. Results further suggested that rubric results might demonstrate increased internal consistency with the removal of the items related to components outside of the immediate educational setting. The correlation coefficient values increased to their highest levels with the removal of the two suspected detracting items.

5.4 Limitations

The Special Education Services Program Delivery rubric (Saskatchewan Ministry of Education, 2008) employed in this thesis was a relatively recent document with no accompanying documentation in relation to validity and reliability measures. The document was created with the goal of facilitating and monitoring effective practice. Instructions accompanying the document suggested that only one assessment would be made for each school division, likely carried out by the Superintendent of Student Services. Unless other forms of data collection have supported an opinion, it is not advisable to collect only one opinion (Stake, 2004). Though the reliability of results suggested an acceptable correlation coefficient, prudent practice suggests that school divisions devise a strategy for more comprehensive data collection that will facilitate accurate reporting on documents such as the rubric employed in this thesis.

It is important to consider possible confounding factors within the process, especially since a portion of the current study relied on secondary analysis, foregoing any type of experimental control. The lack of experimental control created uncertainty in attributing specific cause and effect relationships of actions to results. In addition to the lack of experimental control inherent in the secondary data analysis employed in the current thesis, other possible confounding factors might have been present. In order to increase the generalizability of results, potentially confounding factors must be considered when establishing an experimental design.

In the current study, environment was the main factor being investigated. It is possible that historical outside influences affected the students' results on the CTBS (Nelson Publishing, 1997). However, the target group was compared to itself as well as other groups in an attempt to account for common environmental influences. Though scores suggested that the target group members reacted to their learning environment in very different ways than the other groups, scores also suggested that the target group reacted more positively in the inclusive setting.

It is possible that gains made by the target group were at least partially due to maturation factors or physical or psychological changes in participants. The CTBS (Nelson Publishing, 1997) was employed in an attempt to account for the physical maturation factor. Students were compared to all others of the same grade cohort and compared nationwide. Achievement patterns from the current study could suggest that students with learning difficulties mature earlier, academically, and at a more constant rate but also peak earlier. The results indicated a steady improvement and eventual decline in the target group while the other groups indicated an initial decline followed by a continuous upward trend. Results could suggest differences, not as much in physical maturation, but in maturation of learning processes.

Changes in the measuring instrument or instrumentation, a pretest/posttest was another

factor to consider. The CTBS (Nelson Publishing, 1997) is a standardized test with standardized administration procedures. The target group was administered the test following the same procedures as all other students. Scoring was carried out according to directions accompanying the test. In addition, the CTBS has a long history of development, validation, and reliability checks. The CTBS is, generally, a widely accepted test. However, there is some question as to whether the CTBS is an appropriate measure for students with learning difficulties. It is unclear, given that students with learning difficulties in the current study demonstrated different patterns of achievement in relation to other learners, if the CTBS is optimally designed to assess academic achievement in students with learning difficulties.

A further factor was statistical regression or the tendency for subsequent test scores to move toward the mean. Statistical regression is increasingly inevitable as scores fall further from the mean. This phenomenon tends to be a more serious threat when there are a small number of comparisons. However, in this thesis, scores were compared over a five-year span. The mean scores for two years within one setting were compared with the mean scores for two years within another setting. Using the mean scores in each setting over a relatively long time period was intended to mediate the regression effects that might be obvious if comparing only two scores. Because the mean might change as scores change, it would be potentially more difficult to demonstrate change over a long period of time unless a significant effect was present, as was indicated in some of the comparisons.

A likely confounding factor was differential selection or confounding effects due to the experimenter's inability to hold participant selection and/or experience constant. The current study did not follow a pure experimental design. Group selection was purposeful and site selection was convenient. The intent of the study was to investigate achievement within a

specific group of students within a specific setting. The unique experience of the target group in this particular study was a rare occasion to make a comparison of achievement results within the same group across different treatments. It would likely be unethical to purposely subject a group of students to treatment that could potentially delay or negatively affect their learning. However, the specific situation encountered in this thesis appeared to be a unique opportunity to investigate a phenomenon without imposing upon students or affecting their educational experiences. Within the preexisting situation, there were potentially confounding factors such as teaching style, exposure to changing peer classrooms, and various methods employed.

Experimental mortality, subject attrition or exclusion of subjects because of missing data or other factors, was a factor in the present study. The target group originally contained 10 members, the cohort group originally contained 13 members, and the regular group originally contained 17 members. As a result of inconsistent attendance, withdrawal, and missing test scores, the target group consisted of seven static members over the course of the study, while the cohort group consisted of 10 static members, and the regular group consisted of 10 static members. In a study that was focused on student results over a five-year period, attrition was expected. Though smaller sample size affects the power of inferences, reliability measures attempted to address generalizability issues arising from mortality of subjects.

The current study attempted to address the selection-maturation interaction issue through the use of comparison groups near, or equivalent to, the age of target group subjects. The use of national percentile rankings from standardized test results were used to acquire as fair a comparison as possible in relation to age and grade level.

Experimental treatment diffusion, when treatment procedures diffuse into other group treatments over time was a potential factor in the current thesis. The current study was centered

on the secondary analysis of annual standardized test results. As much as possible, test administration was uniform across years and across groups. There was no intentional treatment that could be diffused between proximate groups. However, it is possible that there was a diffusion of effects between groups when they were routinely placed together as part of the inclusion process. If treatment effects diffused between groups resulting in a regression of results from each group to the shared group's mean, one would expect a smaller effect to be demonstrated. Differing patterns of achievement in this thesis did not support a common positive or negative trend between groups. It is possible that diffusion had a positive effect on the target group in the inclusive setting, which would lend support for inclusive philosophy.

The extent to which one can generalize from the experimental population to a defined population is a major consideration. Given the small sample sizes used in this thesis, the power of inferences and generalizability of results was severely limited. It was the current study's intent to provide a description of a unique situation, with a focus on the accountability process and related measures. The focus was on issues related to the derivation and subsequent use of results as well as on the results themselves. Findings from the current investigation of the process might be generalizable insofar as identifying and addressing measurement issues that could potentially arise in the assessment of accountability. For example, procedures for assessing accountability and related measurement concerns were an important part of the current thesis. Findings might also be valuable for future comparisons at a local level.

The extent to which personological variables interact with treatment effects or how individual abilities and characteristics influence scores was a potential confounding factor in the current thesis. This issue had potential to be a factor in the current study's assessment of accountability through the use of standardized test results. There was a wide array of strategies

employed with the target group students. In that respect, it is impossible to know if a different array of strategies would have produced a more positive or negative effect. It was also impossible to know the magnitude of effects, if any, caused by daily life issues originating outside of the school and the resulting levels of learning and achievement. Because the current study relied upon secondary analysis of data, no manipulation within the treatment setting was possible. Therefore, the confounding factors that accompany the use of historical data must be accepted and acknowledged.

