

**(Re)-Examining the Influence of Program Placement on the
Academic Achievement of Students with Learning Disabilities**

Steven McKibbin

Thesis submitted to the University of Ottawa
in partial fulfilment of the requirements for the

Doctor of Philosophy,

Education

Faculty of Education

University of Ottawa

© Steven McKibbin, Ottawa, Canada, 2020

Abstract

This study explored the relationship between several variables known to influence achievement in Canadian Grade 6 students with Learning Disabilities (LD) who received instruction in either a regular class or specialized program placement. The main independent variable was program placement while the influence of four other independent variables was explored (i.e., level of academic need; prior achievement; socioeconomic status and sex). The dependent variable was a standardized, large-scale assessment of achievement. Hierarchical multiple regression was conducted on a secondary data file in order to address the following research questions: i) Does placement in a regular or specialized program influence the educational outcomes for Grade 6 students with LD, after controlling for the influence of prior achievement in Grade 3?; ii) Is there a relationship between the sociodemographic variables of sex and/or socioeconomic status and achievement for students with LD placed in either a regular or specialized program?; and iii) What influence does the student's level of academic need have on achievement, beyond program placement, and after controlling for the influence of the other variables in the model? Results revealed that the variables of program placement and prior achievement were significant predictors of scholastic success only when the level of academic need variable was not taken into account. When the follow-up analysis focused on a relatively matched group of students with similar academic need, none of the predictors in the regression model significantly influenced achievement -- including program placement. These results provide important insight into the nuanced relationship of the ecological variables known to affect learning in students with LD placed in regular or specialized programs for instruction. Implications are discussed for stakeholders in Ontario's public education system in terms of the optimum service delivery model for students with LD, and the inclusive education debate in Canada and abroad.

Dedication

This dissertation is dedicated to the memory of my late brother, Christopher Lorne McKibbin,
who in his shortened journey of life exemplified the best of us.

Acknowledgments

Above all, Dr. Jessica Whitley's sagacious guidance and unwavering encouragement as my advisor were instrumental in navigating this arduous, but sublime, scholarly journey. Without Dr. Whitley's thoughtfulness, generosity and compassion, this dissertation would not have reached completion (remember, Jess, the walls of hell must be papered with SPSS logos!). My deepest and most genuine gratitude also extends to my committee members: Dr. Maria Rogers, Dr. David Trumpower, and Professor Emeritus, Dr. J. Bradley Cousins, whose fine scholarly example were a paragon of inspiration for me both as a student, and in my professional life. I am eternally obliged for their collective knowledge, technical expertise, and intellectual mentorship. Dr. Martin Barlosky also merits enduring thanks for encouraging me to pursue a doctoral program, and my humblest appreciation goes to Dr. Gina Harrison for acting as my external evaluator. To my colleagues, Drs. June Pimm, Tim Hogan, Lisa Piers, Tsala Mosimakoko, Andrea Rahim, Lauren Figueredo, Marilyn Kasian and Petra Duschner -- as well as Michèle Giroux -- how can I ever repay the inestimable support you have all graciously provided at various stages of this project? I shall be forever indebted to you all. Sincerest thanks to Carol Ann Kelly as well, for her kindly direction throughout this administrative odyssey, and to the OSSTF for their financial awards and professional support at critical junctures (thank you Bonnie Cheung!). My trusted friend, Kevin DeGaust, deserves distinctive mention for always being there for me at the biggest moments in my life, including this one -- as copy editor! Final tribute goes to my wife, Kelly, and two wonderful daughters, Alyssa and Hillary, for their love, support and sacrifice amidst the most difficult of circumstances. Together with my dear friends, J.B., Chris, and Jonesy; as well as my mom, Little Steven, Estelle -- and the rest of my immediate and extended family -- you have all motivated me to persevere throughout my academic formation, and shared with me the little victories of this momentous personal accomplishment. Thank you all for believing with me along the way!

Table of Contents

Chapter One: Introduction	1
Background of the Problem	1
Research Design.....	4
Clarification of Key Terms	6
Study Structure.....	15
Chapter Two: Literature Review	17
Defining Learning Disabilities.....	17
Inclusive Education.....	24
Program Placement and Academic Achievement.....	29
Theoretical Framework.....	42
Chapter Three: Methods.....	49
Demographic Context of Sample.....	50
Instruments and Procedures	51
Screening the Secondary Data File	63
Missing Data	69
Modifying Variables for Multivariate Analysis.....	80
Chapter Four: Results	93
Sample Characteristics.....	93
Assumptions of the Hierarchical Multiple Regression Analyses	98
Not Accounting for Level of Academic Need	100
Adding Level of Academic Need to the Model	103
Chapter Five: Discussion.....	107

Main Findings	110
Implications.....	128
Limitations/Delimitations	133
Directions for Future Research	143
Conclusion	147
References	150
Appendix A: Normality Tests of the Original Data by Variable and Program	167
Appendix B: Histograms of the Original Outcome Data by Subject Area and Program	168
Appendix C: Histograms of the Composite Outcome Variables by Grade and Program.....	177
Appendix D: Scatter and P-P plots for Both Regression Models	180

List of Tables

Table 1. Description of Original Variables in the Secondary Data File	64
Table 2. Means, Standard Deviations and Correlations of the Original Outcome Data	65
Table 3. Missing Values Analysis of the Original Data by Program.....	71
Table 4. Analytical Procedure.....	87
Table 5. Descriptive Statistics by Program and Sex	93
Table 6. Socioeconomic Status Descriptive Statistics by Program	94
Table 7. Independent Samples t-test of LAN by Program.....	95
Table 8. Means, Standard Deviations and Normality Tests of the Composite Variables.....	96
Table 9. Descriptive Statistics and Correlations from the 1st Regression Model	101
Table 10. Regression Model of Grade 6 Achievement Not Accounting for Academic Need....	101
Table 11. Descriptive Statistics and Correlations from the 2nd Regression Model	104
Table 12. Regression Model of Grade 6 Achievement Accounting for Academic Need.....	105

List of Figures

Figure 1. Conceptual Model Examining the Effect of Program Placement on Achievement	5
Figure 2. Bioecological Model of Human Development.....	43
Figure 3. Level of Academic Need (LAN) Rating Scale.....	56
Figure 4. Socioeconomic Status (SES) Index.....	60
Figure 5. Distributions of the Composite Outcome Variables.....	97
Figure 6. Scatterplot and P-P plot of the SUPER6 Variable in the 1st Regression Model.....	180
Figure 7. Scatterplot and P-P plot of the SUPER6 Variable in the 2nd Regression Model	181

Chapter One: Introduction

Background of the Problem

According to D'Intino (2017) the prevalence of Learning Disabilities (LD) in the Canadian school-aged population is estimated conservatively at 3.2%. The most recent data available from the government of Ontario indicates that over 72,000 students were formally identified with the LD exceptionality in 2014-2015 (Ontario Ministry of Education, 2016). This is the highest reported incidence rate for any exceptionality, representing approximately 41% of all formally identified exceptional pupils across the province of Ontario. Notably, many more students will receive special education programs and services without being formally identified with the LD exceptionality by the school system (Bennett, 2009; Ontario Ministry of Education, 2016). Thus, the actual incidence rate for students with LD is likely much higher.

Given the prevalence of LD in school-aged children, educators in the K-12 public education system have long sought to develop special education service delivery models which can address the complex academic, social and emotional needs of these students. Typically, students with LD will receive educational supports, services and resources beyond the regular classroom setting as a function of their LD identification, such as eligibility for a special class, resource support from a special education professional or assistive technology in order to help access the curriculum. Indeed, several different service-delivery models exist, ranging from full inclusion in the regular class, to smaller class sizes supported by special education professionals – and several other permutations and combinations in between.

In recent years, the long-term outcomes for Canadian students with LD have improved, yet these students continue to lag-behind academically and socially when compared to their same-aged peers (Whitley, Lupart & Beran, 2007). For example, students with LD often struggle

with basic reading, writing and math skills, and can present difficulties in other learning skills which affect their achievement, such as asking for help from a teacher. These variables are either directly or indirectly related to the learning disability itself. Thus, understanding the factors that influence achievement for students with LD is important if improvements are to be made to the education system serving these students.

Finally, the word ‘re-examination’ in the title of this study is purposefully presented in parentheses because many authors in the past have reported on the influence of program placement on achievement for students with LD (Andrews & Lupart, 2000; Baker, Wang & Walberg, 1995; Bennett, 2009; Calberg & Kavale, 1980; Cole, Waldron, & Majd, 2004; Ghandi, 2007; Marston, 1996). However, as is demonstrated in the literature review below, there is a paucity of large scale, methodologically rigorous, and recent Canadian data related to this problem. Therefore, re-examining the relationship between program placement and achievement for students with LD remains timely and relevant for contributing to the ongoing academic debate regarding inclusive education in Canada.

Purpose of Research

The purpose of this research was to re-examine the impact of program placement on achievement for students with LD, while controlling for the influence of several other ecological factors that have been identified in the research literature as significant predictors of achievement for this group of students. In addressing the influence of these multiple ecological factors, the academic trajectory of students with LD placed in either a regular or specialized program was followed longitudinally as they moved through elementary school from Grade 3 to Grade 6. Results from this study add to the research base by examining achievement within the same students over time using a repeated measures design, whereby a student acted as their own

control for variance in test scores (Kauffman, Mock, Tankersley, & Landrum, 2008; Rogers, 2006b). Many researchers have also established the relationship between the student's sex and/or socioeconomic status (SES) and achievement for students with LD (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Flannery, Liederman, Daly, & Schultz, 2000; Morgan, Farkas & Wu, 2009; Wei, Lenz, & Blackorby, 2013; Vogel, 1990; Vandenberg & Emery, 2009; Wheldall & Limbrick, 2010). Therefore, the influence of these particular sociodemographic variables was also examined in order to shed light on any differences they may exert on this population across program placement conditions. The use of a standardized, criterion-referenced assessment for the study's outcome measure also lends credence to the research design.

Finally, this research sought to control for the influence of another main variable of interest called, level of academic need (LAN) -- which represents a proxy for LD severity in the current study. The LAN variable has been identified by several researchers as being salient in influencing achievement for students with LD but has proven difficult to conceptualize or measure across program placement conditions (Baker et al., 1995; Bennett, 2009; McLeskey & Waldron, 2011; Zigmond, 2003). To address the challenge of measuring the LAN variable, a 'natural' comparison group was identified in students who were placed on a waitlist for a specialized LD program, but who received their instruction in a regular classroom nonetheless. Achievement data from this waitlist group of students was compared to students placed in the specialized LD program, as well as to another group of students placed in the regular program, in order to highlight the nuanced role of the LAN variable in predicting scholastic success. This analytical approach represents a notable improvement in research methodology by effectively matching the high academic need students with a comparison group when assessing the influence

of program placement on achievement. In sum, the study's purpose was to develop a more methodologically sound research design for evaluating the inclusion model for students with LD.

Research Design

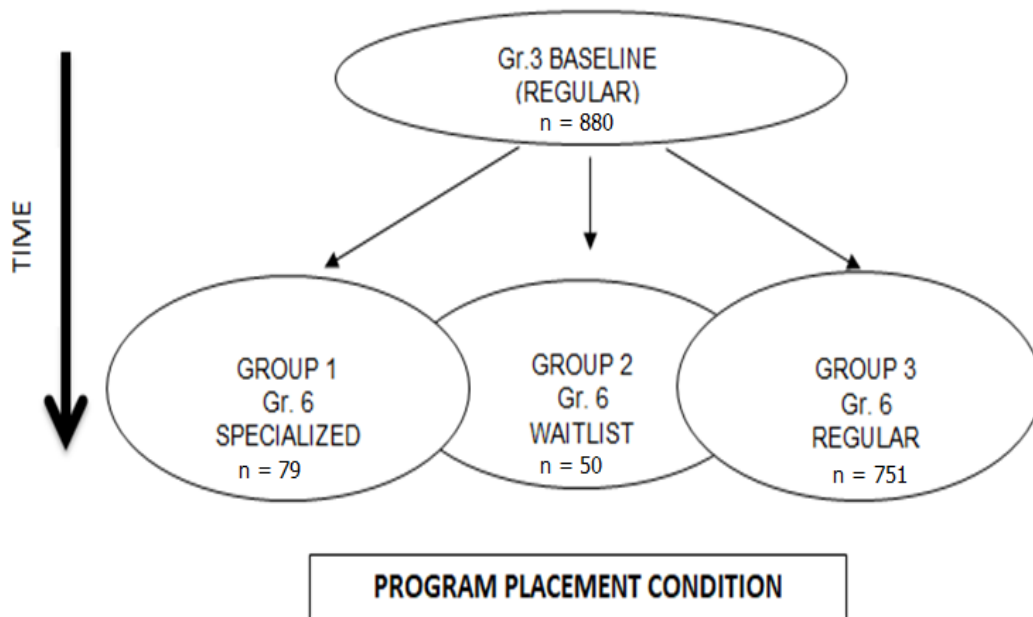
Secondary data. This study was conducted using a secondary data file that was originally compiled in 2014 as part of a review of programs and services for students with LD within a relatively large school district in Ontario, Canada. For the sake of full disclosure it should be noted that I led this review in my capacity as a district-level program evaluator. Access to the secondary data file for purposes of this study was granted through a data sharing agreement with the school district under study. Educational researchers in Ontario have conducted studies based on secondary data for decades (Cousins, Ross & Prentice 1993; Ebsensen, Piper, Vogel, West, Arboit, & Harris, 1996; Robinson & Cousins 2004). Ethics approval for the current study was granted by the University of Ottawa in accordance with ethical standards set forth by the *Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans* (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada, 2018). All personal identifying information contained in the secondary data file was removed prior to analysis. Therefore, the anonymity of the students contained in the secondary data file, as well as the name of the school district under study, was guaranteed.

Conceptual model. This study employed a quasi-experimental, longitudinal correlational research design exploring the degree of relationship between several independent variables known to influence academic achievement in elementary students with LD (Creswell, 2009a). The academic trajectory of three groups of students was followed from Grades 3 to 6 in order to measure the influence of program placement on achievement. Students from all three groups

received their instruction in a regular classroom program in Grade 3. However, three years later one smaller group of these students had subsequently received their instruction in a specialized LD program. A second, smaller group of these students was placed on a waitlist for a specialized LD program, but received their instruction in the regular program nonetheless. This waitlist group served as the quasi comparison group for the study and was used to control for selection factors known to influence the composition of specialized LD programs, such as LD severity. Finally, a third much larger group of students with LD was placed in a regular program for instruction. A full description of how these three groups of students with LD were categorized into different program placements is presented in the methods chapter.

Figure 1

Conceptual Model Examining the Effect of Program Placement on Achievement



The following three research questions were explored in this study:

Research question #1. Does placement in either a regular or specialized program influence the educational outcomes for Grade 6 students with LD after controlling for the influence of prior achievement in Grade 3?;

Research question #2. Is there a relationship between the sociodemographic variables of student sex and/or socioeconomic status and the educational outcomes for these students placed in either the regular or specialized program?; and

Research question #3. What influence does the student's level of academic need have on their achievement beyond program placement, after controlling for the influence of the other ecological variables in the model?

In addressing these research questions, the study sought to provide a better understanding of the impact of program placement on achievement for students with LD in terms of its nuanced relationship with other ecological factors. The information gained from this research study benefits the K-12 education sector in Ontario in terms of providing valuable evaluative feedback on the service delivery model for students with LD. The use of secondary data analysis techniques on a time-series sample with a matched comparison group also presents a novel methodological advantage in terms of re-examining the impact of inclusive educational practices on students with LD in Canada. Finally, the use of large-scale, standardized provincial assessment data also provides more relevant and recent data to inform the Canadian special education context.