An attempt was made in this thesis to provide as complete a description as possible based upon available resources and sources of information. However, the data was historical, employees have left the situation, memories were potentially subject to failure and fabrication, records were missing, limited or poorly kept, and not all information was recorded within the daily routine. It is recommended that the current study findings be considered within the context of the findings of other related research.

Multiple treatment inference could lead to unclear treatment effects because subjects were exposed to more than one treatment. This issue could pose a potential threat to external validity. It was obvious from the description of strategies employed with the target group that it was impossible to attribute success or failure to a specific strategy. The current study followed the general direction of other related studies, but the general nature of the treatment resulted in broad comparisons.

The Hawthorne effect might have been present in the current situation simply because it involved a change of treatment from an isolated setting to a more socially interactive setting. The change of setting could have potentially resulted in improved performance through increased motivation to participate in the new setting. However, the use of scores gathered over a long

period of time was intended to offset or, at least, detect unusual increases or declines in performance. In this thesis, achievement patterns did not appear to support the Hawthorne effect.

In the current study, there was no direct involvement of the experimenter or apparent bias in the procedures. It was earlier acknowledged that teacher traits might have contributed to any effect. The strategies within the inclusive environment were indirectly assessed through standardized test results. The teacher was unaware that the student's scores would undergo a secondary analysis at a later date. However, it was unknown if the teacher intentionally acted in a biased manner in relation to the treatment of the target group.

There was a potential interaction of history and treatment effects in the current study. The treatment of the target group had been drastically changed based upon the principal's support for inclusionary practice. Additionally, a new teacher was hired to facilitate the successful integration of previously isolated students. It was possible that results would have been different if the principal had the philosophy imposed upon her and the teacher had been transferred from what was perceived as a more desirable position to a more difficult task. The change was implemented consistent with recommendations originating from a change in philosophy through the Saskatchewan Ministry of Education. So, it was likely to be seen as a priority and an opportunity to embrace a newly encouraged philosophy.

In the current study, there was potential for effects due to the forms of assessment used for the dependent variable. The CTBS forms vary as students move ahead in grades. The CTBS also consists of several subsets of subject areas. The interrelation of numerous subsets created a problem for analysis, specifically in relation to multicollinearity. It was decided, within the current study, that to control for interrelated variables would have resulted in an extremely

reduced list of possibilities for measurement and consequently a narrow view of student performance. It was also decided that as broad a representation as possible of annual achievement was desirable. Therefore, the composite score was chosen because it represented the overall achievement of the students' combined performances in all subsets. The weakness of that choice was that attribution of gains to any specific area was not possible.

Though significant results were obtained for the academic achievement of the target group, results were not generalizable to other groups of students with learning difficulties. The sample sizes in the current study were too small to make any meaningful inferences with regard to academic achievement. In response to the statistical limitations imposed by the small sample size in the current study, Classical Test Theory methods were employed in an attempt to demonstrate the internal consistency of the rubric results.

A portion of the study involved the use of a recently released rubric that was not accompanied by documentation of validity or reliability. In addition, the sample size of rubric evaluators was small. Though internal consistency measures indicated an acceptable level of internal consistency, caution should be exercised when using the rubric until greater numbers can support the current findings. Given that the rubric (Saskatchewan Ministry of Education, 2008) was recently mandated, initial analyses, such as that performed in the current thesis, have value in establishing a base of reliability. Insofar as the rubric (Saskatchewan Ministry of Education, 2008) applies to the specific situation being evaluated in the current thesis, the correlation coefficient indicated a minimal acceptable level of reliability of results.

Standardized test, CTBS (Nelson Publishing, 1997), results were employed to measure academic achievement in students with learning difficulties. It is unclear as to whether that was an appropriate instrument for assessment of students with learning difficulties. The CTBS

Norms Booklet (Nelson Publishing, 1998) indicated that modifications in test administration might be necessary with students with learning difficulties in order to elicit meaningful results. However, modification undermines the validity and reliability of the standardized test. In the current thesis, the CTBS (Nelson Publishing, 1997) was administered uniformly across all groups.

5.5 Conclusion

In the current study, results indicated that, in the specific location being investigated, students with learning difficulties, identified as *at-risk*, achieved significantly higher results on a standardized academic achievement test, Canadian Test of Basic Skills (Nelson Publishing, 1997), while participating in an inclusive educational setting as compared to results while participating in a pullout setting. Students with learning difficulties, in the current study, benefited significantly, on academic achievement, from being placed into an inclusive setting.

Special education program delivery, as evaluated by the current study, was assessed overall as *emerging/developing* to *evident*. Results suggest that there is room for improvement in the delivery of special education services. The evaluation criteria listed in the mandated rubric (Saskatchewan Ministry of Education, 2008) indicates the general areas and specific facets of delivery to be met.

The results of the mandated rubric (Saskatchewan Ministry of Education, 2008) including all items elicited a coefficient alpha of $\alpha = .69$, which met the minimum acceptable benchmark of $\alpha = .65$ (DeVellis, 2003). Results from administration of the rubric employed in the current study were reliable as they pertained to the specific situation being evaluated. Results of the systematic removal of items from the rubric suggested that *assessment* was an important item within the scale. The correlation coefficient values increased to their highest levels with the

removal of the two suspected detracting items, *parental involvement* and *service coordination*. Results of the analysis, generally, suggested that the rubric, including all items, could elicit reliable results when assessing the quality of special education service delivery.

The use of historical data within a secondary analysis design was associated with numerous potential confounding factors related to the generalizability of results. A more stringent design, if possible to implement, might elicit more generalizable results.

5.6 Implications for Practice

Results from the current study suggested that special education program delivery in the specific location evaluated could be improved by continuing to strive toward higher levels of service delivery as described on the Saskatchewan Ministry of Education (2008) rubric. However, the novelty of the employed rubric (Saskatchewan Ministry of Education, 2008) and the lack of reliability and validity data prompted caution and scrutiny of results. The rubric (Saskatchewan Ministry of Education, 2008) was mandated. Therefore, it could not be altered unless alterations originated from the Saskatchewan Ministry of Education. Though the current study employed six participants in the rubric (Saskatchewan Ministry of Education, 2008) assessment, it was intended for application by only one person, the Superintendent of Student Services, in each school division. The practical implications suggest that reliability measures be carried out on a larger scale, perhaps including special education teachers, students, and parents in the service delivery evaluation.

The rubric results demonstrated an acceptable level of internal consistency in its current form, suggesting its continued use. However, it might be beneficial to revise the items, possibly creating a new subset, or perhaps devising an alternative method of measurement for the items

pertaining to factors outside of the immediate educational setting. A potential procedural change could involve having the outside stakeholders assess that specific area of service delivery.