Clarification of Key Terms

Learning Disabilities. The definition of LD has a long, contentious history in the academic research literature (D'Intino, 2017; Kavale, 2005; Simpson, Rose & Bakken, 2013).

Given that the present study is based on participants from an Ontario school district, and that the outcome measure was provincial assessment scores, the Ontario Ministry of Education's definition of LD was used as an operational definition. The provincial government's definition of LD is clearly stated in *Policy and Program Memorandum No. 8* (Ontario Ministry of Education, 2014) as:

...one of a number of neuro-developmental disorders that persistently and significantly has an impact on the ability to learn and use academic and other skills and that:

- affects the ability to perceive or process verbal or non-verbal information in an effective and accurate manner in students who have assessed intellectual abilities that are at least in the average range;
- results in (a) academic underachievement that is inconsistent with the intellectual abilities of the student (which are at least in the average range) and/or (b) academic achievement that can be maintained by the student only with extremely high levels of effort and/or with additional support;
- results in difficulties in the development and use of skills in one or more of the following areas: reading, writing, mathematics, and work habits and learning skills;
- may typically be associated with difficulties in one or more cognitive processes, such as phonological processing; memory and attention; processing speed; perceptual-motor processing; visual-spatial processing; executive functions (e.g., self-regulation of behaviour and emotions, planning, organizing of thoughts and activities, prioritizing, decision making);

- may be associated with difficulties in social interaction (e.g., difficulty in understanding social norms or the point of view of others); with various other conditions or disorders, diagnosed or undiagnosed; or with other exceptionalities; and
- is not the result of a lack of acuity in hearing and/or vision that has not been corrected; intellectual disabilities; socio-economic factors; cultural differences; lack of proficiency in the language of instruction; lack of motivation or effort; gaps in school attendance or inadequate opportunity to benefit from instruction.

This definition of LD brings to light one of the main issues in the academic debate concerning the ‘discrepancy model’, which refers to the measurement gap between standardized tests of intelligence (IQ) and academic performance. Despite the fact that this issue has been noted by several authors (D’Intino, 2017; Fletcher, Francis, Shaywitz, Foorman, Stuebing, & Shaywitz, 1998; Klassen 2002; Kozey & Siegel, 2008; Vellutino et al., 2000), the Ontario Ministry of Education’s definition of LD clearly contains the main elements of the discrepancy model. A consensus definition of LD is further complicated by the 2013 revision of the American Psychiatric Association’s, *Diagnostic and Statistical Manual of Mental Disorders, 5* (DSM-5), and the ‘clinical fall-out’ that emerged from the more broad-based diagnostic criteria provided by the new definition of Specific Learning Disabilities. This matter is discussed in more detail in the literature review chapter. Suffice it to say here that other researchers contend that the definition of LD should be grounded in developmental theories (Waber, 2010). Thus, it should be acknowledged that the group of students identified with LD using the approach described by the Ontario provincial policy may be different from those students who might be identified using a more developmental approach.

Level of academic need. An important aspect of the current study focusing on inclusive education was that students with LD were grouped on the basis of their LAN before comparing achievement by placement type in either the regular or specialized program. This methodological issue has been long identified as requiring further research by many authors in the field (Baker et al., 1995; Ghandi, 2007; Marston, 1987; McLeskey & Waldron, 2011; Zigmond, 2003). Most authors argue for a better construct – one that considers the severity of the LD in academic terms – to be used in research designs comparing program placement models. Thus, the LAN variable was conceived as a proxy for LD severity in the current study and operationally defined in terms of a composite of seven different psychometric, academic and ecological factors, including:

- i) highest IQ score, as measured by the Wechsler Intelligence Scale for Children – IV (WISC-IV);
- ii) extent of cognitive processing deficit using normative data from WISC-IV;
- iii) difference between academic performance on Wechsler Individual Achievement Test - III and the WISC-IV (i.e., ability-achievement discrepancy);
- iv) extent of the academic delay measured using normative data from the Wechsler Individual Achievement Test – III (WIAT-III);
- v) evidence (or diagnosis) of an additional disorder;
- vi) extent to which the academic delay is attributable to other ecological factors (e.g., absenteeism; cultural factors; and/or family issues); and
- vii) evidence of an emotional/behavioural/social problems related to the LD

Notably, only students in the specialized and waitlist groups were assigned a LAN score as these scores were calculated strictly for those students who applied to the specialized program at the school district under study. For this reason, students in the regular group were not assigned

a LAN score. More details about the LAN instrument and the procedures used to develop the LAN score are provided in the methods chapter.

Socioeconomic status. Many researchers have examined the influence of SES on academic achievement for students with LD (Morgan et al., 2009; Sirin, 2005; Vandenberg & Emery, 2009; Wei et al., 2013). However, the SES variable has not been examined to date in conjunction with the program placement variable for its influence on achievement in this vulnerable group of students. Sirin (2005) provides a useful definition of SES for understanding how the construct has been employed for the current study: “SES describes an individual’s or a family’s ranking on a hierarchy according to access to or control over some combination of valued commodities such as wealth, power, and social status” (p. 418). The author concludes that family SES, both at the student level and school levels, are strong correlates of academic performance in the general student population.

The index used to assign SES scores to individual students for this research was originally developed as a school-level measure of socioeconomic need within the district under study. These school-level scores were used as a proxy measure of SES for the families of individual students with LD who attended the school. This operationally consisted of assigning a standardized z score to all students in a specific school based on their neighborhood schools’ SES score. The SES variable itself was defined as a composite of twelve different indicators, organized into the following five factors: i) Poverty; ii) Family/Community; iii) Mobility; iv) Cultural/Linguistic; and v) Readiness to Learn. The methods chapter provides complete details about the SES instrument and the procedures used to assign SES scores to individual students in the secondary data file.

Inclusive education. This study adopted the definition of inclusive education proposed by Winzer (2008 p.g., 43), “as a system of equity for students with exceptionalities that expresses a commitment to educate each child to the maximum extent through placement, instruction and support in the most heterogeneous and appropriate environment”. An important obscurity must be noted here in adopting a definition of inclusive education that purports a system of equity for programs and services that students with LD receive in various program placements. Essentially, there is no effective way to determine whether or not this criterion was met. The difficulty in measuring or even operationalizing inclusive educational settings is an issue that has been reported by many researchers and is further elaborated in the literature review chapter. The Ontario Ministry of Education operationalizes inclusive education in terms of a continuum of services model, ranging from program placement in the regular classroom to a specialized class placement. For this reason, the term ‘regular program placement’ was used as a proxy for inclusive education for the purposes of the current study.

Regular program placement. One of the most challenging methodological issues with studies examining the influence of program placement and achievement for students with LD is the lack of sufficient detail provided about the instruction model used in the research design (McLeskey & Waldron, 2011). In this study, placement in the regular program refers to students with LD who received instruction for all subjects, for a significant portion of the instructional day in a regular classroom with same-age peers. In some cases, these students received additional support ranging from monitoring and withdrawal services provided by a school-based special education teacher or educational assistant, to occasional classroom support in the form of an itinerant special education teacher. The level of support each student with LD in the regular program received depended on the individual student’s educational needs and, of course, the

availability of support within the local school context. As is typical for most students with LD placed in the regular classroom, monitoring and withdrawal support for these students in the district under study was in the areas of language and math, and was focused on addressing learning goals or expectations contained in the student's Individual Education Plan (Bennett, 2009). The actual length of time students with LD spent out of the regular class for withdrawal support was unknown.

Specialized program placement. In the current study a specialized program placement referred to students with LD who were placed into a special education class where the student-teacher ratio was lower than in a regular class (i.e., 8 students and 1 special education teacher), and the amount of instruction delivered in the specialized class ranged from at least 50 per cent of the school day to the entire school day. Students with LD who received a specialized program may have been included in a regular class for participation in: i) subjects in areas of a given student's strength; ii) non-academic curriculum areas such as physical education, art, music, and drama; and iii) other school-based activities. Again, the actual proportion of the instructional day that individual students with LD received support in the regular class was not specified for the current study. Nevertheless, the focus of the specialized program was to develop academic, communication, social, organizational and self-advocacy skills. As such, intensive instructional support was provided for language and math within the specialized program placement. These students may have also received support from district-level specialized personnel (e.g., teacher consultants; psychologists; speech-language pathologists) specifically trained in intervention strategies that are beneficial for the LD exceptionality.

In sum, the construct of a specialized program placement is understood for the purposes of the current study in relation to LAN. As will be demonstrated later, offers for a specialized LD

class placement were determined following a school-level application process involving information provided by educators. A pool of applications were then pre-sorted by a district level committee using a rank order procedure before any formal placement recommendations were made with respect to any individual student. This critical procedural information is important to understand so that the nuanced differences between the LAN of students placed in the waitlist or specialized program conditions could be highlighted. More details on the program placement procedures used for the study is provided in the Methods chapter.

Pedagogical support for students with LD. In accordance with Ontario Ministry of Education direction, the district under study offers programs and services for students with LD through the tiered approach to intervention. This particular service-delivery approach is akin to the Multi-Tier System of Support (MTSS), and is highlighted in *Learning for All* (2013), as an evidence-based approach designed to meet the needs of students with special education needs through increasing the intensity of instructional supports for a particular student in response to their rate of learning. According to *Learning for All* (2013), the tiered approach complements other pedagogical strategies for students with special education needs such as Universal Design for Learning (UDL), Differentiated Instruction (DI) and progress monitoring through Response to Intervention (RTI). In this way, the tiered approach to intervention forms part of an overall pedagogical strategy designed to support all students identified with special education needs, including students with LD.

Notably, the tiered intervention approach operates independently of program placement. Instructional supports at the Tier 1 and Tier 2 levels specifically designed for students with LD are typically delivered to groups of students in the regular classroom (e.g., small group reading activities), whereas Tier 3 interventions are normally targeted at individual students with LD

through more intensive personalized instruction, such as individualized academic support from a special education teacher. However, Tier 3 interventions can also occur within either a regular classroom or specialized program placement. An example of this type of targeted literacy intervention which takes place either within a regular class or specialized program placement is the Empower Reading Program.

The Empower Reading program was employed on a large scale in the district under study in order to remediate the large number of students in the primary grades who were experiencing reading difficulties in both inclusive and non-inclusive settings. This remedial program was the main literacy intervention operating within the district's specialized LD programs during the time period when the data for the current study was collected. Additionally, the focus of the specialized program was to develop academic, communication, social, organizational and self-advocacy skills in students with LD. As such, intensive instructional support was provided for language and math within the specialized program placement. This group of students would have also received support from district-level specialized personnel such as, teacher consultants; psychologists; speech-language pathologists who are specifically trained in pedagogical intervention strategies that are beneficial for the LD exceptionality.

In sum, students with LD are supported in the regular classroom through the pedagogical strategies of UDL and DI that are encapsulated in the province of Ontario through the *Learning for All* (2013) quasi-policy document. The tiered intervention approach ensures that more targeted and precise instructional and assessment strategies are employed for students with LD in both inclusive and non-inclusive educational settings, such as the Empower Reading program. In this way, the tiered intervention approach is comparable to contemporary MTSS and RTI practices.

Study Structure

This thesis is organized into four remaining chapters. The literature review chapter situates the study within the wider academic debate on inclusive education by addressing the ‘methods gap’ in the research base, and in so doing, lays the groundwork for a more methodologically sound research design for understanding several ecological factors influencing achievement in students with LD in inclusive and specialized educational settings. Next, the methods chapter describes the demographic context of the original secondary data file upon which the study is based, and presents the instruments and procedures that were used to prepare the secondary data file for the modifications required in order to carry out the analytical plan. A complete missing values analysis is also provided along with a description of the subsequent multiple imputation procedures used for handling missing data. Finally, the study’s two-step analytical plan is presented along with the justification for conducting hierarchical multiple regression data analysis techniques.

The results chapter presents the final sample characteristics and discusses how it meets the assumptions for the subsequent, two-step hierarchical multiple regression analyses. Results from both steps in the analysis are also presented. In the first analytical step, the influence of the LAN variable was not taken into account in the prediction model; whereas in the follow-up analysis, the influence of this variable was properly considered in the regression model. The final chapter includes a discussion of the main findings from both steps of the analysis and their implications within the context of the study’s theoretical framework and three research questions. Limitations and delimitations of the study are also presented along with several methodological considerations for future research examining the nuanced influence of program

placement and achievement for students with LD. Finally, concluding remarks are provided based on the study's findings.

Chapter Two: Literature Review

The following review of the academic literature highlights the difficulties in defining the construct of LD, and describes the typical learning problems which students with LD face in school. Next, a summary is provided regarding the scholarly research on the philosophical, legislative, and policy underpinnings of inclusive education as it relates to program placement for students with special education needs in general. The literature specifically concerning academic outcomes for students with LD in both regular and inclusive educational settings is also discussed, paying particular attention to the methodological limitations inherent in the existing research base. Similarly, the relationship between the sociodemographic variables of student sex and SES for students with LD is also presented. The literature review concludes with a description of the theoretical framework used for the study (Bronfenbrenner, 1979), and suggests how the present study addresses the methodological issues that were identified in the academic literature review.

The purpose of this review of literature is to situate the current study in the wider academic debate on inclusion in Canada. In so doing, it addresses an important gap in the research base related to the methods and variables typically considered in measuring the influence of program placement on achievement for students with LD. Thus, the review forms the basis for a more methodologically sound research design for understanding the ecological factors influencing achievement in this population of students.

Defining Learning Disabilities

As noted earlier, one of the major points of contention surrounding the definition of LD centres on the ‘discrepancy model’, which posits that when a significant discrepancy (or gap) exists between a students’ cognitive ability and their academic ability, that student can be

considered to have a learning disability (Fletcher et al., 1998; Kozey & Siegel, 2008). Typically, this discrepancy is assessed as the difference in performance on a norm-referenced intelligence test and a standardized measure of academic achievement. Using this definition, the discrepancy between these two particular constructs would provide the evidentiary basis for diagnosing a student with LD. The discrepancy model is practically ubiquitous in the field of special education, as evidenced by the official definition of LD adopted by the Learning Disabilities Association of Canada (2019):

...Learning Disabilities refer to a number of disorders which may affect the acquisition, organization, retention, understanding or use of verbal or nonverbal information. These disorders affect learning in individuals who otherwise demonstrate at least average abilities essential for thinking and/or reasoning. As such, learning disabilities are distinct from global intellectual deficiency.

Several authors have noted that there is no consensus amongst researchers or practitioners with respect to the operational definition of LD, and that there is no theoretical support for a definition of LD on the basis of the discrepancy between IQ and achievement alone (D'Intino, 2017; Klassen 2002; Kozey & Siegel, 2008). The main issues concerning the use of IQ in the operational definition of LD are: i) its weakness in reliably distinguishing between disabled and non-disabled students; and ii) its practical irrelevance in the remediation process for students with learning difficulties. The contention here is that the discrepancy model cannot adequately serve to identify students with LD.

Critics of the discrepancy model note that its use makes early identification and intervention of children with suspected learning difficulties challenging, asserting that it is a 'wait-to-fail' approach (Fletcher et al., 1998; Kozey & Siegel, 2008). These researchers note that

children who struggle in the primary grades do not necessarily present with the IQ-achievement discrepancy necessary to meet the eligibility requirements under the model. Another major criticism of the discrepancy model is the inconsistent manner in which both school-level clinicians and educators apply this approach in their assessment and diagnostic practices. Adding to the shortcomings of the discrepancy model is that the diagnostic cut-scores can ‘miss’ students with below average intellectual ability (i.e., slow learner); therefore, these students do not qualify to receive special education services. By relying on the discrepancy model for identification purposes, children with lower average cognitive capacity risk falling through the service cracks. Finally, there is little empirical evidence in support of either the reliability or validity of the discrepancy model to identify students with LD.

Vellutino, Scanlon, and Lyon (2000) further articulate that the discrepancy model cannot effectively distinguish between children with LD and those students whose academic delay is attributable to other variables, more ecological in nature, such as cultural or linguistic capital, absences from school, or even poor instruction. In this sense, any definition of LD which vaults the gap between IQ and academics as the ideal measure, proves too narrow, and misses a host of other important ecological variables. Given this ambiguity in defining the LD construct, Klassen (2002) argues that other ecological approaches may be more suitable for definitional purposes. An ecological perspective of LD would define the construct in terms of learning processes or other contextual factors, rather than relying solely on the discrepancy between IQ and performance as the evidentiary basis. For instance, standardized language decoding measures such as a reading comprehension assessment would be used instead of IQ tests to identify the LD. An ecological definition of LD, might also consider the impact of factors external to the student, such as SES.

Proponents of the ecological perspective argue that child-centric definitions of LD such as the discrepancy model limit the ability to acknowledge the influence of other important systemic factors operating at the school or family levels. For example, sociodemographic variables such as SES or sex have been shown to predict the incidence rate for students with LD (Vandenberg & Emery, 2009; Wei, et al., 2013; Whitley, et al., 2007) and, thus, should be considered in an operational definition of LD. Similarly, Florian and McLaughlin (2008) examined the classification of disabilities in both the UK and the United States, and claim that definitions for any exceptionality, including LD, must take into account other environmental factors affecting the individual student. In other words, the context counts too. Supporting this position, Waber (2010) favours a more developmental approach to defining LD, where it is described as the interaction between the student and their environment, rather than in terms of individual skill deficits within the student. These types of ecological, systems or developmental approaches to a definition of LD gain credibility when applied to the educational context referred to as inclusion in Ontario's public school system.

Many provincial governments in Canada continue to ascribe to the discrepancy model's central role in their policy definitions of LD (D'Intio, 2017; Kozey & Siegel, 2008). In 2014, the province of Ontario revised its policy and memorandum No. 8 defining LD; however, two key components of the discrepancy model were retained in the definition: i) severe learning difficulties as measured on a standardized language assessment; and ii) average to above - average intellectual ability as measured on a psychological assessment (Ontario Ministry of Education, 2014). The revised provincial policy acknowledges 'exclusionary factors' such as SES and cultural differences. Nevertheless, it remains focused on a child-centric, discrepancy between achievement and intelligence in defining LD.

The lingering persistence of the discrepancy model apparent in Ontario's provincial policy is likely rooted in the 2013 revision of the DSM - 5 (Diagnostic and Statistical Manual of Mental Disorders, 5th edition), which redefined and renamed the clinical definition of LD to Specific Learning Disorder (SLD). The revised definition of SLD takes a neurodevelopmental diagnostic approach, and in so doing, attempts to eliminate the discrepancy model from the definition of LD. This new neurodevelopmental focus addresses academia's main criticisms concerning the overreliance on cut-scores from standardized cognitive assessments for diagnostic purposes. However many clinical practitioners in Ontario continue to find the relationship between IQ status and achievement under the new DSM-5 definition of SLD an ambiguous one. For example, in the absence of any cognitive processing scores, the neurodevelopmental diagnostic criteria for SLD could be inadvertently understood more broadly by practitioners. This practice could in turn result in slow learners being misdiagnosed as having LD. In response, the Ontario Psychological Association recently developed a consensus statement stressing that measuring the intellectual profile of the student should remain an important step in the process used by clinical practitioners to diagnose LD (Ontario Psychological Association, 2018). Clearly, the definition of LD is an evolving landscape in Ontario, and this definitional complexity should be taken into consideration when discussing the multitudinous learning problems for students with LD.

Learning problems for students with LD. Research has shown that students with LD manifest problems in several key learning areas, including language and mathematics. For example, studies in cognition have shown that when students with LD are compared to their same-age peers without disabilities, they display significant differences in selective attention, and time on task (Sousa, 2006; Wolfe & Nevills, 2004), as well as deficits in both short and

long-term memory (Liddell & Rasmussen, 2005). Discrepancies in visual-spatial and working memory skills have also been noted in other studies (Murphy, Mazzocco, Hanich, & Early, 2007). Researchers have also documented deficits in important language-related skills for these students, particularly in the areas of reading and writing, such as: phonemic awareness, syntax, pragmatics, word recognition, reading fluency, reading comprehension, spelling, handwriting, and written expression (Shaywitz & Shaywitz, 2006; Stillman & Scott, 2006).

Several studies have also demonstrated that students with LD can manifest difficulties in mathematics, including spatial abilities, performing complex operations, and solving word problems. McLeskey and Waldron (2011) note that students with LD make significantly less progress in math when compared to their general education or low-achieving peers. Other researchers have also demonstrated that these students suffer from mathematics underachievement in the important area of number sense, the mathematical equivalent of phonemic awareness, and problem-solving (Chard, Baker, Clarke, Jungjohann, Davis, & Smolkowski, 2008; Devine, A., Soltesz, F., Nobes, A., Goswami, U., & Szucs, D., 2013; Krawec, Huang, Montague, Kressler & de Alba, 2013). Finally, research by Whitley et al., (2007) demonstrated that when compared to their peers, students with LD struggle with several other issues related to social or emotional characteristics, and when added to their specific learning problems, contribute to overall difficulties in school.

Social or emotional characteristics of students with LD. Research related to social and emotional characteristics of students with LD has shown that these students have lower self-concept in relation to their same-age peers without disabilities, and a higher external locus of control (Ring & Reetz, 2000). Students with LD also exhibit increased levels of anxiety when compared to their peers (Lufi, Okasha & Choen, 2004), and are more prone to loneliness,

depression, and suicide (Pavri & Monda-Amaya, 2000). In school settings, fewer self-advocacy skills, such as asking for help or asking clarifying questions have been noted in the research (Test, Fowler, Brewer, & Wood, 2005), as well as persistent patterns of maladaptive behaviours such as attention problems, difficulties in classroom management and withdrawal (McKinney, 1989). Students with LD also tend to engage in risky behaviours with more frequency when compared to students without LD (Beitchman, Wilson, Douglas, Young, & Adlaf, 2001; Cosden, 2001). Kavale and Mostert (2004) demonstrated that social skill development is delayed in students with LD, particularly in building social networks and adaptive competence. These students also suffer more from low feelings of social acceptance and lack social competence when compared to their peers (Al-Yagon & Mikulincer, 2004; Wiener & Tardif, 2004). In sum, students with LD commonly struggle in school due to learning problems related to cognition, language and math, and these difficulties are often exacerbated by other social or emotional characteristics.

Arguing from a developmental perspective, Waber (2010) explores possible explanations for the difficulties students with LD face in school. The author notes that at least some of the learning problems that these students face are rooted in the fact that their educational, social or personal needs are not being met by the education system. Indeed, in a pan-Canadian study, Whitley (2008) points to several critical student, school and family characteristics which help to explain the social and academic problems that students with LD face. Whitley (2008) further notes that in order to understand the nuanced nature of school outcomes for these students, contextual factors such as parental involvement or student self-concept, must be considered in relation to one another. This is particularly true in inclusive educational settings where placement type is one of these contextual factors. Indeed, Bennett (2009) also discusses the effect of the

environment and culture of the Ontario school system on social or academic outcomes for students with special education needs, particularly the central role of teacher beliefs or attitudes about inclusion. Clearly, there are several important ecological factors influencing educational outcomes for students with LD in inclusive and non-inclusive educational settings.

Inclusive Education

Inclusive education as a general term has several meanings in the academic literature. Harrington (1997) argues that inclusive education is typically perceived as an organizational model, rather than as an educational intervention, and that the concept is most commonly stated in terms of placing special needs students in their age or grade appropriate class. In this sense, defining program placement in the context of inclusive education may be understood in terms of: i) the degree of consultation that takes place between a regular classroom teacher and special education staff; ii) a particular type of pull-out model in place for certain subjects (e.g., reading or math); or iii) a continuum of program placements, ranging from a fully self-contained special class, to a pull-out model, or a fully integrated setting. Thus, inclusive education from a service-delivery perspective can represent a wide variety of educational experiences and organizational models.

The term inclusive education is often mistakenly confused with or used synonymously with several other related terms by Canadian policy makers and other educational stakeholders. For example, 'integration' and 'mainstreaming' were terms in popular use from the 1950's to the 1980's to describe the placement of students with special education needs in the regular classroom (Lupart & Webber, 2012; Winzer, 2008). In the United States, the term inclusion is closely associated with the 'Least Restrictive Environment' mandate, posited in the Individuals with Disabilities Act (IDEA). This watershed American legislation dictates that students with

special education needs have a right to be educated with their non-disabled peers to the greatest extent possible (Felder, 2019; Hocutt, 1996).

Inclusion is also sometimes referred to as the Regular Education Initiative (REI) which situates the notion of inclusion in terms of the larger school reform debate, and in particular, merging the parallel systems of regular and special education (Kauffman, et al., 2008).

Proponents of the REI argue that the regular classroom teacher should take responsibility for any student to whom they are assigned, regardless of the diversity in educational need that student represents (Andrews & Lupart, 2000; Kavale & Forness, 2000). In this sense, inclusive education is seen as a moral imperative, or the 'right thing to do' (Mansett & Semmel, 1997). This idea fundamentally challenges the parallel systems view of special education service delivery, where inclusive and non-inclusive systems run in tandem with one another. In contrast with the full-inclusion notion, the parallel systems view of special education assumes that a separate program delivery model is more appropriate for students with certain types of special education needs. The parallel systems view would, by definition, include specialized program placements for students with LD.

Regardless of the nomenclature used to describe the term 'inclusive education', there are several different philosophical positions which are generally taken respecting the concept of inclusion (Felder, 2019; Cigman, 2007). These positions represent a range of beliefs that exist in the professional, educational, and advocacy communities regarding the benefits or detriments of the inclusive education model for students with special education needs. One position, known as 'radical inclusion', polarizes the inclusion debate, resisting the construct of disability itself (Norwich, 2008). The radical inclusion position argues for a fully inclusive education system, and justifies its position on legislative grounds, or as a completely human rights issue. Radical

inclusionists challenge the operation of a parallel education system which marginalizes and discriminates against students with special education needs (Felder, 2019; Lindsay, 2003).

Specialized classes for students with LD would be unthinkable from this perspective.

On the other hand, moderate inclusionists believe there is a philosophical dilemma inherent in the inclusion model itself in terms of treating all students with special education needs the same (equality), and differently, simultaneously (equity). Essentially, how can a regular classroom teacher treat one student as special, while at the same time attend to whole-class instruction? Further, moderate inclusionists underscore the tensions between individual and collective rights in a pluralistic, multi-value society (Felder, 2019; Cigman, 2007; Fuchs & Fuchs, 1995). Such a viewpoint requires a compromise between legitimate social needs of the many, and equally valid educational needs for specific individuals (Lindsay, 2003; Norwich, 2008). These tensions and their associated socio-political dynamics must be considered in any discussion on inclusive education. Yet, although many of these important legal and philosophical considerations are outside the scope of the present study, at least some discussion of the legislative landscape of disability rights in Canada must be addressed so that the term inclusive education can be operationally defined and properly contextualized for the present study.

Legislative foundations of inclusive education in Canada. The *Canadian Charter of Rights and Freedoms* was adopted in 1982, and this important legal document had significant implications for educational systems across the country. Foremost, the Charter specifically includes people with disabilities, and defends their rights with respect to the equal benefit of the law without discrimination based on mental or physical disability (Hutchinson & Martin, 2012). Consequently, several amendments related to special education were made to the *Education Act* in Ontario (i.e., Bill 82 and Regulation 181/98), which defines the responsibilities of school

boards when making decisions about the identification, placement and delivery of services for students with special education needs, including those with LD.

Historically, human rights activists lobbying the provincial government on behalf of students with LD have been buoyed in their efforts mainly through an important national advocacy group called, the Learning Disabilities Association of Canada (LDAC). This powerful parent advocacy organization, which has a provincial chapter in Ontario, is actively working for policy improvements that better support the multitudinous needs of students with LD across the province. More recently the Ontario Human Rights Commission (OHRC) launched a public inquiry entitled, *Right to Read*, to assess the approach that Ontario school districts are following with respect to the delivery of programs and services for students with LD using the following five benchmarks: i) Universal design for learning; ii) early screening procedures; iii) evidence-based reading intervention programs; iv) educational accommodations for instruction and assessment; and v) psycho-educational assessments (Ontario Human Rights Commission, 2019). Among the primary goals of the OHRC inquiry will be to shape provincial policy respecting the service-delivery model for students with reading difficulties, including those diagnosed with LD.

Provincial policy on inclusive education. The legislative advances in recent decades did much for setting the policy groundwork for inclusive education in terms of a human rights issue in Ontario; however, they did not specifically address the policy imperatives of inclusion as an educational practice in the province's publicly-funded school system. In other words, what exactly should inclusion look like in Ontario? In 2017, the Ontario Ministry of Education released the policy and resource guide, *Special Education in Ontario Kindergarten to Grade 12*. This document superseded the former provincial policy which emphasized accountability in developing and implementing individual education plans for special needs students on *Individual*

Education Plans (IEPs): Standards for Development, Program Planning and Implementation (2000). Nevertheless, the IEP section of the new policy remains unchanged since 2000 and the province of Ontario still has no explicit policy on inclusive education. Instead, *Learning for All: A Guide to Effective Assessment and Instruction for All Students Kindergarten to Grade 12* (herein referred to as *Learning for All*), was released in 2013 to serve as a resource guide for school boards across the province in delivering special education programs to students with special needs placed in various educational settings. Notably, the philosophy of inclusion is implied in the title of the province's '*Learning for All*' document.

In 2013, the Ministry of Education also released *Program and Policy Memorandum #119, Developing and Implementing Equity and Inclusive Education Policies in Ontario schools*. This companion document to *Learning for All* focuses on the pedagogical foundation required for instructing students with special education needs in inclusive settings, such as Universal Design for Learning (UDL), Differentiated Instruction (DI) and progress monitoring through Response to Intervention (RTI). In effect, these instructional approaches assist with planning and delivering instruction and assessment for the entire range of student abilities including both high achievers and those who require extra pedagogical support, such as students with LD.

In Ontario, the *Learning for All* resource guide forms the backbone of the inclusive education strategy for special needs students, including students identified with the LD exceptionality. Although not an official policy document, it aligns with other Ministry policies, such as, *Growing Success Assessment, Evaluation, and Reporting in Ontario Schools* (2010), which specifically addresses the modifications, accommodations, or alternative programs required for assessing, evaluating and reporting on students with special education needs. Finally, the province's strategy on inclusion entitled, *Realizing the Promise of Diversity:*

Ontario's Equity and Inclusive Education Strategy, refers not just to students with disabilities, but rather espouses a broader notion of educational reform supporting inclusion respecting all forms of student diversity in the province's publicly funded schools (Ontario Ministry of Education, 2009).

In sum, the term 'inclusion' refers to a philosophical shift from the parallel systems view of special education service delivery to a more unified view of program delivery where the educational needs of all students, regardless of disability, can be met in the regular classroom (Andrews & Lupart, 2000; Hutchinson & Martin, 2012; Kavale & Forness, 2000). Proponents of the inclusive education movement view it as a human rights issue, where the burden of proof respecting the benefits of segregation rests with those who believe students with special education needs cannot be properly supported in the regular classroom (Cole et al., 2004; Fedler, 2019; Norwich, 2008). This contrasts sharply with the view that students with special education needs require a separate education system in order to best meet their learning needs. The main philosophical implication of inclusive education is the dismantling of the parallel systems view of special education. The corollary of this position is that separate, special education systems, and their associated exclusionary practices of discrimination and marginalization, would be abolished. In this view, inclusive education implies that all students, regardless of ability, will be instructed in a regular program placement (Fedler, 2019; Lindsay, 2003; Norwich, 2008).