Given the differing achievement patterns demonstrated by the target group as compared to the regular and cohort groups, an alternate form of assessment, other than the CTBS (Nelson Publishing, 1997) might be considered for students with learning difficulties. Students with learning difficulties do not demonstrate the same achievement patterns as other students. Therefore, it appears counteractive to measure them with the same processes and instruments. Efforts could be made to apply alternate assessment instruments for tracking the academic achievement of students with learning difficulties.

5.7 Implications for Future Research

Academic achievement results for students with learning difficulties, as collected through the use of standardized tests, have historically been sparse and inconsistent (Lindsay, 2007). The assessment of academic achievement in students with learning difficulties could be facilitated with the development of new assessment instruments that are more closely tied to the adapted or modified instructional strategies employed. Similarly, new instruments require validation, especially when program decisions with subsequent widespread effects are the consequence of evaluation. Future research could contribute to the development and validation of new assessment instruments and strategies.

Researchers and stakeholders could benefit from coordinated efforts. Sample size is a major concern when considering varying forms of statistical analyses and attempting to produce generalizable results. Researchers could potentially increase the array of statistical analyses available to them, and potentially increase generalizability of results, by combining data from various jurisdictions. Such an undertaking would likely require coordination between several

stakeholders including, schools, school divisions, researchers, and the Saskatchewan Ministry of Education. Such collaboration might be achieved through a centralized data collection agency where all stakeholders, at all levels, share data and results. All stakeholders could potentially benefit from establishing common forms of assessment that have been validated and shown to produce reliable results.

The goal of this thesis was to provide information, to the Living Sky School Division and to the Saskatchewan Ministry of Education, on the implementation and success of the inclusion model in a rural Saskatchewan setting and to provide information on assessment instruments employed in the study. Current thesis results are important because they have provided a basis for the quantitative study of academic success in rural Saskatchewan education. The current thesis has also applied the responsive model, combining quantitative and qualitative methods, to provide an example of corroboration of evidence when performing program assessment. It has also established a beginning point for future empirical studies of educational programming in Saskatchewan and, potentially, other rural educational settings. Though results suggest favorable outcomes for academic achievement of students with learning difficulties while in an inclusive setting, future research might involve common assessments on a larger scale and alternate common assessments. Rubric results indicate that special education service delivery is not implemented at an exemplary level. Future research might contribute to the development of methods of implementation and measurement of program effectiveness. Scale validation appears to be a specific area of need. More centralized collection of data and sharing of common resources is encouraged.

References

Alberta Learning. (2004). *Alberta initiative for school improvement: Improving student learning.* Provincial Report for Cycle 1 (2000-2003). School Improvement Branch, Edmonton: AB.

Anderson, J. O., Bernier, J. J. & Hebert, M. (2004). *Review of the canadian tests of basic skills, forms 7 & 8.* Buros Institute Mental Measurements Yearbook: 11 (AN: 11011006), [Electronic version]. Retrieved January 28, 2008, from the Ebsco Host Research Databases website: http://web.ebscohost.com.cyber.usask.ca/ehost/results?vid=2&hid=115&sid=fd3a97328f164183b8617f16c1d6acb8%40sessionmgr106

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27-30. Retrieved July 6, 2008, from ProQuest Education Journals database. (Document ID: 761703241).

Begeny, J. C., & Martens, B. K. (2007). Inclusionary education in Italy: A literature review and call for more empirical research. *Review of Remedial and Special Education, 28*(2), 80-94. Retrieved January 17, 2008, from ProQuest Education Journals database. (Document ID: 1252813121).

Benard, B. (1991). *Fostering resiliency in kids: protective factors in the family, school, and community*. Western Center for Drug-Free Schools and Communities, Portland, OR. Retrieved on July 23, 2008 at http://nwrac.org/pub/library/f/

Blum, R. W., & Libbey, H. P. (2004). School connectedness—strengthening health and education outcomes fore teenagers: Executive summary. *Journal of School Health*, *74*, 231–232. Retrieved November 1, 2007, from ProQuest Education Journals database. (Document ID: 700576051).

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *Vol. 73*, No. 2. (Summer, 2003), 125-230.

Bost L. W., & Riccomini P. J. (2006). Effective instruction: An inconspicuous strategy for dropout prevention. *Remedial and Special Education, 27*(5), 301-311. Retrieved November 1, 2007, from ProQuest Education Journals database. (Document ID: 1143555191).

Bunch, G., Lupart, J., & Brown, M. (1997). *Resistance and acceptance: educator attitudes to inclusion of students with disabilities*. Social Sciences and Humanities Research Council of Canada. Ottawa, ON. Retrieved October 21, 2008 from: http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/b9/f6.pdf

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in scale development. *Psychological Assessment, 7*(3), 309-319.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed). Thousand Oaks, CA: Sage Publications.

Dirling, J. (1999). Inclusion: Enhancing resilience. *Preventing School Failure, 43*(3), 125. Retrieved October 31, 2007, from ProQuest Education Journals database. (Document ID: 42993507).

Frattura, E., & Capper, C. (2006). Segregated programs versus integrated comprehensive service delivery for all learners. *Remedial and Special Education, 27* (6), 355-364.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *The Journal of Secondary Gifted Education, 18*(3), 402-423. Retrieved November 7, 2007, from ProQuest Education Journals database. (Document ID: 1322893721).

Frisbie, D. A. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice 24* (3), 21–28. doi:10.1111/j.1745-3992.2005.00016.x

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Toronto, ON: Allyn and Bacon.

Government of Canada. (1985). *Canadian Human Rights Act (R.S., 1985, c. H-6)* Department of Justice Canada. Retrieved October 20, 2008, from Government of Canada Web site: http://laws.justice.gc.ca/en/showdoc/cs/H-6///en?page=1

Government of Saskatchewan. (1995). *The Education Act, 1995 (sec. 85(1))E-0.2 - Education Act, 1995.pdf*. Retrieved November 30, 2007 from Government of Saskatchewan. Publications Centre. Legislation. Queen's Printer: http://www.publications.gov.sk.ca/details.cfm?p=487&c=5

Government of Saskatchewan. (2007). *Provincial funding manual 2007-2008.* Education Finance and Facilities Branch Department of Learning March 2007. Retrieved November 2, 2007, from Government of Saskatchewan Web site: http://www.learning.gov.sk.ca/Default.aspx?DN=0a214474-b29f-4b18-9f67-0254e54fb6e6/funding_manual_07-08.pdf

Graf, D. (2004). Rubrics: Score one for the instructor and one for the student. *Distance Learning, 1*(4), 43-45. Retrieved November 7, 2007, from ProQuest Education Journals database. (Document ID: 809428291).

Holahan, A., & Costenbader, V. (2000). A comparison of developmental gains for preschool children with disabilities in inclusive and self-contained classrooms. *Topics in Early Childhood Special Education, 20*(4), 224. Retrieved November 1, 2007, from ProQuest Education Journals database. (Document ID: 65934950).

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa test of basic skills: Form A*. Rolling Meadows, IA: Riverside Publishing.