Program Placement and Academic Achievement

The academic literature is replete with studies discussing inclusive education in terms of its impact on outcomes for students with all types of disabilities. These studies range in focus on students with socio-emotional difficulties, physical disabilities such as deaf-hard of hearing and blind-low vision, gifted-talented students, and students with LD (Andrews & Lupart, 2000;

Hocutt, 1996; Kalambouka, Farrell, Dyson, & Kaplan, 2007). Typically, the impact of program placement is examined in relation to various social, emotional or academic outcomes, and for students with or without disabilities. Results from the research generally indicate that the inclusion of students with special education needs in the regular classroom has no deleterious effect on social or other educational outcomes for students with and without disabilities.

For example, two separate meta-analyses conducted by Calberg and Kavale (1980), and Baker et al., (1995), show small to moderate effect sizes for inclusive practices in terms of their beneficial effect on academic outcomes for students with special education needs. These are important findings in support of inclusive education. However, these authors note that improved methods are required in academic research in order to establish the relative merits of various service delivery models for students with special education needs. In fact, several methodological problems were identified in the studies under their review. These problems centre mainly on the difficulty in matching groups in terms of LD severity before making between-groups comparisons of achievement. In a Canadian review, Bennett (2009), also points to the favourable orientation for inclusive education practices on achievement for special needs students in general, with the caveat that due to the extreme variability in nature and type of disabilities, reliable and accurate information related to these outcomes is difficult to determine. Again, several methodological issues inherent in the research designs included in the analysis are underlined in Bennett's (2009) review.

Indeed, Baker et al., (1995) argued twenty-five years ago that improved research methods were required in order to establish the relative merits of various service delivery models for students with special education needs. These researchers calculated a small effect size in favour of inclusion using a meta-analysis approach that allowed for better comparisons of service

delivery models in inclusive vs. non-inclusive settings. However, the authors noted that several methodological issues latent in the research designs included in their meta-analysis should be considered. For example: i) lack of random assignment of subjects between specialized and regular classroom settings; ii) differences between the groups in terms of the types of instructional or pedagogical strategies that were employed in either setting; iii) poor methods for measuring the degree of implementation of inclusive practices in those with pre-post treatment designs; and iv) inadequate measures taken to ensure that the psychological profiles of students in either group were matched prior to comparing placement. These problems stem from the fact that quasi-experimental research designs, by nature, violate the assumptions inherent in inferential statistics such as random assignment, and homogeneity of variance across comparison groups (Marston, 1987). The authors concluded that program placement as an IV does not lend itself well to measurement.

Several international researchers have examined the impact of inclusive educational practices on students with and without disabilities and found no adverse effects on academic outcomes for either group of students (Gandhi, 2007; Kalambouka et al. 2007; Ruijs, Van der Veen & Peetsma, 2010). For example, in a large-scale study of over one million students in the United Kingdom, Dyson and Farrell (2007) found that inclusion did not impact either negatively or positively the academic attainment of students without special education needs. Kalambouka et al., (2007) reported similar findings in the U.K. related to the absence of any negative impact of inclusion on students without special education needs. These particular studies were conducted in jurisdictions outside Canada, and their findings suggest that inclusive educational practices have no deleterious effects on disabled or non-disabled students. However, one must

look province by province to see if the findings from the international research base are applicable in Canada, as each provincial education system is unique.

In the Ontario context, Demeris, Childs and Jordan (2007) used large-scale provincial data from the Education Quality and Accountability Office (EQAO) to explore the relationship between the number of students with special needs in Grade 3 classrooms and the academic achievement of students without special needs. They found that the presence of students with special needs in the regular classroom actually had a small, but positive impact on the scholastic performance of their non-disabled peers. The findings from this study using large-scale, provincial assessment data is aligned with other international research studies comparing the academic achievement of students with special needs who were included in a regular classroom with their same-age peers (Waldron & McLeskey, 1998; Zigmond, 2003). Thus, there is evidence in Ontario that including students with special needs in the regular class has a positive educative influence on students without disabilities in the classroom.

Nevertheless, there remain few research studies in Canada specifically comparing the outcomes of students with LD in different program placements. In the venerable Canadian textbook, *Children with Exceptionalities in Canadian Classrooms*, Winzer (2008) concludes that there is no existing evidence which shows the benefits of segregated or specialized educational practices in Canada when compared to those of inclusive practices for students with special education needs. Winzer (2008) further contends that no differential effects can be expected as a function of the interventions in either setting. These interventions may range from modifying the curriculum in both placement types, to redistributing resources in either setting, or establishing formal means of collaboration between teachers in either setting. In other words, the research suggests that there is no relationship between program placement and academic outcomes for

students with special education needs generally. The following section examines the research literature on program placement, and its relationship with achievement specifically for students with LD.

Literature reviews on placement and achievement for students with LD. Over the last several decades, many authors have conducted comprehensive reviews of the research literature examining the efficacy of program placement on academic outcomes specifically for students with LD. For example, Harrington (1997) argued that the evidence with respect to the influence of placement on achievement favours specialized class placement for LD students - even when a substantial amount of resources are allocated to students in the inclusive setting (i.e., special education teachers and educational professionals providing targeted support). However, it should be noted that the conclusions drawn from Harrington's (1997) study are based on a review of the literature conducted over two decades ago. The researcher states that the inclusion model in place at the time of the review failed to meet the needs of students with LD because teachers in these classrooms lacked adequate resources and training in understanding the LD exceptionality. In recent years, the importance of considering these teacher-level factors in assessing the impact of inclusive education have been well documented by other researchers in the field (Bennett, 2009, Whitley, 2010).

Not all researchers have been as conclusive about their findings on the impact of inclusion for students with LD. For example, in an influential review of literature conducted by Hocutt (1996), the author states that the research base does not support inclusion for all students with LD, noting that two important methodological issues must be considered in future research designs: i) the academic profile of the students placed in either inclusive or specialized settings (addressing the level of academic need in each group); and ii) the assessments selected for

measuring outcome variables in each group. The reason for addressing these methodological issues is that students with LD placed in more restrictive settings (i.e., specialized classes) often differ from those educated in comparison groups (regular class) on many important variables other than academic ability such as, disruptive behaviours or emotional difficulties. These particular methodological issues call into question the conclusions which can be drawn from these studies on the merits of inclusive education. Hocutt (1996) further notes that the academic measures used in many efficacy studies examining placement are not always standardized tests, and even when these measures are available, schools routinely exclude students with special education needs from participating in the testing. Consequently, the results from these studies are based on a significant amount of missing data. This limitation represents another major methodological consideration in research design.

Manset and Semmel (1997) also reviewed the results from eight different studies comparing full inclusion with pull-out programs; however, methodological concerns precluded the researchers from making any definitive statements about the superiority of either model. Specifically, these authors note that the replicability of the studies included in their review was virtually impossible due to the wide variability within the implementation of the inclusion models operating in each study of the review. They further contend that, given the complexity of inclusive educational programs, experimental designs that focus solely on individual achievement outcomes are simply not sufficient for evaluating program differences. Thus, these authors argue for more clarity in defining, and assessing the independent variables when comparing outcomes between inclusive and regular classroom models.

In an important review of the literature on program placement and achievement for students with special needs (including LD), Zigmond (2003) suggested that the research base

comparing inclusion vs. specialized models has been inherently methodologically flawed, and therefore, leads to spurious conclusions about program efficacy. The author asserts that the efficacy studies examining placement often fail to take in to account individual differences within the groups of students under study. That is, the different academic profiles of these students were seldom matched in comparisons between those placed in inclusion and regular classrooms. This is particularly problematic when reporting group averages, which tend to favour the intrinsic achievement bias for students in the inclusion groups. In other words, placement may not be the critical factor in predicting achievement for students with LD. For this reason, the author concludes that new research designs are required which explore data using techniques which can help settle the ambiguity of efficacy findings on inclusion.

According to Zigmond (2003), “the efficacy research reviewed here which spans more than three decades provides no compelling research evidence that place is the critical factor in academic or social progress of students with mild/moderate disabilities” (p. 195). Kauffman et al., (2008) concur, claiming that few methodologically rigorous studies have been conducted to date which focus specifically on students with LD in inclusive settings. More recently, Judge and Bell (2011) also emphasized the importance and difficulty of accurately measuring the academic need variable in describing the relationship between placement and achievement. However, the fact that past research has been unable to identify suitable variables to adequately assess the influence of class placement on achievement does not preclude these variables as being important to study. Indeed, the research question regarding whether or not program placement is a critical factor in predicting achievement for students with LD remains largely unanswered due to the methodological constraints of previous studies.

In their review, McLeskey and Waldron (2011) specifically examined the influence of program placement on academic outcomes for students with LD. The authors note that one of the main design weaknesses in their investigation was the lack of random selection of the sample or random assignment of treatments used in the studies under review. The researchers also argue that inclusive programs marked by a substantial amount of instructional support are not necessarily sufficient to achieve academic progress for all students with LD. Results revealed that about two-thirds of students with LD did not have their needs met in full inclusion programs, and continued to make limited educational progress in reading and math. Therefore, the authors conclude that the research base does not provide a justification for full-time inclusion programs for most students with LD.

Single studies on placement and achievement for students with LD. As noted above, a number of researchers have conducted studies exploring the influence of class placement on the academic progress of students specifically with LD. These studies are generally inconclusive regarding the impact of placement on achievement variables and point to several methodological challenges in addressing the research problem. A few of these studies are presented here for the purpose of highlighting the methodological issues inherent in the research base. For example, Marston (1987) outlined several threats to internal validity of most quasi-experimental designs examining inclusion, such as selection bias when forming comparison groups, which in turn threaten replication and generalizability of the research findings. This author notes that time-series analysis techniques can be a much more effective alternative because they eliminate the need for random assignment of subjects into groups. In a time-series design, each subject acts as their own control. Thus, researchers in this area have recognized that a more technically valid or

methodologically sound research approach is required for comparing the efficacy of special education programs.

In addressing this particular issue, Marston (1996) conducted a repeated measures analysis of covariance on fall and spring reading scores for 240 Grade 4 students with LD, using a curriculum based reading measure in three groups: i) inclusion only; ii) combined services; and iii) pull-out only. Results from this study showed that the combined services model was significantly better for reading progress than either the inclusion only or pull-out only model. Notably, the instruction was targeted on specific curriculum areas, and was provided by teachers in both the general education setting and in a pull-out resource room. This collaboration between the special education teacher and the regular classroom teacher, combined with the pull-out resource when required, was deemed most beneficial for academic outcomes for students. Findings from Marston's (1996) study help elucidate the complexities of the program placement question in terms of describing and measuring optimal service-delivery models; however, the generalizability of the study's findings is limited given that the sample size was small, particularly for the inclusion only and combined services placement conditions (i.e., $n = 33$ and $n = 36$, respectively). Moreover, there was no procedure used to match students in either educational setting on the basis of academic need prior to making comparisons on academic progress. What if the differences in achievement that were attributed to the instructional model (i.e., inclusion) were really only a function of the different psychological profiles of students in each group?

Banerji and Dailey (1995) also compared the academic achievement of U.S. fifth graders with LD in inclusive settings to a group of normally achieving peers using standardized reading tests (i.e., running record), and non-standardized writing tests. The authors noted that students

with LD in inclusive settings made academic gains comparable to their same-aged peers with no negative effects on academic outcomes for normally achieving students. However, the generalizability of the findings from this study are also limited due to the small sample size, where the two comparison groups consisted of $n = 13$ students with LD, and $n = 17$ normally achieving peers, bringing the total sample for the study to only $N = 30$ students. Again, students in this study were not matched in terms of academic need prior to making the between-group comparisons on achievement. Clearly, larger sample sizes and better matched comparison groups are required in order to make more definitive claims about the findings regarding program efficacy in either inclusive or specialized program settings.

In a larger U.S. study, Cole et al., (2004) compared the academic achievement of nearly 1,000 students with LD in Grades 2- 5 who were situated in inclusive and non-inclusive settings. The overall non-randomly assigned sample was relatively large with approximately $n = 500$ students in both placement conditions. Inclusion was defined as receiving reading and math instruction in an age appropriate classroom, whereas the pull-out model was defined as students receiving instruction from a special education teacher in a resource setting. Results revealed no significant differences between the inclusive setting and the pull-out resource model, using a curriculum based measure of reading and math (Basic Academic Skills Samples). According to the authors, their investigation provides evidence that, for students with LD, the inclusion model offers an instructional experience that may be superior to the traditional resource model. However, the instrument used to measure academic progress in this study (BASS) has been criticised for limited construct validity and potentially over-identifying students with LD (Jenkins & Jewell, 1992). Moreover, the study's authors acknowledge the main limitation of their research design that it did not examine teacher beliefs, values, understandings or practices

related to inclusive education. Therefore, the lack of information about the actual program delivered to students with LD in either program placement condition may confound the conclusions that can be drawn from the results of this particular study.

By contrast, Rea, McLaughlin, and Walther-Thomas (2002) conducted a descriptive investigation of program offerings in different placements and explored the academic outcomes of Grade 8 students with LD in two middle schools in the southeast U.S. According to the authors, analysis on performance data from a state proficiency assessment of reading, mathematics and writing, revealed no significant differences between students with LD placed in inclusive vs. specialized settings, while controlling for SES. The researchers contend that the methodology used in their study could serve as a model for those interested in evaluating service delivery models. However, this contention may be overstated as it is not clear from the description of the study's procedure how the matched sample was attained. In fact, the school attendance data (a proxy for SES) revealed that the sample was not properly matched as the average number of days absent for students was much higher for the non-inclusion school. It is entirely possible that the students at the inclusion school might have performed as poorly as the non-inclusion site if they missed as many school days. Thus, there may be limitations to making generalizations about the inclusion model based on a study involving a comparison of only two schools. The researchers acknowledge this limitation by stating that replication of their research model would be challenging in other settings.

Sociodemographic variables and achievement in students with LD. Population-based studies have consistently reported student sex differences in reading disability that favour girls (Berninger et al., 2008; Devine et al., 2013; Wheldall & Limbrick, 2010). For example, Flannery et al., (2000) found that sex differences in reading disability were evident in students with LD

irrespective of level of severity of the disorder, socioeconomic and racial differences, and even when referral biases were eliminated. The authors further contend that their findings are not an artefact of the IQ-based reading discrepancy model typically used for identifying students with LD, or greater heterogeneity of reading scores in male vs. female populations. Berninger et al., (2008) also found that in children with dyslexia, the writing problems of boys are more severe than girls in measures of automatic letter naming and written expression. Finally, Wei et al., (2013) examined math growth trajectories by disability (including LD), using a large, representative sample of U.S. students. The researchers found that math achievement gaps were significant and stable over time for students with disabilities favouring boys, and that the gap persisted from elementary to secondary school.

Although there is much evidence to suggest that sex differences in language and math achievement exist in students with LD, none of these studies examined the variable of sex relative to program placement or in terms of its relationship with academic outcomes. The central variable of program placement was not considered in the research designs. Specifically, is there a relationship between sex and achievement students with LD placed in a regular or specialized program for instruction? In sum, there is a dearth of scholarly research focusing on the relationship of achievement in students with LD, by sex and by program placement.

Several researchers have examined the influence of SES on incidence rates for students with special education needs (Blair & Scott, 2002; Coutinho et al., 2002; Morgan et al., 2009; Skiba, Poloni-Staudinger, Simmons, Feggins-Azziz, & Chung, 2005; Vandenberg & Emery, 2009; Whitley, et al., 2007). The findings from these studies generally indicate that students from lower SES backgrounds are significantly more likely to be identified with LD. In fact, Coutinho et al., (2002) investigated the relationship between students with LD and socioeconomic factors,

and using odds-ratios to analyse the data, found that SES was important in determining the probability of LD identification.

Unpacking the influence of SES on general incidence rates in special education, Skiba et al., (2005) argue that poverty itself does not fully explain the variance in identification rates for students with LD. Using multivariate analysis techniques, the authors purport that other factors such as developmental readiness may confound typical analyses of SES and its ability to predict LD incidence rates. Similarly, Whitley, et al., (2007) used structural equation modelling in comparing families of students with and without LD, and found that lower SES backgrounds were related to LD placements at a greater rate than middle or higher SES backgrounds.

Vandenberg and Emery (2009) examined the effect of SES on the remediation of intermediate-aged students with LD, and found that this variable was not a significant predictor of change over time on standardized achievement tests. The researchers note, however, that limitations in their methodology might impact the confidence that can be drawn from their conclusions. For example, different measures of achievement were used for different students among the sample population (the sample spanned a 16-year period), and thus, assumptions regarding homogeneity or treatment fidelity could not hold over time. By contrast, Wei et al., (2013) used Hierarchical Linear Modelling in their analysis, and found that math achievement levels for students with disabilities (including LD) was positively associated with SES, as measured by a mother's education and family income. Similarly, in a large sample of U.S. students in Grades 1 to 5, Morgan et al., (2009) used multilevel regression analyses, and found that socio-demographic variables (including sex and SES) were significant predictors of mathematics skills growth for students with LD.

Methodological considerations. As was shown above a plethora of studies over the last 40 years have identified significant methodological concerns in examining the impact of placement on achievement in students with LD (Baker et al., 1995; Marston, 1987; Judge & Bell, 2011; Zigmond, 2003). Consequently, results from the existing research base are generally considered equivocal with respect to the impact of program placement on the academic achievement of students with LD. Studies in this area are often outdated, irrelevant to the Ontario context, or the instructional model is not clearly specified in the research design. In other studies, the number of participants is minimal, or not matched in terms of the academic need of students with LD in either program setting. Also, the influence of other important sociodemographic variables, such as student sex or SES, has not been previously considered in relation to program placement.

To summarize, the relationship between academic achievement and program placement for students with LD remains insufficiently understood in academe, necessitating further research which specifically address the methodological limitations inherent in the research base. In short, a better research design is required. Therefore, the current study attempts to present a more methodologically sound approach for measuring the influence of program placement on achievement for students with LD by accounting for the influence of other important variables beyond program placement in the prediction model. In so doing, it applies an ecological perspective to the problem, and thereby posits a more suitable theoretical framework for understanding the factors influencing educational outcomes for students with LD.

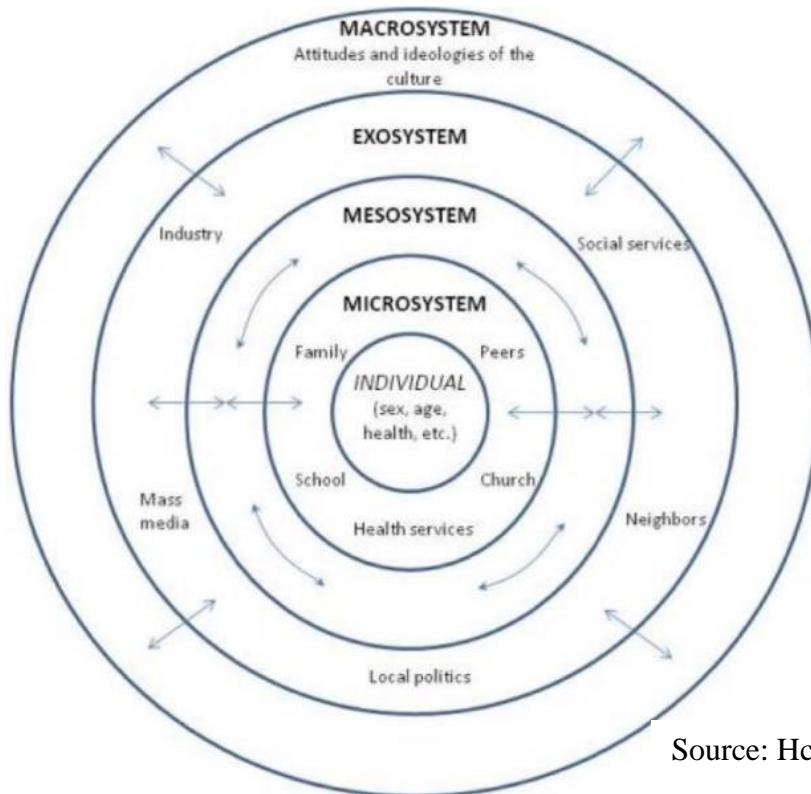
Theoretical Framework

This study applied Urie Bronfenbrenner's bioecological model of human development (Bronfenbrenner & Morris, 2006) to the Canadian special education context, specifically for

students with LD. Bronfenbrenner (1979) first developed his theory several decades ago in response to a perceived fundamental weakness inherent in other psychological theories to adequately explain human development (Lewthwaite & Weibe, 2011; Shaffer, 1993). The bioecological model of human development is an example of a systems theory (Shaffer, 1993). Under this model learning is understood in ecological terms, and is influenced by several complex interactions or systems between the individual and their environment (see Figure 2).

Figure 2

Bioecological Model of Human Development



Source: Hchokr, 2012

The model itself has evolved over time (Bronfenbrenner & Morris, 2006), but originally described human development as a multifaceted phenomenon that can be influenced within the following five sub-systems ranging from the most immediate to the most remote contexts in

relation to the individual: i) Microsystem - the closest unit of analysis to the individual, which encompasses the immediate relationships and interactions the individual student has with their family, school, or peers; ii) Mesosystem - a system of two or more interconnected Microsystems (for example a student's teacher and parents); iii) Exosystem –the connection of two or more systems where, at least one system does not actually contain the individual, but nonetheless indirectly influences that individual (e.g., a parent's work schedule); iv) Macrosystem – the system furthest removed from the individual, representing socio-cultural mores, beliefs or laws, and through cascading influence, directly affects human development (for example, a passage of law; a new curriculum); and v) Chronosystem – the dimension of time as a property of the individual's surrounding environment, and how changes to factors within this system affect development over the lifespan.

The bioecological model of human development also includes two key propositions which serve as defining properties of the theory. In applying these two distinct propositions to the current study, the first proposition implies that learning takes place through a series of progressively complex interactions between the student with LD and the elements of their environment. This proposition further contends that, in order to be effective, these interactions must take place regularly, and for an extended period of time. These types of enduring interactions are referred to as “Proximal Processes”. The second proposition purports that the form, power, content and direction of proximal processes will vary as a function of the student's own characteristics, and the environment within which they are acting (Bronfenbrenner & Morris, 2006). This proposition focuses on the fundamental interaction between the individual and their environment.

Finally, the most recent version of the bioecological theory emphasizes four key elements of the model for understanding human development, and when applied to the current study, refer to the following, respectively: i) Process - the proximal processes of learning for the student with LD (for example interactions at school or at home); ii) Person - the personal characteristics attributable to the student with LD (for example, sex or SES); iii) Context - the various systems within which the student with LD is interacting; and iv) Time - changes over time which influence the students' learning. The authors emphasize that these propositions are theoretically interdependent, and are subject to empirical test in what are called developmental science in the discovery mode (Bronfenbrenner & Morris, 2006).

The field of special education affords several examples of the application of the bioecological model of human development (Lewthwaite & Weibe, 2011). In fact, Sontag (1996) applied the model as a comprehensive framework for general disability research, and noted the model's utility for that purpose. Other researchers have applied the theory to understanding how various factors influence outcome variables for students with LD, such as the way parental expectations influence achievement (Patrikakou, 1996), and found that it is an extremely effective theoretical model for exploring these types of research questions. Keogh and Speece (1996) also published a volume from a systems perspective, examining the influence of classroom ecologies on students with LD in inclusive educational settings. The authors argued that the highly intricate processes of interactions between teachers, students and schools, particularly at the classroom level, call for a more sophisticated model for explaining these interactions. The bioecological theory of human development was that model.

Educational researchers from Manitoba also applied the bioecological model in their work, and extolled the value of the model for defining development as a complex, interactive

process, and its utility for organizing research activity (Lewthwaite & Weibe, 2011). Thus, there is precedence for using the bioecological model in the Canadian educational research context for special needs students, particularly for testing variables related to students with LD. The present study extends the application of the bioecological theory of human development by examining the influence of several ecological variables intrinsic to the individual student with LD and in their extrinsic environment as well, and exploring the relative influence of these factors on educational outcomes.

In the present study, the bioecological model of human development was applied such that the student with LD is at the center of a series of concentric spheres of influence, and achievement was the developmental aspect against which the model was tested. At the centre of the model is the student with LD, encompassing several variables germane to the individual (i.e., sex; age/grade; and level of academic need related to the LD). The next layer of the model involved two microsystems, including the main school variable of program placement, as well as the family characteristic of SES. As such, the effect of program placement in a regular or specialized program setting and the influence that this microsystem had on learning was examined. A second microsystem under study was the effect of the SES variable on the individual student's achievement. The relationship of these two microsystems was examined at the mesosystemic level.

From a theoretical lens, the prior achievement variable's relationship to the model was seen as important for understanding the influence of individual-level factors combined with system-level factors such as program placement in inclusive and non-inclusive settings. The influence of the LAN variable on achievement (synonymous with development in the bioecological model) was of particular importance to explore its differential interaction with the

school micro-system of program placement within inclusive or non-inclusive instructional settings. The sociodemographic variables of SES and SEX were related to the bioecological model because they were grounded in previous research in the area, and were well-aligned with the spheres of influence characterized in the theoretical framework. These particular variables, one germane to the individual student and the other a school-level factor, were juxtaposed against the other micro-systemic variables in the model (i.e., program placement) to create a mesosystemic research problem. In this way, each of the variables included in the research design linked explicitly to the multiple spheres of influence in the theoretical framework explaining development for students with LD within regular or specialized placements. Recognizing that there are many levels of interaction conceivably affecting the student with LD beyond the micro and mesosystemic levels, such as those operating at the macro and exo-systemic levels (e.g., policy changes regarding inclusion or changes to the diagnostic criteria in the DSM – 5), it should be acknowledged that the current study limits the application of the bioecological model of human development to the first two layers of the theory only. That is, explicitly measuring relationships at the micro and mesosystemic levels. This application of the model is not inconsistent with its original conception as it was designed to be applied and tested in such a manner (Bronfenbrenner & Morris, 2006).

Through this systems-level theoretical lens, the study's three main research questions were explored. The first question sought to determine whether or not the independent variable of program placement influenced the educational outcomes for Grade 6 students with LD after controlling for the variable of prior achievement in Grade 3. The second research question sought to understand the relative importance of the sex and SES variables in influencing achievement across placement types. The third research question specifically explored the

influence of the LAN variable on achievement, beyond the program placement variable, and after controlling for the influence of the other ecological variables in the prediction model. In addressing these three research questions, a more rigorous methodological framework was employed that better accounted for the nuanced relationship between the individual student with LD, and the other ecological factors that are influenced by, and influenced, the student at both the micro and mesosystemic levels.

Chapter Three: Methods

An Ecological Model for Exploring Achievement in Students with LD

As noted earlier, inclusive education studies are generally equivocal regarding the impact of program placement on achievement for students with LD (Andrews & Lupart, 2000; Baker et al., 1995; Bennett, 2009; Calberg & Kavale, 1980; Cole et al., 2004; Ghandi, 2007; Marston, 1996). But would this conclusion still be true after taking into account the student's prior achievement, SES, sex, or LAN in the research design? Perhaps group differences existed in these variables before the effect of program placement was even tested? Clearly, multiple ecological variables should be considered when measuring the influence of program placement on achievement for students with LD.

This study sought to follow the academic trajectory of students with LD through their elementary school journey in order to determine the unique variance in achievement attributable to the individual student across different program placement conditions. This methodological feature represents the repeated-measures aspect of the research design. The value of longitudinal analysis in comparison to traditional cross-sectional analysis has been identified by other researchers in the field as particularly important for examining the impact of program placement on achievement in students with LD (Marston, 1996; Zigmond, 2003). Creswell (2009a) also notes that the main advantage of a repeated-measures design is that it reduces the overall variability in a data set by measuring the dependent variables (DVs) in the same subjects over time. This facet of the study represents a significant methodological strength as changes in individual students over time could account for the variance in outcomes.

Many educational researchers have shown a link between SES and achievement in students with LD (Coutinho et al., 2002; Morgan et al., 2009; Vandenberg & Emery, 2009; Wei

et al., 2013). Similarly, researchers have also demonstrated the relationship between sex and achievement for students with LD, where boys are overrepresented in the population and academic outcomes typically favour girls (Berninger, et al., 2008; Devine, et al., 2013; Flannery, et al., 2000; Vogel, 1990; Wheldall & Limbrick, 2010; Young, Kim, & Gerber, 1999). For this reason these two sociodemographic variables were examined in conjunction with the program placement variable in order to assess their unique contribution to the variance in Grade 6 test scores on a standardized, criterion-referenced assessment (EQAO).

Finally, the absence of the LAN variable in research designs has been identified as a significant methodological weakness in most reviews of the literature (Carlberg & Kavale, 1980; Hocutt, 1996; McLeskey & Waldron, 2011). Studies in the area typically compare achievement and placement type without recognizing the fact that the academically weakest students are often placed in the specialized program. The current study sought to partial-out the variance associated with the LAN variable while controlling for the other independent variables (IVs) that have been shown to influence achievement for students with LD. In sum, multiple ecological variables were included in the prediction model for the current study. The sections which follow describe the methods, instruments, procedures and analyses used to develop an ecological prediction model for exploring achievement in students with LD.

Demographic Context of Sample

The current study was based on a secondary data file that was compiled as part of a program review originally conducted in 2014 for a relatively large school district in Ontario, Canada. The district under study serves over 70,000 students in nearly one hundred and forty elementary schools located in both urban and rural locations. According to proxy demographic data that was obtained from the EQAO, the population of Grade 6 students in the district was

comparable to the province for the time period under study (i.e., 2010-2014). This contextual information provides important details about the overall district enrollment; the proportion of students designated with any type of special education needs; and the number of students identified specifically with the LD exceptionality, relative to the province. In sum, the school district under study contained a population of students with LD that was similar to other school districts in Ontario. This fact has important implications for the generalizability of research findings to other jurisdictions across the province.

Participants. Participants in the secondary data file represented four separate cohorts of Grade 6 students (N = 880) who were clinically diagnosed with LD by a registered psychologist in the province of Ontario. All students in the data file had also been formally identified with LD by the school system at some point prior to Grade 6 using the Ministry of Education's formal mechanism for identifying exceptional pupils called the Identification, Placement, and Review Committee (IPRC). Historical program placement information and achievement data for the four cohorts of students was gathered for these students when they were registered in Grade 3 (i.e., 2007-2011), and then again three years later when these same students were enrolled in Grade 6 (i.e., 2010-2014). No other demographic information about these students was contained in the original secondary data file (e.g., date of birth; language spoken at home etc.).

Instruments and Procedures

EQAO assessments. Academic outcome data for the study was compiled for each student in the secondary data file through information provided by the EQAO, responsible for administering large-scale assessments in Ontario's publicly funded elementary schools. The EQAO assessments are conducted at two elementary grade levels. The primary (Grade 3) and junior (Grade 6) assessments are standardized, criterion-referenced assessments specifically

designed to assess a student's level of achievement respecting the Ontario provincial curriculum in reading, writing and mathematics. The primary assessment served as the baseline outcome measure for all students with LD for the study, while the junior assessment provided the post-test measure of achievement. As such, the EQAO assessments served as the time series, longitudinal outcome data for the study.