Hopson, J., Newton, J., Ward, D., Beriault, L., Moen, C., Garrity, F., et al. (1999). *Report on the Committee on Integration of Students with Special Needs in the Classroom*. Retrieved November 29, 2007 from http://saskschoolboards.ca/Communications/is_spnd.html

Huber, K. D., Rosenfeld, J. G., & Fiorello, C. A. (2001). The differential impact of inclusion and inclusive practices on high, average, and low achieving general education students. *Psychology in the Schools. 38*(6). 497-504.

Jenkins, J. R., Jewell, M., Leicester, N., O Connor, R. E., et al. (1994). Accommodations for individual differences without classroom ability groups: An experiment in school restructuring. *Exceptional Children, 60*(4), 344. Retrieved November 1, 2007, from ProQuest Education Journals database. (Document ID: 1638909).

Kane, M. T. (2001). Current concerns in validity theory. *JEM, 38*(4), 319. Retrieved October 16, 2007, from ProQuest Education Journals database. (Document ID: 99529673).

Klem, A. M., & Connell, J. P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273.

Kline, T. (2005). *Classical test theory: Assumptions, equations, limitations, and item analyses*. Retrieved on July 3, 2008 at http://www.sagepub.com/upmdata/4869_Kline_Chapter_5_Classical_Test_Theory.pdf

Langsrud, O. (2003). ANOVA for unbalanced data use Type II instead of Type III sums of squares, *Statistics and Computing 13*, 163–167.

Lewsey J.D., Gardiner W.P., & Gettinby G. (2001). A study of Type II and Type III power for testing hypotheses from unbalanced factorial designs. *Communications in Statistics-Simulation and Computation 30*, 597–609.

Lindsay, G. (2007). Educational psychology and the effectiveness of inclusive education/mainstreaming. *British Journal of Educational Psychology, 77*(1), 1-24. DOI: 10.1348/000709906X156881; (*AN 24519626*).

Living Sky School Division No. 202. (2006). *1040 Guiding Principles and Beliefs.* Board of Education Policies. Retrieved on July 12, 2008 from the Living Sky school Division website: http://www.lskysd.ca/board/documents/MicrosoftWord-1040GUIDINGPRINCIPLES.pdf

Living Sky School Division No, 202. (2008). Retrieved August 6, 2008, from Living Sky School Division Website: http://www.lskysd.ca/

Lupart, Judy. (1998). Setting right the delusion of inclusion: implications for Canadian schools. *Canadian Journal of Education, 23*(3), 251-264. Retrieved February 5, 2008, from ProQuest Education Journals database. (Document ID: 413120331).

Manset, G., & Semmel, M. I. (1997). Are inclusive programs for students with mild disabilities effective? A comparative review of model programs. *The Journal of Special Education*, *31*(2), 155-180.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist,56*(3), 227-238.

McDonnell, J., Thorson, N., Disher, S., Mathot-Buckner, C., Mendel, J., & Ray, L. (2003). The achievement of students with developmental disabilities and their peers without disabilities in inclusive settings: An exploratory study. *Education & Treatment of Children, 26*(3), 224-237.

Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher 18*(2), 5-11.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice 14*(4), 5–8. doi:10.1111/j.1745-3992.1995.tb00881.x.

Miller, J. J. (2004) *Citizenship Education Policy at the School District Level.* Education Commission of the States' National Center for Learning and Citizenship. 2004 Policy Brief Collection, Issue Paper, April 2004. Retrieved at http://www.ecs.org/html/educationIssues/StateNotes/2004_PolicyBrief_Collection.pdf

Ministers of Social Services of Canada. (1998). *In unison: A canadian approach to disability issues.* Released at the Federal-Provincial-Territorial Meeting of Ministers of Social Services. Toronto, Ontario, October 27, 1998. Retrieved on October 29, 2008 at http://www.scics.gc.ca/cinfo98/83063002_e.html

Moskal, M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, *8*(14). Retrieved November 7, 2007 from http://PAREonline.net/getvn.asp?v=8&n=14

Nelson Publishing. (1997). *Canadian tests of basic skills: Directions for administration; Form K, level 6.* Scarborough, ON, Canada: ITP Nelson.

Nelson Publishing. (1997). *Canadian tests of basic skills: Directions for administration; Form L, levels 9-14*. Scarborough, ON, Canada: ITP Nelson.

Nelson Publishing. (1998). *Canadian tests of basic skills: Norms booklet; Form K, levels 5-8.* Scarborough, ON, Canada: ITP Nelson.

Nelson Publishing. (2002). *Canadian tests of basic skills: CTBS form K, levels 5 to 17/18 - correlations to CCAT form K.* Scarborough, ON, Canada: ITP Nelson.

Nelson Publishing. (2002). *Canadian tests of basic skills: CTBS form K, reliabilities of differences among tests, levels 5-8.* Scarborough, ON, Canada: ITP Nelson.

Nelson Publishing. (1997). *Canadian tests of basic skills.* Test Contents. Retrieved January 28, 2008, from Psychtest.com website: http://www.psychtest.com/curr01/CATLG033.HTM

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis. 21.* 127-142.

Oliver-Hoyo, M., & Allen, D. (2006). The use of triangulation methods in qualitative educational research. *Journal of College Science Teaching. 35*(4). 42-47.

Policies and procedures for ethics in human research. (n.d.). University of Saskatchewan. Retrieved on February 10, 2008 from University of Saskatchewan Research, Policies and Procedures website: http://www.usask.ca/research/ethics_review/policies.php

Rea, P. J., McLaughlin, V., & Walther-Thomas, C. (2002). Outcomes for students with learning disabilities in inclusive and pullout programs. *Exceptional Children, 68*(2), 230.

Reithaug, 2002. *Orchestrating success in reading*. West Vancouver, BC, Canada: Stirling Head Enterprises.

Rolland De Denus, J. (1995). *Inclusion: A description of the learning experiences of students with special needs in mainstream classrooms*. Unpublished master's thesis, University of Manitoba, Winnipeg, Manitoba, Canada.