Several researchers have reported misgivings about the use of large-scale assessment data such as those administered by EQAO (Kohn, 2001; Nichols, Glass, & Berliner, 2006; Popham, 1999; Rogers, 2014; Volante, 2006; Wolfe, Childs, & Elgie, 2004). For example, in their external evaluation of the EQAO's assessment process, Wolfe, Childs, and Elgie (2004) highlight the unintended consequences of the use of large-scale assessment data as a way to assess school quality, rather than as an aid to formative assessment for individual students. In arguing for an alternate vision for large-scale assessment in Canada, Volante (2006) summarizes the key arguments against current practices in the use of large-scale assessment data, emphasizing the potential harm to student self-concept or disengagement in the learning process by 'teaching to the test'. Rogers (2014) also notes several changes required in order for the results from large-scale assessments to be better utilized; however, the benefits of standardized assessments to compare achievement over time were also noted. Several positive aspects of the assessments were also identified by Wolfe et al., (2004), such as their utility in informing policy-makers at the district, provincial and national levels.

Indeed, other types of outcome data have been used in research studies to determine the differential impact of program placement on achievement for students with LD, such as classroom-based assessments or standardized educational assessments (Judge & Bell, 2012; Cole et al., 2004). However, the use of non-standardized, classroom-based assessments for measuring

the impact of placement on achievement has been criticized by some researchers as a less rigorous option due to the subjective nature of these assessments (Baker et al., 1995; Kauffman et al., 2008; Marston, 1987). In addition, Thorndike and Thorndike-Christ (2010) note that the use of standardized achievement assessments such as the Wechsler Individual Achievement Test (WIAT) for measuring academic outcomes lacks the content validity required to adequately align with classroom instruction let alone provincial curriculum.

In fact, standardized achievement assessments such as the WIAT are not designed to assess a student's understanding relative to provincial curriculum, but rather to measure their ability to perform specific academic tasks against developmentally equivalent, same-age peers. For this reason, it can be argued that the use of a standardized, criterion-referenced measure such as the EQAO assessment is a more methodologically sound outcome measure for the study. Thus, the study claims increased content validity through the use of a standardized measure of curriculum, and at the same time, minimizes the subjectivity bias traditionally associated with classroom-based assessments.

LAN rating scale. The nomination process for applying to a specialized LD program in the district under study represents an unknown variable in the secondary data file. However, it is known that the process for selecting a child for a specialized program is typically formalized through an IPRC, stipulated by *Education Act*. According to Ministry of Education guidelines, the information used by the IPRC in its program placement deliberations involves input from both parents and school-level personnel. Many parents will advocate heavily for an IPRC and seek a specialized program placement for their child, while other parents will advocate for a regular class placement and resist the specialized program application entirely. Typically, the

specialized program placement advocates are stronger. Thus, parent pressure in program placement decision-making is a consideration; however, it is not always the deciding factor.

Other variables related to the nomination process that are equally relevant to parent input are school level variables. For example the degree of precision of the early diagnostic information put forward by educators at the child's school for consideration by the IPRC is not always consistent. Was the application based on anecdotal information, or were other more formal assessment and intervention processes implemented? These school level factors were also not available in the secondary data file. In any case, it is generally understood that the students who are nominated by the school for consideration for a specialized LD class placement are usually those who struggle most with basic academic skills. These students are considered more severe in terms of academic need.

In the district under study, there were many more applicants to the LD specialized program than spaces available during the period under investigation (i.e., 2010-2014). In an attempt to build equity into the program placement procedure, applicants were rated in order of priority by a team of educational professionals composed of special education staff and psychological personnel. As a means to prioritize placement offers, this team developed the LAN instrument that purported to assess several dimensions of the LD student profile, and therefore, could be used as an overall measure to rank applicants in terms of their level of academic need. Essentially, students were rated on the LAN scale and then ranked for selection purposes into the specialized program, where students with the highest LAN scores were offered placement options first. The majority of students for whom specialized placements were available took their spot; however, there were several students each year for whom no space was available. These students were placed on a high-priority waitlist for the specialized program.

For the purposes of this study, the construct of LAN was used as a proxy measure for the severity of a student's learning difficulties. This aspect of the study is particularly important as it addresses the issue of selection bias in the sample related to the severity of the LD. After pointing to differences in achievement that usually favour the regular program, many authors argue that the inclusion model is superior to the specialized program model (Bennett, 2009; Hocutt, 1996; Simpson et al., 2013). However, the inherent sampling bias in this type of research design creates an essential weakness in the comparisons of performance data. Despite the fact that not examining the LAN variable has been identified as a significant methodological gap in most reviews of the literature, it has not been previously explored by other researchers in the field (Carlberg & Kavale, 1980; Hocutt, 1996; McLeskey & Waldron, 2011). In fact, most studies in this area compare achievement and placement without recognizing the fact that the neediest students are typically placed in the specialized program.

The LAN rating scale represents a composite of several psychometric, academic and ecological variables which provide information about the applicant and their family. The various standardized test scores and other ecological variables summarized by the LAN instrument were converted into seven different Likert rating scores with ranges combined to create a composite score out of 41 for each student (see Figure 3). Notably, the LAN variable was collected for students in the specialized and waitlist groups only. Students in the regular program did not apply for a specialized placement; therefore they were not assessed with the LAN instrument and did not have a LAN score in the secondary data file.

Figure 3

Level of Academic Need (LAN) Rating Scale

Name _____ Date of Birth _____ School _____

Disqualifiers: IQ below 25thile _____ 2 academic areas not below 15thile _____ No LD diagnosis _____

A.- Highest IQ score

0 1 1.2 1.4 1.6 1.8 **2** 2.2 2.4 2.6 2.8 **3** 3.2 3.4 3.6 3.8 **4** 4.2 4.4 4.6 4.8 **5** 5.2 5.4 5.6 5.8 **6** 6.2 6.4 6.6 6.8 **7**
 <90 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 ≥120

B- Processing Deficit

0 1 1.2 1.4 1.6 1.8 **2** 2.2 2.4 2.6 2.8 **3** 3.2 3.4 3.6 3.8 **4** 4.2 4.4 4.6 4.8 **5**
 <15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

C. Ability Achievement Discrepancy

0 1 1.2 1.4 1.6 1.8 **2** 2.2 2.4 2.6 2.8 **3** 3.2 3.4 3.6 3.8 **4** 4.2 4.4 4.6 4.8 **5** 5.2 5.4 5.6 5.8 **6** 6.2 6.4 6.6 6.8 **7**
 <20 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

D. Academic Delay

- 0** 3 areas at 15thile; 2 areas ≤15thile & 1 at ≤10thile; 2 areas; 10thile & 1 ≥15thile
- 1** 2 areas ≤15thile & 1 ≤5th; 2 areas ≤10thile & 1 ≤15thile; 3 areas ≤10thile; 1 area ≤15thile & 1 area ≤10thile & 1 ≤5th
- 2** 2 areas ≤10thile & 1 ≤5thile; 2 areas ≤15thile & 1 area ≤1stile; 2 areas ≤5thile & 1 ≥15thile
- 3** 1 area ≤1stile & 2 areas ≤10thile; 2 areas ≤5thile & 1 ≤15thile; 2 areas ≤5thile & 1 ≤10thile; 1 ≤5thile & 1 ≤10th & 1 ≤15th
- 4** 1 area ≤1stile and 2 areas ≤5thile; ; 3 areas ≤5thile; 2 areas ≤1stile & 1 ≥15thile
- 5** 2 areas ≤1stile and 1 area ≤10thile; 2 areas ≤1stile & 1 ≤15thile
- 6** 2 areas ≤1stile and 1 area ≤5thile
- 7** 3 areas ≤1stile

E.

0	0.5	1
No evidence of an additional Diagnosis/Disorder	Symptoms of an additional disorder but no diagnosis	Additional diagnosis _____

F.

1	2	3	4	5	6	7
Academic delay is clearly explained by factors other than LD (e.g. absenteeism, ESL ASD, family issues)		Academic delay is due to a great extent to factors other than LD		Academic delay may be a result of factors other than LD		Academic delay cannot be explained by any other factor than LD

G.

1	2	3	4	5	6	7
No emotional/ behavioural/social problems as a result of ld		Some emotional behavioural/social problems as a result of ld		Evident emotional behavioural/social problems as a result of ld		Severe emotional behavioural/social problems as a result of ld

The district-level selection committee made its recommendations for a specialized LD program placement before an IPRC for any specific child was initiated. The LAN rating scale formed the basis for these recommendations. Thus, because LAN scores were not available for the regular group, it cannot be said with certainty that the learning disabilities for students in this group were less severe than the other two groups. Insofar as achievement scores can be considered symptomatic of a learning disability, however, the outcome data contained in the secondary data file does provide some indication of the LD severity differences across groups. Descriptive statistics from the original outcome data presented later in this chapter demonstrates that the EQAO scores for the regular group were higher than for either of the other two groups. This disparity in achievement provides solid evidence to support the assumption that students in the regular group were less severely LD than those in the specialized or waitlist groups.

In sum, the LAN score contained in the secondary data file was a numeric representation of each student's overall level of academic need for students included in the specialized and waitlist groups. Students did not require identical LAN scores or to be matched on a 1:1 basis in order to be included in the comparisons of achievement between placement groups. Instead, the LAN scores were used as a continuous independent variable suitable for hierarchical multiple regression analysis exploring the relationship of this variable with academic outcomes. Finally, it should be noted that I had no control over the nature of the LAN scale scores contained in the secondary data file. These scores, which combine standardized (e.g. IQ scores) with non-standardized scores (e.g. likert-type scale of emotional/behavioural/social problems) were provided in the secondary data file as part of the original data compilation process. No validation exercises were ever conducted on the LAN scale or on the ways in which the individual sub-scale items were brought together to create a final LAN scale score. It may have

been useful to have access to the raw IQ scores, or the standardized achievement data for analysis purposes; however, access to the raw sub-scale data summarized by the LAN was not possible due to the secondary nature of the data. This limitation was also true for the SES sub-scale scores used to create the SES index.

SES index. In a seminal review of the literature, Sirin (2005) illustrates that measuring the relationship between SES and academic achievement conceptually or empirically can be fraught with complexities related to the aggregation of different combinations of variables and applying these to a single measure. Morgan et al., (2009) also note that the effects of SES on outcomes have been measured in some studies using such limited variables as school reports of free or reduced lunch programs. Even when census variables associated with a school's location are used as a proxy for SES, these authors note that the rendering of this data into a discrete variable is less precise in terms of estimating effects (i.e., low income vs. not low income). Further, Sirin (2005) warns that the misuse of SES data could lead to an 'ecological fallacy' in data interpretation, where individual-level inferences are made on the basis of group aggregated data.

The SES instrument used for the current research project was developed by the school district involved in the study (see Figure 4). The SES data contained in the secondary data file was based on an update of the index that occurred in 2010-2011. Given that the study analyzed historical achievement data over the four-year period beginning in 2010, the data contained in the SES index was considered suitable for analysis as it reflected the most recent data available for each indicator. In other words, the index was a reasonable reflection of SES at the time it was calculated. The SES measure itself aligns with the more robust instruments discussed in the research literature in the sense that it is a composite of multiple sources of data (i.e., census vs.

family/cultural factors) combined at the individual level, and then aggregated across students within a particular school (Morgan et al., 2009; Sirin, 2005). The scores from each indicator in Figure 4 were standardized and a composite z score was calculated for each child in the school. These scores were then averaged for all students in the school to create a school-level, standardised SES score, ranging from $z = -3$ to $+3$, or a standard normal distribution. Notably, the individual level SES scores were not available due to the secondary nature of the data file. The limitation of this particular procedure is discussed later.

Figure 4*Socioeconomic Status (SES) Index*

Factor	Indicator	Source
Poverty	Average Median Income (Elementary & Secondary)	Statistics Canada: student postal codes were matched against 2009 Revenue Canada tax file data for families with school-aged children. A median income was calculated for each postal code and assigned to each student having that postal code; an average of the median income was then calculated for each school.
	Social Assistance (Elementary & Secondary)	Statistics Canada: the percentage of families receiving social assistance. The matching of student postal codes, as above, was used to calculate this measure.
	Low Income Measure (Elementary & Secondary)	Statistics Canada: the percentage of families identified as living below the Low Income Measure for their family type and size. The matching of student postal codes, as above, was used to calculate this measure.
Family/ Community	Single Parent Families (Elementary & Secondary)	Trillium: the number of students living with only one parent as of 31 October 2010 (i.e., mother OR father OR stepmother OR stepfather).
	Students Living in Foster Care/Group Homes (Elementary & Secondary)	Trillium: the number of students identified as living in foster care or group homes as of 31 October 2010.
Mobility	Absenteeism (Elementary & Secondary)	Trillium: the average number of full days missed by active students in a school as of 31 January 2010.
	Entries/ Withdrawals (Elementary & Secondary)	Trillium: the number of student entries and withdrawals between 1 October 2010 and 31 January 2011 as a percentage of October 31st enrolment (Note: the number was used for the secondary index).
	New Students (Elementary & Secondary)	Trillium: the number of students actively enrolled in an elementary/secondary school on 31 October 2010 who were not enrolled in that school on 31 October 2009.
Cultural/ Linguistic	Needs ESL Support (Elementary & Secondary)	Enrollment & Staffing Data Unit: the Funding Factor applied by Human Resources for the purpose of allocating ESL/ELD overlay staff to schools.
	New Immigrants (Elementary & Secondary)	Trillium: the percentage of students identified as not having Canadian Citizenship or Landed Immigrant status.
	First Language Other than English (Elementary & Secondary)	Trillium: the number of elementary students who have indicated their first language is not English. (Note: the percentage of students was used for the secondary index).
Readiness to Learn	Learning Skills and Work Habits (LSWH) (Elementary & Secondary)	Trillium: a single score was computed for each student based on the six Learning Skills and Work Habits identified on the report card, then an average was calculated for each school (Note: at the elementary level only one LSWH score was computed given that they are not broken down by subject area; at the secondary level, LSWHs for English and Math courses (grades 9-12) were treated as two separate indicators).

Assigning school SES scores to individual students. Although the individual socio-economic circumstances of participants were not represented by the SES variable used in the study, the ecological fallacy identified by Sirin (2005) was addressed in part by assigning the SES score for the school located within the student's actual neighbourhood, referred to as home school. For example, if a student attended an LD specialized program school with low SES but lived in a neighbourhood with high SES, that student would be assigned the SES score from their home school. In this example the student would get a high SES score. In the rare case that there was more than one school in a student's postal code catchment area, the mean of multiple schools in that area was used. This procedural modification was necessary given that students who attended the LD specialized program were frequently transported out of their home school in order to attend the school in which the program was offered. Often these schools were located in another neighbourhood with a different SES score. The study attempted to address this methodological concern by assigning a more appropriate school-level SES score to individual students at the specialized program sites. Students in both the waitlist and regular groups were assigned the SES score from their home school because these students did not relocate schools to receive their instruction.

Program placement procedure. As noted above, it is the child's home school which puts forward the application for a specialized program on behalf of the child. The committee in charge of specialized class placements in the district under study reviewed application information from a pool of approximately 100 students with LD each year. Each student was prioritized for placement in the specialized LD program based on the student's overall LAN scale rating; however, the LAN score was not the only criterion used to place a student in the specialized program. In some cases, parental preference in terms of the program's proximity to

the school, or disruption to a child's peer group were taken into consideration in making program placement recommendations. In general, however, students were rank-ordered from highest to lowest LAN score in order to develop a list of students who were recommended for a specialized LD program placement in any given year. The remaining students were placed on the waitlist for a specialized program. Notably, the LAN scores for students on the waitlist were considered sufficiently high for placement in a specialized class; thus, these students represented the quasi-experimental comparison group for the study. Barring space constraints, these students would have been placed into a specialized program; hence the term, 'waitlist' group. Service provision for the waitlist group of students was identical to all other students with LD placed in the regular program for instruction.

Procedure for merging the secondary data file. The district's centralized student information system provided the required demographic data contained in the secondary data file, specifically: i) the sex of students who had been formally identified with LD; and ii) the essential program placement information used to identify which students were placed in the regular or specialized program at Grade 6. All students were educated in the regular program at Grade 3. As noted earlier, academic outcome data was obtained from the EQAO in the form of an Individual Student Data File (ISDF) which contained raw achievement scores in reading, writing and math for both the primary and junior assessments. In order to generate a sufficient sample size for each group in the analysis (Thorndike & Thorndike-Christ, 2010), four different cohorts of students were required over four separate administrations of the EQAO assessments.

The district's student database also provided the unique numeric identifier (i.e., Ontario Education Number) required for merging the academic and achievement information in the secondary data file. This unique identifier was used to link the Grade 3 ISDF for each student

during the data compilation process and provided the baseline scores in reading, writing and math. A similar process was undertaken with the Grade 6 ISDF in order to generate the post-test data required for the analyses. The information for the LAN variable was stored in the district's paper-based, central records compiled by the placement committee each year. As such, LAN scores were assigned by-hand for each student after the achievement and demographic data was merged into the secondary data file. This data was essential for coding which students were placed in a specialized LD program, and which students were included in the quasi-comparison waitlist group. Finally, the information required for the school-level SES scores was assigned to all students in secondary data file according to their home school, regardless of program placement. Thus, the program placement, achievement and sociodemographic data for all three groups of students with LD was compiled across the four cohort years so that the time-series, longitudinal analysis could be undertaken.

Screening the Secondary Data File

Accuracy of values. Table 1 provides a full summary of the study's variables, including their type, range and other relevant properties. Given that the study is based on secondary data, an important step in the pre-analysis stage involved screening the information contained in the secondary data file in order to determine its suitability for multivariate analysis. After screening for accuracy, it was determined that all values contained in the data file were within the appropriate range for the continuous variables being employed in the study. Specifically: i) the LAN variable contained values ranging from 8.6 to 35.1; ii) the EQAO outcome data for reading, writing and math ranged from 0.6 to 4.1; and iii) the SES variable reflected standard z score values ranging from $z = -2.00$ to $z = 3.44$. Also, no out of range values were found for either of the discrete variables used in the study (i.e., PROGRAM and/or PLACE and SEX).

Table 1

Description of Original Variables in the Secondary Data File

Variable	Type	Properties	Range/Levels
Grade 3 Achievement (G3READ; G3WRITE; G3MATH)	Independent (Predictor)	Continuous	0 to 4.9
Program Placement (PROGRAM)	Independent (Predictor)	Discrete	1 = LDP; 2 = WAITLIST; 3 = REGULAR
Level of Academic Need (LAN)*	Independent (Predictor)	Continuous	Composite score (max score 41)
Student Sex (SEX)	Independent (Predictor)	Discrete/ Dichotomous	1=Male; 2=Female
Socioeconomic Status (SES)	Independent (Predictor)	Continuous	Composite score Standard Normal Distribution (i.e., $z = -3$ to $+3$)
Grade 6 Achievement (G6READ; G6WRITE; G6MATH)	Dependent (Criterion)	Continuous	0 to 4.9

*Note. The LAN variable was collected for students in the LDP and WAITLIST groups only

Inflated/deflated correlations. Another step in the data screening process examined any possible correlations between the variables contained in either the LAN scale or SES index in terms of whether or not they were inappropriately inflated or deflated. The LAN variable was operationally defined in terms of a combination of several psychometric, academic and other

ecological factors, resulting in a composite score out of 41. The SES variable was a standardized index of z -scores at the school level, assigned to each student in the data file. Although these two scales were composite measures of their respective constructs, none of the items used to develop the sub-scales within either measure were identical. Each item in the composite scale measured different aspects of their respective constructs; therefore, the correlations emerging from them cannot be considered inflated (Tabachnick & Fidell, 2007). Further, the range of responses in the continuous variables for these instruments was also not restricted in the sample. On the dichotomous variable of SEX, scores were also distributed such that there was not a larger than expected proportion of cases falling into either category.

Bivariate correlations were also conducted on the original EQAO scores in reading, writing and math. Results showed statistically significant Pearson correlations between the outcome variables at both grades, $p < .01$ (see Table 2); however, the correlations were stronger with the DVs at Grade 6. These results call into question the normality of the original outcome data, even before examining their relationship with the other predictor variables in the hierarchical regression models planned for the study (i.e., SES, SEX, LAN and PROGRAM). For this reason normality tests were conducted on the original outcome data in order to determine their suitability for the planned analyses.

Table 2

Means, Standard Deviations and Correlations of the Original Outcome Data

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. G3 READ	2.68	.86	--				
2. G3 WRITE	3.06	.46	.54*				

3. G3 MATH	2.94	.64	.55*	.42*			
4. G6 READ	3.00	.67	.42*	.20*	.36*		
5. G6 WRITE	2.98	.57	.28*	.18*	.26*	.65*	
6. G6 MATH	2.54	.77	.34*	.18*	.49*	.61*	.54*

* $p < .01$

Normality tests on the original data. Multivariate normality tests were conducted on the original outcome data for all of the achievement variables at Grade 3 and Grade 6, across the PROGRAM level IV (i.e., REGULAR; WAITLIST; and LDP). These tests were conducted as part of the data screening process in order to ensure that the outcome variables in the secondary data file were appropriate for multivariate parametric statistical analysis using hierarchical multiple regression (see Appendix A). As such, measures of skewness and kurtosis as well as their standard errors were calculated for all the original primary and junior EQAO scores in reading, writing and math across the three placement conditions.

Results indicate that non-normal distribution patterns were most pronounced in the LDP group for Grade 3 writing (i.e., skew = - 9.33; kurtosis = 24.81); in the WAITLIST group in Grade 6 writing (i.e., skew = - 6.20; kurtosis = 15.37) and in the REGULAR group in Grade 6 writing (i.e., skew = - 13.82; kurtosis = 38.98). Normality tests also showed that the WAITLIST group was approximately normally distributed in both grades, and in all three subject areas, with the exception of Grade 6 writing. Junior math scores were also slightly kurtotic for this group. In the REGULAR group, scores were positively skewed in Grade 3 reading, and not normally distributed in either Grade 3 math or writing. Grade 6 scores for the REGULAR group were also not normally distributed in reading and writing, and slightly leptokurtotic in math. In the LDP

group, writing scores were not normally distributed in either Grade 3 or Grade 6, and the Grade 6 math scores were also slightly kurtotic. Initial investigations of the visual outputs from the histograms reinforced that several of the original outcome data distributions were not normal for certain subject areas across the level of program (see Appendix B).

Addressing normality of the original outcome data. While some of the original outcome distributions approximated normality, others clearly did not. Tabachnick and Fidell (2007) suggest that in considering the normality of distributions for multiple regression techniques, the values for skewness and kurtosis measures should be as close to zero as possible; however, small departures from zero are tolerable provided that the measures are not too large to compare against their standard errors. Other authors point out that the use of non-normal distributions in multiple regression analysis is completely valid depending on the variance in the distributions themselves, and provided that the sample size is sufficiently large (Kneif & Forstmeir, 2018; Lemeshko, Lemeshko & Gorbunova, 2010; Lumley, Diehr, Emerson, & Chen, 2002). The reason for this latitude is that the normality assumption being tested in multiple regression analysis relates to the random error between the IV and the DV in the model - not the distribution of the DV itself (Bishara & Hittner, 2012).

Further, it is generally acknowledged that the consequences of violating the normality assumption in a hierarchical linear regression are minimal (Kneif & Forstmeier, 2018; Tabachnick & Fidell, 2007). For example, making a Type I error in rejecting a null hypothesis that is actually true is unlikely, particularly if the sample size is sufficiently large (i.e., $N > 200$). Although concerns stemming from violating the normality assumptions of the original outcome data were minimal, it was important nonetheless to understand the nature of the skewness and/or kurtosis values underlying these distribution patterns when considering the planned analyses for

the study. These non-normal distribution patterns in the original outcome data might be rooted in the nature of criterion-referenced assessments.

Criterion-referenced vs. norm-referenced assessment. The normality tests on the original outcome data may have been influenced by two important factors related to the nature of criterion-referenced assessment: i) the relatively large sample size in the REGULAR group; and ii) the ceiling effect acting on the DV which had an upper score limit of 4.9. According to Cramer and Howitt (2004), larger samples are prone to the ceiling effect with certain types of data which can, in turn, impact normality testing. The ceiling effect acting on the original outcome data might be explained by the fact that the EQAO assessments are standardized, criterion-referenced assessments, and as such, are meant to test students against a fixed set of criteria which students are expected to achieve at grade's end. It stands to reason that students with a milder form of LD placed in the regular class were more likely to meet these provincial expectations, whereas the more severely LD students placed in the specialised program or on the waitlist were not. In other words, the skew in the original outcome data might be explained by the criterion-based nature of the assessment where students in the inclusive setting (i.e. REGULAR group) were more likely to achieve the provincial standard.

Contrast the expected distribution patterns for a criterion-referenced assessment with those of a norm-referenced assessment, such as the Wechsler Individual Achievement Test (WIAT), which is designed to compare a student's performance against same-aged, or developmentally equivalent peers. The expected shapes of the distributions for these two types of assessments are quite different. For this reason it is not unexpected that the distributions for the criterion-referenced, EQAO tests used to measure achievement in the study did not follow the normal curve, and tended to be positively skewed in certain cases. This ceiling effect was

particularly evident in the REGULAR group where the students' learning difficulties were not as pronounced as in the LDP or WAITLIST groups.

In sum, the original outcome data were probably not normally distributed because of ceiling effect acting on the DV that was related to the criterion-referenced nature of the EQAO assessments. This fact alone does not pose a serious problem respecting assumption testing for multiple regression analyses because the normality assumption in this statistical technique applies to the residuals only. Given that the sample size for the current study was sufficiently large, any violations of the normality assumption in the regression analyses would likely have little effect on the efficiency of the prediction models to demonstrate the strength of the relationships between the variables included in the model. Further, normality results confirmed that the original data represented non-normal distributions that were mainly limited to one subject area (i.e., writing) and one group (i.e., REGULAR).

It is also possible that the normality tests of the original outcome data may have been influenced by the large amount of missing data for the Grade 3 EQAO assessment. Although the amount of missing data was not as pronounced at Grade 6, this pattern of missing data at Grade 3 is likely associated with other variables in the data file. For this reason it was important to address the issue of missing data before any reasonable modifications of the variables for subsequent multivariate analyses could be considered.

Missing Data

Given the large proportion of missing outcome data, a complete missing values analysis was conducted on the secondary data file using the SPSS Missing Values Analysis function. The parameters for inclusion were set at .01%, and included all variables in the secondary data file. Overall analysis at the variables level revealed that the SEX and PROGRAM variables were

complete with no missing data; whereas the SES variable was missing < 1% of data across cases. The remaining variables in the dataset were incomplete to lesser or greater extents, meaning that there was at least one missing value in each of the variables across all cases in the dataset. As datasets are rarely without missing data (Rubin, 1996), these results were not unusual, as will be discussed further below. Results from the case level analysis also indicated that over 90% of cases were incomplete, with some degree of missing data across certain variables. These results were also not unexpected considering both the number and nature of the variables in the study. Nevertheless, an important pattern of missing values emerged at the values level of analysis, particularly with the PROGRAM level IV.

Missing data at the values level. A general rule of thumb for researchers in assessing the ignorability of missing data at the values level of analysis is wherever there is less than 5% missing values across the entire dataset (Enders, 2016; Holmes-Finch, 2016). In the original secondary data file, the overall proportion of missing data for the DVs at Grade 6 was actually close to the acceptable range, where the missingness was approximately 5%. However, an obvious pattern of missing data was revealed within the outcome variables of reading, writing and math at Grade 3, where approximately 15% of values were missing in each level of the DV. Table 3 shows that when the PROGRAM IV is considered, the proportion of missing data for several of the original outcome variables greatly exceeds the ignorable threshold, particularly in the WAITLIST and LDP groups. When this important variable is considered, the proportion of missing data at Grade 3 represents about one third of all cases for these two groups.

Table 3

Missing Values Analysis of the Original Data by Program

	LDP (n=79)	WAITLIST (n=50)	REGULAR (n=751)	OVERALL (N=880)
<i>Variable</i>	<i>% Missing (n)</i>	<i>% Missing (n)</i>	<i>% Missing (n)</i>	<i>% Missing (n)</i>
G3READ	38.0% (28)	36.0% (18)	12.9% (97)	16.5% (145)
G3WRITE	31.6% (25)	32.0% (16)	12.0% (90)	14.9% (131)
G3MATH	35.4% (30)	34.0% (17)	12.0% (90)	15.3% (135)
G6READ	12.7% (10)	8.0% (4)	4.4% (33)*	5.3% (47)
G6WRITE	12.7% (10)	8.0% (4)	4.1% (31)*	5.1% (55)
G6MATH	17.7% (14)	6.0% (3)	4.7% (35)*	5.9% (42)

* missing data did not exceed the 5% threshold for ignorability.

In sum, overall patterns of missing data were lower for the DVs in the REGULAR group in Grade 6 and the missingness for these particular variables might be ignored without serious consequences to the interpretability of the results. However, when broken down by grade and program the overall pattern of missing data in Grade 3 was pronounced, particularly in the LDP and WAITLIST groups. For this reason, it was important to understand the scenario of the missing data.

Scenarios of Missing Data. As noted by several researchers, missing values in a data set can seriously compromise the integrity of a study's results (Enders, 2016; Holmes-Finch, 2016;

He, Zaslavsky & Landrum, 2010; Manly & Wells, 2015; Rogers, Anderson, Klinger, & Dawber, 2006). By contrast, randomly dispersed missing values in a data matrix pose fewer problems for analysis. Therefore, finding patterns of missing data is an extremely important step in the pre-analysis stage, as non-randomly missing data can have serious implications on the generalizability of results. Several authors point out that specific guidelines have not been established for assessing the tolerability of missing data for samples of various sizes particularly in secondary data sets (Rogers et al., 2006). Nonetheless, there are three main scenarios which are generally applied for understanding scenarios of missing data, and which were considered before analysis of the secondary dataset was conducted for the current study.

The first scenario is the missing completely at random (MCAR) scenario, which presents the researcher with a completely unpredictable distribution of missing data. This scenario poses no major problems for analysis, is entirely ignorable, and can be effectively dealt with through simple listwise deletion of cases (Enders, 2016; He et al., 2010; Holmes-Finch, 2016; Manly & Wells, 2015; Rogers et al., 2006a). By contrast, the missing not at random (MNAR) scenario cannot be ignored by the researcher without implications to the analysis of the results. The MNAR scenario of missing data is non-ignorable because the missing values are somehow linked to a phenomenon that is not related to the dependent variable in the dataset. Thus, in the MNAR scenario it is important to understand how much data is missing, and why it is missing. Too much missing data, or any missing data patterns, may render the dataset un-analyzable (He et al., 2010; Rubin, 1996).

The third scenario to consider is the missing at random (MAR) situation, which unlike MNAR, could be ignorable for the researcher provided that the missing data is somewhat predictable from the other variables in the dataset (Manly & Wells, 2015). According to Rubin

(1996), with the MAR scenario, there is a relationship between how the data is missing and the actual observed data. In other words, the relationship between the observed data and the missing data is not evident due to a variable not included in the dataset, as is the case with MNAR. By contrast, MAR implies that there are systematic differences between the missing data and observed values, and that these differences can be explained by other observed variables in the dataset. With the MAR scenario the researcher has several options available to treat missing data, not the least of which is simple listwise deletion. However, Klinger et al., (2006) point out that listwise deletion may not be desirable with smaller datasets as the resulting reduction in cases leads to concerns about the representative nature of the sample compared to the population, particularly when drawing inferences based on a small sample. Thus, when it can be established that the missingness is centered on patterns within the dependent variable (i.e., MAR), other methods are recommended for treating the missing data.