Sackney, L. (2008). *Systemic reform for sustainability.* Retrieved June 28, 2008, from Government of Saskatchewan Web site: http://www.sasklearning.gov.sk.ca/branches/aar/ci/Systemic_Reform.pdf

Saskatchewan Ministry of Education. (2000). *Directions for diversity: Enhancing supports to children and Y\youth with diverse needs*. Final Report of the Saskatchewan Special Education Review Committee. Regina: SK. Retrieved May 24, 2008, from Government of Saskatchewan Web site: http://www.sasked.gov.sk.ca/admin/pub_pdf/committee.pdf

Saskatchewan Ministry of Education. (2008). *Minister's mandate.* Retrieved February 7, 2008, from Government of Saskatchewan Website: http://www.gov.sk.ca/adx/aspx/adxGetMedia.aspx?DocID=1580,1579,617,534,206,Documents&MediaID=2167&Filename=MandateLetter-MinKrawetz.pdf

Saskatchewan Ministry of Education. (2007). *Special education teacher qualifications: Course verification catalogue.* Retrieved June 2, 2008, from Government of Saskatchewan Web site: http://www.learning.gov.sk.ca/SpecEdTQ

Saskatchewan Ministry of Education. (2008). *Intensive supports.* Regina: SK. Retrieved October 22, 2008, from Government of Saskatchewan Web site: http://www.education.gov.sk.ca/IntensiveSupports

Saskatchewan Ministry of Education. (2008). *School division student services, service delivery model rubrics 2007-2008: Facilitating and monitoring effective practice.* Regina: SK. Retrieved October 21, 2008, from Government of Saskatchewan Web site: http://www.education.gov.sk.ca/adx/aspx/adxGetMedia.aspx?DocID=211,107,81,1,Documents&MediaID=4584&Filename=Student+Services+Service+Delivery+Model+Rubric+2007+2008-+English.pdf

Schulte, A. C., Osborne, S. T., & McKinney, J. D. (1990). Academic outcomes for students with learning disabilities in consultation and resource programs. *Exceptional Children, 57*, 162-172.

Seethaler, P. M., & Fuchs, L. S. (2005). A drop in the bucket: Randomized contact trials testing reading and math interventions. *Learning Disabilities Research and Practice*, *20*, 98–102.

Seymour, H. N., & Seymour, C. M. (1979). Ebonics and public law 94-142. *Journal of Black Studies 9*(4), 449-468.

SPSS Inc. (2006). *SPSS 14 for Windows*. Chicago, IL.

Stake, R. E. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage Publications. ISBN 0-7619-2665-8.

Success for All Foundation. (2008). *About SFAF: History.* Retrieved October 25, 2008, from the Success for All Foundation web site: http://www.successforall.net/about/about_history.htm

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics.* (5th ed.). Toronto, ON: Allyn and Bacon.

Tanguma, J., & Speed, F. M. (2000). *Interpreting the four types of sums of squares in SPSS*. Paper presented at the 28th Annual Meeting of the Mid-South Educational Research Association, Bowling Green, KY, November 15-17.

Townsend, T. (1993). *School effectiveness: Identifying the complexities*. Paper presented at the Annual Meeting of the International Congress for School Effectiveness and Improvement. Norrkoping, Sweden, January 1993.

Tse, J. (2006). Research on day treatment programs for preschoolers with disruptive behavior disorders. *Psychiatric Services*, *57*(4), 477-486.

Van der Linden, Wim J. (2005). *Linear models for optimum test design*. New York, NY: Springer. Retrieved on July 3, 2008 at http://www.springerlink.com/content/q45156727330541w/fulltext.pdf

Wang, M.C. (1984). *An investigation of the implementation and effects of a full-time mainstreaming program in a large urban school system.* New York City Board of Education, NY. National Inst. of Education (ED), Washington, DC.

Wang, M. C., & Birch. J. W. (1984). Comparison of a full-time mainstreaming program and a resource room approach. *Exceptional Children, 51*, 33-40.

Wang, M. C., Peverly, S., & Rudolph R. (1984). An investigation of the implementation and effects of a full-time mainstreaming program. *Remedial and Special Education, 5*, 21-32.

Wielkiewicz, R. M. (2000). *SPSS online guide.* Retrieved on February 9, 2008 from the McGraw Hill Publishing website: http://www.mhhe.com/socscience/psychology/runyon/spss/factorialanova.html

Wiener, J., & Tardiff, C. Y. (2004). Social and emotional functioning of children with learning disabilities: Does special education placement make a difference. *Learning Disabilities Research and Practice*, *19*, 20–32.

Willrodt, K., & Claybrook, S. (1995). *Effects of inclusion on academic outcomes.* Research Paper, Sam Houston State University, Huntsville, Texas.

Woloschuk, D. M. (2003). *The implementation of inclusive education in three Saskatchewan secondary schools.* Unpublished master's thesis, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Zigmond, N., Jenkins, J., Fuchs, L. S., Fafard, M., et al. (1995). Special education in restructured schools: Findings from three multi-year studies. *Phi Delta Kappan, 76*(7), 531. Retrieved October 31, 2007, from ProQuest Education Journals database. (Document ID: 1761416).

Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools, 45*(3), 206-216.

LIST OF APPENDICES

Appendix A

Permission to use Reliability of Differences Table

Hello Austin,

Permission is granted to use the CTBS Reliability of Differences Table in your thesis.

We would be very interested in receiving a copy of your thesis, for our files.

Thanks!

Rick

From: Norman, Rick (Nelson CAN)

Sent: Thursday, May 08, 2008 7:11 AM

To: 'Austin D.'

Subject: RE: Statistical background

Message received -- thanks Austin!
Please watch for reply within 24 hours...

Rick Norman
Manager of Assessment Resources and Services
Nelson Education Ltd.
1120 Birchmount Road
Scarborough, ON
Canada M1K 5G4

Nelson Web Site: www.nelson.com
On-line Catalogue: www.assess.nelson.com/catalogue.html
Customer Support: 1-800-268-2222

e-Mail: Rick.Norman@nelson.com
Voice Mail: 1-800-914-7776 ext 5547
Cell Phone (when in range): (905) 269-0344
Fax: (416) 752-9646

Appendix B

Permission to use Data

Date: December 3, 2007

To: Whom it may concern

From: Ron Ford, Director, Living Sky School Division

Re: Austin Degenhardt, research project

I have communicated recently with Austin regarding use of data from Living Sky School Division. Please consider this letter as permission for him to access the data pool within the school division for purposes of advancing his project. We have a common understanding as to the ethics involved.

I wish Austin all the best as he progresses with his project.

Yours truly,

Ron Ford

Appendix C

Ethics Application

UNIVERSITY OF SASKATCHEWAN

Application for Approval of Research Protocol

Behavioural Research Ethics Board (Beh-REB)

---

1. Name of Supervisors

a) Dr. Ivan Kelly, Thesis Supervisor
Department of Educational Psychology and Special Education

1a. Name of Student

a) Austin Degenhardt, Master of Education student
Department of Educational Psychology and Special Education

1b. Anticipated start and completion date of the study

Start: May 15th 2008
Anticipated end date of research: September 2008

2. Title of Study

A Case Study of Accountability for Special Education Delivery Services: A Mixed Model Analysis

3. Abstract

The purpose of this study will be to compare and critique two recognized methods of accountability measures and to identify potentially confounding issues or variables when studying the progress of special learners and program delivery. The researcher will address recommendations for measures of accountability as outlined in the Saskatchewan Special Education Review Committee report from January 2000. The report acknowledges the use of standardized testing as a measure of accountability but also encourages the development of other measures. The Saskatchewan Ministry of Education Service Delivery Model Rubrics: 2007-2008 regarding inclusionary philosophy and beliefs released on April 14, 2008 recommends a rubric-based assessment for program delivery.

The analysis will be conducted as a mixed-methods case study that includes 2 parts. The first part will assess the progress of a group of special learners through analysis of standardized test scores. A variety of measurement issues in relation to using academic achievement results in special learners as a measure of accountability will be addressed. A second assessment will be made to determine the degree of conformity that is perceived by a Superintendent of Special Education Services, a Principal, and a Special Education Teacher. Information will be gained through responsive evaluation and will address the categories defined in the Service Delivery Model Rubrics. Measurement issues in relation to the use of rubrics in performance assessment will be addressed.

4. Funding

There is no external source of funding used on this project.

5. Participants

The current study is conducting a secondary data analysis on data to be provided by the Living Sky School Division. The participants of this study are students enrolled in grades one to six that have participated in standardized testing sessions. All data will be **deidentified** before being received by the researcher. No direct contact has been, nor will be, made with the participants.

6. Consent

A signed letter of permission to use the scores and a statement of support for the research from Mr. Ron Ford, Director of Education for the Living Sky School Division, accompanies this application.

7. Methods/Procedures

A literature review will be completed in relation to methods that have been recognized as appropriate for assessing program effectiveness. Investigation and application of recognized methods will be conducted.

The analysis will be conducted as a mixed-methods case study that includes 2 parts. The first will explore issues related to the use of results of standardized testing for assessing accountability. The second will explore issues of rubric employment for assessing accountability. Statistical analysis of standardized scores and responsive evaluation will be employed to analyze received data.

No contact will be made with students whose scores will be used in the assessment as the present study employs secondary data analysis. All data will be **deidentified** and no attempts will be made to identify any record.

8. Storage of Data

All research data including consent forms, response forms, transcripts and tapes will be stored in a locked file cabinet in the Department of Educational Psychology and Special Education office at the University of Saskatchewan. All accumulated data from the study will be kept for five years upon study completion by Dr. Kelly at the University of Saskatchewan. All data will be destroyed after five years of storage.

9. Dissemination of Results

The results of this study will be disseminated in the form of a thesis.

10. Risk, Benefits, and Deception

The project presents no risk to participants as there is no direct interaction involved. The results of this study will be communicated to the Living Sky School Division.

11. Confidentiality

Any identifiable information regarding the participating schools and classes will be removed prior to obtaining the data from the school division. The focus of the present study is on measures of program accountability of programming within the school division. Therefore no individual results will be reported or needed. The participants of the study will not be contacted. Data will be **deidentified** prior to receipt by the researcher. The division has agreed to provide data containing only generic student numbers, standardized test scores, and a generic group identification score.

12. Data/Transcript Release

The dataset used in the present study will not be made public and will only be used for the purpose of the current study. Any subsequent data files created or modeling files created will be deleted as stipulated in section 8.

13. Debriefing and feedback

Results of the study will be made available to the Electronic Thesis Database with the University of Saskatchewan upon completion.

Appendix D

Ethics Exemption

UNIVERSITY OF SASKATCHEWAN

Research Ethic

Room 302
117 Scien
Saskatoon SK S7N 5C8
Telephone: (306) 9
Facsimile: (306) 9

To: Ivan Kelly – Educational Psychology and Special Education
Austin Degenhardt

Date: June 19, 2008

Re: A Case Study of Accountability for Special Education Delivery Services: A Mixed Model

---

The study entitled, "A Case Study of Accountability for Special Education Delivery Services: A Mixed Model" is exempt from the Research Ethics Board review process. This decision is based on the information provided in your application to the Ethics Office on May 27th, 2008.

Article 3.3 of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (1998) specifies that REB review and approval is not required to conduct a secondary analysis of data that cannot be linked to individuals, and for which there is no possibility that individuals can be identified in any published reports.

It should be noted that though your project is exempt of ethics review, your project should be conducted in an ethical manner (i.e. in accordance with the information that you submitted). It should also be noted that any deviation from the original methodology and/or research question should be brought to the attention of the Behavioural Research Ethics Board for further review. Please ensure that a full application is submitted to the Research Ethics Office prior to starting part II of this study.

Sincerely,

Dr. John Rigby, Chair
Behavioural Research Ethics Board
University of Saskatchewan

Appendix E

Participant Consent Agreement

PARTICIPANT CONSENT AGREEMENT

Title of Study:
A Case Study of Accountability for Special Education Delivery Services: A Mixed Model Analysis

Researcher and Supervisor:
Austin Degenhardt, Master of Education candidate in the Department of Educational Psychology and Special Education at the University of Saskatchewan.
E-mail: austindeg@sasktel.net
Home Telephone: 975-1952

Dr. Ivan Kelly, Professor, Department of Educational Psychology and Special Education, University of Saskatchewan.
E-mail: kelly52@shaw.ca
Office Telephone: 966-7715

Purpose of the Study:
You are invited to participate in a study, the purpose of which is to apply the newly mandated Saskatchewan Ministry of Education rubric. The purpose for applying the rubric is to assess the current level of special education service delivery in your school division. There are no known risks in this research study. The results would potentially be used for this research thesis, scientific publications, and presentations to teachers, parents, and professionals. Only aggregate data will be reported. Therefore, it will not be possible to identify any individual participants in any documents resulting from this research.

As a participant in this study:

1. You are provided with an invitational letter to participate in this study that provides project information, contact information, and research procedures.

2. You are asked to indicate consent by returning this letter via email with the accompanying completed rubric. Data will be kept confidential. Identifying information will be removed and replaced with code letters, so it is not possible to associate a name with any given set of responses. Arbitrary identification codes will be used that will not allow the identification of individual participants. Therefore, researchers will only have access to anonymous information.

3. Scores will serve as the data used in the statistical analyses on which the results and discussion of this study will be based. Data will be kept confidential. The researcher intends to begin data analysis by July 15, 2008.

4. Your data will be stored in a locked cabinet accessible only by the researchers' supervisor, and safeguarded for at least five years. Information identifying participants will be destroyed.
If you have any questions concerning the study, please feel free to contact the researcher at the number provided. The University of Saskatchewan Behavioural Research Ethics Board (Beh-REB) exempted this study from Research Ethics Board review on June 19, 2008. Participants interested in the results of the study will receive an executive summary upon request by contacting the researcher by phone or e-mail.

5. **You have the right to withdraw from this study at any time. If you choose to withdraw, the data you provided will be removed from analysis and destroyed. Withdrawal from this study will not result in any sort of penalty.**

I have read and understood the description above. I have been provided with contact information to have any questions addressed. By returning the completed rubric, I signify my consent to participate in the study, as stated. I understand that I may withdraw this consent at any time by notifying the researcher.

Appendix F

Instructions for Rubric Completion

INSTRUCTIONS FOR SURVEY COMPLETION

Please complete the following survey and return by **JULY 15, 2008**. There are 12 principles of service delivery to be evaluated. This is the actual rubric issued by Saskatchewan Education, April 14, 2008 for evaluation of Special Education Service Delivery across the province. Please be honest and open about your assessment. Your identity will be kept confidential.

Your category selection should reflect your opinion of where Special Education service delivery rates at the present time (Spring 2008) from your particular perspective.

- Please respond to all principles.
- If you are unsure, indicate what you perceive or believe is the appropriate category.
- Your perceptions are important.

Please scroll down to the survey and follow these directions.

1. Read the **principle** in the left column.
2. Read each of the descriptors to the right of each principle.
3. Indicate the category that you believe best describes the level of service for that particular principle.
    - If you have Microsoft "WORD", highlight the box and click on the highlighter tool
    - The highlighter tool can be activated by right-clicking on the gray area at the top of the document and checking off the "formatting" feature. Then just click the highlighter icon.
    - You could also right-click on the gray area at the top of the document and check off the "drawing" feature. Then just click the underlined letter A on the bottom bar to change the font color.
    - If you do not have these features, you could highlight the selected box and use "Control u" to underline your choice.
    - Your method is not important as long as your selection is clear.

4. Check the survey over to insure that EVERY principle has a category selected. The words "END OF SURVEY" indicate that you have addressed all principles.
5. I would not expect the survey to take more than 1 hour.
6. Return the survey by email to:
    - austindeg@sasktel.net

You may opt out of the survey process at any time.

Glossary and definitions

Inclusion
A particular philosophy of education with a set of related principles.
The principles are: inclusionary practice, differentiated instruction, parental involvement, assessment, team/collaboration, fostering independence, and assistive technology

Special learners
This term was chosen rather than "student with special needs" or other similar terminology in order to broaden the scope of intention. The term 'special learner' denotes students who may, indeed, have special needs historically known in the province as "designated" disabilities and now as students requiring intensive supports or with intensive needs (ie: Intensive Supports funding recognition). However, it also includes those students who are at risk, have mild disabilities or who have needs arising from environmental effects (ie: Diversity funding recognition).

Parents/caregivers
In recognition that children and adolescents may live within configurations of 'family' that differ from the traditional interpretation, the term 'caregivers' is added. This term affirms the role that others, whether foster parents, youth workers, grandparents or other individuals responsible for that child or adolescent, have in working with the school-based personnel.

Agency personnel
This term is used to denote personnel who have an interest in service provision to the child or adolescent who is a special learner but who are within a community-based or government- based agency not connected with the school division. This is intended to describe agencies with a mandate that includes health, corrections, social services, or supplementary supports related to the disability or at risk circumstances experienced by the child or youth.

Qualification guidelines given in the Saskatchewan Learning Special Education Teacher Qualifications Course Verification Catalogue (January, 2007).
Acceptable qualifications include a Master's Degree in Special Education from any university OR a Saskatchewan Professional A Teaching Certificate AND have successfully completed a minimum of eighteen credit hours of specified courses in special education with courses from each of the following areas: Speech and Language – three hours, Individual Assessment of Students with Exceptional Needs – three hours, Programming for Students with Exceptional Needs – three hours, and nine additional credit hours in approved special education courses.

Norm-referenced Testing
When an individual's test achievement is measured in comparison to the performance of all others that have written the same test providing a rank or placement in comparison to each other.

Difficulties in interpreting or responding to categories because of unclear definitions or meaning:
Questions:
Please contact austindeg@sasktel.net **or 306-975-1952**

THANK YOU for your time and participation on this survey.

School Division Student Services – Service Delivery Model Rubrics: 2007-2008
Inclusionary Philosophy and Beliefs

Principles	*Not Evident*	*Emerging/ Developing*	*Evident*	*Exemplary*
Inclusionary Practice	Special learners receive educational programming primarily in a special education classroom, a resource room or a therapy room; special education and regular education operate as separate enterprises	Special learners receive educational programming primarily in a special classroom, a resource room or therapy room, are in process of being transitioned into regular classroom setting with support and relevant programming	Special learners receive instruction in heterogeneous groups in the regular classroom and other school settings with support and relevant programming; specialized programming outside of the regular classroom is utilized when learner outcomes not achieved within classroom setting	Special learners supported in age-appropriate classrooms in neighborhood schools; sufficient supports exist within regular classroom for successful achievement of appropriate curricular outcomes for special learners; classroom teachers engage in instructional planning based on inclusionary practices and beliefs
Differentiated Instruction	Awareness of differentiated instructional practices is not evident; student achievement valued as acquisition of factual knowledge; assessment confined to summative information; special learners seen as outside of regular planning for classroom of students; special learners viewed as the responsibility of the special education teacher	Classroom teachers recognize the need for differentiated instruction; attempts at programming do not reflect consistency of planning; little enunciation of process by which to facilitate various levels of content for intended student products; special learners accommodated through parallel activities related to activities of regular class	Differentiated instruction is a component of school division and school strategic planning; adaptations made for student differences; modifications to content, process, and product; instructional strategies and curriculum linked; strategic planning for differentiated instruction crosses student commonalities and differences; evidence of efforts to effect successful curricular, interpersonal, and technology outcomes	School division/school plans to support differentiated instruction are comprehensive and successfully implemented: special learners are accepted unconditionally; special learners' work focuses on essential concepts and skills generalized across curricula and reflecting differing learning modalities, pacing and complexity; assessment, classroom management, and instruction interconnected; special learners' participate in work that supports their identity and are partners in collaborative learning; special learners learn from work aimed at their competencies; materials are used flexibly and there is adapted pacing
Parental Involvement	Parent/caregivers are not involved in selecting, developing, contributing to, and/or monitoring learner outcomes; parents/caregivers feel there has been no effort to communicate, listen or provide information	Parents/caregivers have familiarity with multidisciplinary team members; parent/caregiver input sought in monitoring learner outcomes; contact occurs less frequently than reporting periods; procedures that foster equal partnership with parents/caregivers and the school-based team are encouraged	Interest, willingness to invest time/energy in developing relationship with parents/caregivers evident; involvement of parents/ caregivers, parallel to the reporting periods, in contributing to, and/or monitoring special learners' outcomes; procedures that foster equal partnership with parents/caregivers and the school-based team are initiated	Regularly-scheduled involvement of parents/ caregivers, in addition to the reporting periods, in meetings focused on selecting, developing, contributing to, and/or monitoring learner outcomes; procedures that foster equal partnership with parents/caregivers and the school-based team are common accepted practice; shared responsibility for supporting learner outcomes
Assessment	Assessment focus on deficits; norm-referenced assessment instrument data preferred; assessment information is outdated; students referred for supports prior to assessment and/or development of individualized programming; progress related to activities rather than objectives	Diagnostic and summative assessments conducted; limited evidence of programming based on special learner strengths; limited informal assessment occurring to guide ongoing programming; annual goals developed but assessment information not directly linked; some norm-referenced assessments current	Majority of norm-referenced assessments are current; informal and formal assessment promoted; consistent effort made to make programming decisions based on special learner's demonstrated current level of performance; annual goals, assessment information and program planning directly linked	Informal and formal assessment is common practice and used to direct programming adaptations to content, process and products; assessment data from variety of sources used as basis for annual update of program planning

Inclusionary Philosophy and Beliefs

Principles	*Not Evident*	*Emerging/ Developing*	*Evident*	*Exemplary*
Team/ Collaboration	Supporting professionals provide services in isolation from regular classroom; each identifies, plans, monitors, reviews learner outcomes based on own discipline; students, classroom teachers supported through single or consecutive consultative response(s); protocols to connect interagency supports are not developed	Supporting professionals consult with teacher on ad hoc basis; no shared responsibility and accountability; some planning, monitoring and reviewing of learner outcomes occurring; need for collaborative interaction identified but no training; *reactive* multidisciplinary responses prevalent; protocols are in development to connect interagency supports	Collaboration occurs among multidisciplinary teams to identify, plan, monitor and review learner outcomes on scheduled basis; completion of initial inservice training on multidisciplinary collaboration; broader-based assessments; students, classroom teachers supported through *proactive* multidisciplinary responses; protocols are developed to connect interagency supports and implemented on ad hoc basis	Interdependence exists among multidisciplinary team members: mutual goal(s) and shared report writing; systematic approach to collection/analysis of diagnostic information; clear, effective decision making/planning process; interdisciplinary trust; clear role definitions and accountability parameters; supporting professionals involved with student in regular classroom; team members support complementary skills of each other; team members learn new skills and help one another to communicate accurately; conflicts are resolved constructively
Fostering Independence	Special learner has one-on-one support for entirety of school day; allocation of support focuses on responding to current level of dependency	Special learners provided with consistent one-on-one supports for significant portion of school day; multidisciplinary team, including classroom teacher, in process of assessing potential reduction in time allocation in developmentally appropriate sequence; transition planning in process	Program planning and staffing allocations address change in needs for independence for special learners; attention is paid to developmentally appropriate dependence/independence; PPP outlines transition steps to greater independence	Special learner accesses staffing allocation for physical or academic support only if required; regular monitoring of level of need occurs; appropriate skills are taught to special learner to facilitate independence in own decision making; student is able to navigate the educational environment with minimal individual or group support; positive interdependence with peers fostered to provide support for inclusive activities as needed; independent living needs considered, planned
Assistive Technology	Individual assistive technology not explored; knowledge of types or uses or advantages of assistive technology not known; no evidence of school division plan to enhance access	Limited individual assistive technologies considered; school division plan to enhance access to variety/applicable assistive technologies in development but limited pilot implementation; limited or no training provided to staff	Assessments by qualified personnel to determine appropriate assistive technology conducted for particular special learners; school division has developed comprehensive plan to enhance access to variety of assistive technologies for special learners; requests individually submitted – are congruent with school division plan; some level of training provided to school-based personnel	Individual requests for assistive technology are reviewed, fulfilled and updated on on-going basis; comprehensive plan to facilitate and enhance access to a range of assistive technologies to support educational outcomes for special learners is communicated and implemented; technical support is available to special learners and/or school personnel

Intervention Plan

Principles	*Not Evident*	*Emerging/ Developing*	*Evident*	*Exemplary*
Referral Process	A clear set of procedures is not established; pre-referral processes are not required, or not documented, or communicated by classroom teacher	A set of school division procedures has been established; inconsistent use and documentation of pre-referral processes; parents are notified of referral; results of formal assessments are communicated to parents/caregivers	Division-wide referral process established and followed; pre-referral processes practiced as part of referral process when uncertain of presence of needs; results formal assessments are communicated to parents/caregivers	School-based implementation guidelines established for division-wide referral process, including pre-referral; classroom teachers knowledgeable of, and alert to, students unresponsive to classroom interventions; parents are part of assessment process; assessment plans are developed, and communicated and monitored to parents/caregivers; all forms of assessment results communicated to parents/ caregivers; assessment plans for monitoring progress established
Personal Program Plans (PPP)	Development process for PPP not established at division/school levels; no procedures for review process in place; PPP developed by resource teacher	Development process of PPP and yearly review expectations established at division level; development process and review inconsistent in implementation; PPP developed by classroom teacher and resource teacher; PPP communicated to parents/caregivers; PPP reviewed by school-based personnel at end of school year	School-based team part of PPP development meeting with parents/caregivers; assessment information presented; programming needs outlined; services and supports are identified, included in PPP; team-identified priority annual outcomes are established and meet the standards identified by the Ministry of Education; review dates for PPP determined	PPP development meeting held among in-school team, supporting professionals, parents/ caregivers, outside agency personnel; team-identified priority annual outcomes are established and align with areas within the Impact Assessment Profile and with the standards identified by the Ministry of Education.
Transition Planning	No school division process or procedures outlined for transition planning to the next environment; transition considered a separate event	School division process and procedures for transition outlined; implementation at school level inconsistent; transition planning from grade to grade, between schools, into grade one from kindergarten, as well as post-school	School division process and procedures translated into guidelines at school level; consistent on-going process for transitioning into school, from grade to grade, between schools, as well as post-school; parents/caregivers involved in process; critical factors that must be in place identified.	School division process and procedures of transitions integrated as common practice at school level; indicators of successful transition developed and monitored.
Service Coordination	No common philosophy, language, perspective or focus; No evidence of interagency coordination	Agency groups providing services and programs documented; information shared between school and agencies; gap analysis of needs completed followed by identification of resources available; referrals to other complementary agencies made if necessary; separate procedures, policies, activities determined but are complementary	Agency and school-based personnel work together to complete a gap analysis and align resources effectively and efficiently; case management strategies evident	Protocols are mutually developed and implemented with shared leadership among agency and school-based personnel; joint planning for the identification and elimination of gaps occurs on a regular and pre-determined basis; shared case management strategies evident

Staffing Profile

Principles	*Not Evident*	*Emerging/ Developing*	*Evident*	*Exemplary*
Special Education Teacher Qualifications	Less than 50% of special educators have special education qualifications according to Ministry of Education requirements.	At least 50% of special educators have special education qualifications according to Ministry of Education requirements.	At least 75% of special educators have special education qualifications according to Ministry of Education requirements	All school division special educators have special education qualifications according to Ministry of Education requirements

END OF SURVEY

Please complete the survey and return by JULY 15, 2008.

VITA

Austin Degenhardt was born in Goodsoil, Saskatchewan, Canada. His hometown is located in Saskatoon, Saskatchewan. He received his Bachelor of Arts Honours degree specializing in Psychology from the University of Saskatchewan in 1987, his Bachelor of Education degree with distinction from the University of Saskatchewan in 1993, and is currently a Master of Education candidate at the University of Saskatchewan.