The case for MAR. As noted earlier, the most salient patterns of missing data were limited to the Grade 3 outcome data, particularly in the WAITLIST and LDP groups. Why is this pattern emerging more pronounced in Grade 3, and why only in certain levels of the program IV? Knowing this reason might help to explain the relationship between this predictor variable and other variables in the secondary data file. The question here is whether or not the distribution of missing data in G3 would have been different, if this could be observed, to the distribution of scores in students with complete data. For this reason, it is important to explore the other observable data in the data file as knowledge about these variables may help to explain the differences in missing academic achievement at Grade 3. In sum, understanding the relationship between variables in the dataset may help explain the patterns of missing data.

Classroom assessment practices for exempting students with LD. Reasons for the pattern in the missing data at Grade 3 may be partly related to the nature of classroom-based assessment practices that teachers typically adopt in assessing children with LD who struggle with test-taking in earlier grades. While it is true that all students in the secondary data file were formally identified with LD by the time they reached Grade 6, many of these same students had not been formally identified with LD at Grade 3. These particular children all received instruction in a regular classroom setting, and most of them were likely exhibiting academic difficulties while awaiting their formal LD diagnosis. This timeframe represents the lag between onset of the LD and formal diagnosis. During this pre-diagnosis period these students would likely have experienced greater amounts of test-anxiety related to their un-diagnosed and un-treated learning difficulties (Lufi et al., 2004). Once these same students were formally identified through the IPRC process, they would have then received accommodations to enable them to participate in the provincial tests. Therefore, it seems conceivable that teachers may have exempted some of these students from participating in the primary EQAO assessment in order to avoid any possible anxiety that participation in the tests may have caused.

The systematic exemption of primary-aged students with LD from large-scale assessments is congruent with typical classroom-based assessment practices, where younger students may be permitted alternatives to taking tests which research suggests may exacerbate their anxiety (Al-Yagon & Mikulincer, 2004; Lufi, Okasha & Choen, 2004; Martin et al., 2006; Test et al., 2005; Wiener & Tardif, 2004). Further, given that the EQAO assessments do not form part of the official Ontario elementary report card, some teachers may have opted for their students not to participate in the test. This is particularly true if the exemption practice is consistent with the student's regular classroom routines, and is documented in their

Individualized Education Plan (IEP). The EQAO's policy on exemptions and accommodations states that a student may be excused from the provincial assessments if the student is regularly exempted from classroom-based assessments, on the condition that this fact is reflected in the student's IEP (EQAO, 2019).

In 2018, the exemption rates for both the primary and junior EQAO assessments were approximately 2% of eligible students within the district under study, and these figures mirrored those across the province. Without a formal diagnosis of LD a teacher may simply state in the IEP that the student is not required to participate in the EQAO assessment due to their test anxiety. In this light, the difficulty of the primary EQAO assessment may have been beyond the abilities of younger, undiagnosed students with LD, and so teachers likely exempted them from the assessments. This knowledge about the patterns of missing data and their relationship with other variables in the data file provides strong support for considering the missingness in the secondary data file used in the study to be missing at random, or MAR.

Missing data across grade or program. In considering teacher exemption practices as a possible reason for missing EQAO data, it is striking that fewer students seemed to have been exempted in the REGULAR group, where the proportion of missing data was much lower in this group compared to the LDP and WAITLIST groups. Thus, it gives plausibility to the idea that the missing data for the students with higher academic need was due to systematic exemptions from the EQAO tests occurred when their diagnosis of LD was relatively recent, or not yet determined. The grade at which the students in the data file were formally identified is not known; however, due to the level of their academic need, many of the students in the LDP and WAITLIST groups would likely have received a diagnosis sooner than a student in the REGULAR group. Therefore, students in a specialized class placement, or those waiting for one,

may have been more likely to be exempted from taking part in classroom-based assessments when they were in Grade 3.

One possible explanation for less missingness in the Grade 6 outcome data is that by the time these same students were age-appropriate for the junior EQAO test they were formally identified with the LD diagnosis, and properly supported in their test-taking activities through the appropriate use of accommodations. This might explain why the same students in Grade 3 were not exempted from the EQAO assessments in Grade 6. Indeed, the proportion of missing data for all three groups is much lower in the junior panel. In any case, in order to be considered MAR, the missingness in the DVs should theoretically be related to some other variable in the dataset which can explain the patterns of missing data. Given that a student in either the LDP or WAITLIST groups with missing achievement data at G6 may likely (but not necessarily) have missing data in G3, the decision to treat the missing data as MAR was considered reasonable as the missingness in the outcome data was likely related to the LAN variable.

To sum, results from the missing values analysis confirmed that the sex and program variables accounted for no missing cases, while the SES variable accounted for less than 1% of missing cases in the dataset. This is clearly an ignorable scenario of missing data, where listwise deletion is suitable prior to analysis. The patterns of missing data for the DVs at Grade 6 were also not as extreme (i.e., approx. 5% overall), and were limited to the LDP and WAITLIST groups. Nonetheless, patterns of missing data for these two groups show a higher proportion in Grade 3 compared to Grade 6. This missing data pattern at Grade 3 might be explained by other variables in the data file, and for this reason, was considered MAR for the purposes of handling the missing data.

Handling missing data. Missing data is a common problem in social science research, and the field of education particularly, due to the quasi-experimental nature of the enterprise (Creswell, 2009a; Holmes-Finch, 2016; Tabachnick & Fidell, 2007; Rubin, 1996). For example, in longitudinal research designs such as the current study, subjects are more prone to attrition from treatment programs over time, and this can result in large amounts of missing data. Educational research designs comparing group means on one or more outcome variables is also prone to missing values because there are multiple variables involved in the group level analysis (Manly & Wells, 2015; Pampaka, Hutcheson & Williams, 2016). Other researchers analyzing EQAO results have encountered significant amounts of missing data in their studies (Klinger et al., 2006), and thus, were required to consider several possible techniques in order to deal with the missing data, including simple listwise deletion.

Listwise deletion is an appropriate method for handling missing data if the pattern of missingness is random (i.e., MCAR); if there are few cases which have missing data on different variables; or if the variable in question is not critical to the analysis (Creswell, 2009a; Rubin, 1996; Tabachnick & Fidell, 2007). The major drawback of listwise deletion is loss of valuable data for analysis. When listwise deletion is used to handle missing data, the power of statistical tests can also be severely compromised (Rubin, 1996). One of the major problems of analyzing data sets with large amounts of missing data are the biased parameter estimates or inflated standard errors which may result (Holmes-Finch, 2016; Manly & Wells, 2015; Pampaka et al., 2016;). Considering the patterns of missingness which emerged at Grade 3 in the original data file, where nearly one-third of cases are missing data in the LDP and WAITLIST groups, listwise deletion was not considered an appropriate method for handling the missing data. Too much information would be lost.

There are several other possible options for dealing with datasets where more than 5% of the values are missing for any variable. For example, methods of handling missing data that preserve all cases for analysis are preferable if there is some non-randomness in the pattern of missing data. Other techniques also exist that enable researchers to make estimates of missing data, where the missing values are imputed with a replaced value (Wang & Johnson, 2018; Kang, 2013). For example, single imputation methods such as mean substitution are used where the average score of a variable replaces a missing value for a given case in the dataset. However, several authors point out that single imputation methods equate to smaller standard errors of measurement and increase chance of Type I error (Enders, 2016; Kang, 2013; Marshall, Altman, Holder & Royston, 2009). These authors note that mean substitution and other single imputation methods should be avoided if the proportion of missing values is large, or if no other options for handling the missing data are available (Kang, 2013). In cases where the missingness is disproportionately large researchers typically use a more advanced and generally accepted technique called, the Multiple Imputation (MI) method.

Rationale for MI. The MI method has become an increasingly popular approach for researchers to estimate statistically unbiased data between the sample and the population it represents. The MI method is essentially a data estimation strategy generally attributed to the venerable statistician, Donald Rubin, in his early work on estimating missing data in large data sets. Many authors agree that MI is a respectable method of choice for dealing with missing data because it does not require the data to be MCAR or even MAR (Enders, 2016; Kang, 2013; Lee & Huber, 2011; Marshall et al., 2009). Indeed, MI makes no assumptions about randomly missing data and provides similar results to other regression estimation methods such as Maximum Likelihood Estimation, or Expected Maximization techniques.

The MI procedure itself replaces each missing value in a data file with a set of plausible, estimated values that each represents the value they impute (Rubin, 1996). In other words, the MI method uses a number of iterations of the observed data in order to estimate missing values with values that ‘fit best’ with those that are actually present in the dataset. Essentially, the MI method is a way of estimating data based on actual values in the dataset, and then reporting combined estimated results. Contrast the MI method with attempting to analyze a data set with a large proportion of missing information, or even using lesser methods of handling missing data such as listwise deletion or single imputation, and the rationale for using MI becomes readily apparent.

According to Rubin (1996), MI is the ideal method for handling missing data in secondary datasets that are not collected by the individual conducting the research. MI can also be effectively used when: i) the missingness is not random but related to other variables associated with the sample; ii) research designs using longitudinal or time-series data, which are prone to missing data; and iii) the proportion of missing data is high or the data set is small. Several researchers have also employed the MI technique on data sets where the proportion of missing values exceeded a threshold of 30% (Lee & Huber, 2011; Rubin, 1996). Finally, MI is also amenable to any form of the General Linear Model, such as multiple regression. Given that the current study is based on secondary data analysis; contains a large proportion of missing data; employs a longitudinal research design; and is appropriate for the subsequent planned multivariate analyses, the MI method was considered a sound method choice for estimating the missing values in the secondary data file.

Modifying Variables for Multivariate Analysis

Imputing missing data. There are essentially three steps involved in a MI (Lee & Huber, 2011; Rubin, 1996). The first step involves checking the original data file for patterns of missingness. The second step involves imputing the missing values themselves, and the final step requires pooling the imputed data in order to create parameter estimates and standard error estimates for the missing data. The main advantage of the MI method is that it accounts for the total variance estimate, as well as means for each parameter estimate, across multiple data sets or imputations. By using multiple predictions for each missing value, MI yields better estimates of variance and standard errors (Enders, 2016; Marshall et al., 2009; Rubin, 1996; Wang & Johnson, 2018). In this sense the MI method provides a truer measure of uncertainty caused by the missing data. For these reasons MI was used to replace the missing outcome data at Grade 3 and Grade 6 for subsequent hierarchical multiple regression analyses in the study.

The missing values analysis above showed that patterns of missing data were evident in the outcome data at Grade 3 as well as the DV's at Grade 6, particularly in the LDP and WAITLIST groups. The proportion of missing values was lower in the REGULAR group but still bordered on the threshold of ignorability. Therefore, the MI technique was used to multiply impute the reading, writing and math scores for the students in both Grades and across each level of the PROGRAM IV. Per standard practice, the number of imputations was set to five in order to provide sufficient estimates of the missing values for the outcome variables in each subject and grade. This repeated imputation procedure was completed in order to ensure that the pooled results for each of the estimated variables was reasonably similar between the original data set and the five multiply imputed data sets.

The resulting multiply imputed data file contained values for the following variables: G3READ; G3WRITE; G3MATH; G6READ; G6WRITE; and G6MATH. The multiply imputed data file also included original data from several variables where the missingness was considered tolerable for listwise deletion during the pre-analysis data screening stage. This list of original variables included: PROGRAM; SEX; SES; and LAN. Notably, the original data for these variables were retained for all subsequent multiply imputed data sets used in the hierarchical multiple regression analyses.

Creating the program placement variable. In accordance with the two-step plan of analysis for the study (see below), students in the WAITLIST group were combined with students in the REGULAR group before conducting the first hierarchical multiple regression analysis. The reason for collapsing these two particular groups was that, despite the fact that the WAITLIST students qualified for a specialized program, students in both groups actually received their instruction in a regular classroom setting. This methodological approach of treating all students with LD equally when comparing academic outcomes is typical in most studies examining the inclusion model (Bennett, 2009). Moreover, this approach has been criticized by several researchers primarily because the psychological profile of the students was not taken into account prior to comparing achievement outcomes in inclusive vs. specialized settings (Bennett, 2009; Hocutt, 1996; Simpson et al., 2013). This method was replicated in the first step of the analysis for the confirmatory purpose of highlighting the unique relationship of the LAN variable with the other variables included in the follow-up analysis.

Thus, in the first step of the analysis, the original discrete variable of PROGRAM, which originally comprised three categories, was recoded into a dichotomous variable called PLACE with only two categories: LDP and REGULAR. Students in the WAITLIST group were simply

added to the REGULAR group in order to create a placement condition that was congruous with other studies analyzing the influence of placement on achievement for students with LD receiving instruction in the regular program, regardless of academic need. In short, the PLACE variable grouped students into either inclusive or non-inclusive settings in the first regression model. This procedure was central for demonstrating the nuanced influence of the LAN variable with the program placement variable in the follow-up analysis.

Rationale for combining the original outcome variables. Several researchers have analysed large-scale, provincial assessment data for students with LD using multi-level regression techniques (Demeris et al., 2007; Klinger et al., 2009; Klinger et al., 2006; Pang & Rogers, 2013). These studies examined the influence of multiple independent variables on only a single outcome measure such as reading or math, or the influence of these variables on achievement in a single grade, such as Ontario's Grade 10 literacy measure (Ontario Secondary School Literacy Test), or the province of Alberta's elementary reading program. The current study originally proposed to examine the influence of multiple independent variables, simultaneously acting on multiple DVs across two separate points in time (Grade 3 and Grade 6). However, many authors note that the use of multiple DVs in multivariate analysis is discouraged if it can be demonstrated that the DVs in the model are highly correlated, or there is some theoretical or statistical rationale for combining the variables (Grice & Iwasaki, 2007; Osborne, 2017; Tabachnick & Fidell, 2007).

As Grice and Iwasaki (2007) indicate, the goal in multilevel statistical analysis is to determine how the combined variables in the analysis can maximally discriminate the influence of specific groupings of IVs on specific outcome variables. For example, with MANOVA inter-correlations amongst the independent variables are examined to interpret the underlying

multivariate effect of their linear combinations. As such, the unique contribution of math mean scores combined with reading mean scores can be measured across groups. Stated another way, a multivariate approach to understanding the IVs is not required if there is no intent to interpret the linear combinations of the variables involved in the analysis. In such cases the researcher is theoretically interested in the linear combinations of specific IVs, not a general relationship with the DV based on multiple correlations of the IVs (Grice & Iwasaki, 2007).

As noted above, the bivariate correlations conducted for both sets of outcome variables were statistically significantly correlated in the original outcome data. The correlations at Grade 3 were mild to moderate but fairly strong in the Grade 6 DVs nonetheless. Thus, from a strictly statistical lens there was a legitimate argument for combining the correlated DVs into a single outcome variable prior to analysis. Moreover, from a theoretical lens, it was also important to underscore that the study sought to understand the relationship between several ecological variables in predicting achievement generally for students with LD. There was no theoretical impetus for exploring the specific relationship between Grade 3 reading vs. writing, or Grade 3 vs. Grade 6 math in the inclusive or specialized program settings. In other words, the current study was interested in the theoretical construct of academic achievement generally, not a granular exploration of the linear combinations of different subject areas at different grade levels. Therefore, there was both a statistical and theoretical rationale for combining the original, multiple DVs into a single composite outcome variable at each grade.

Many authors note that the use of composite variables is common practice in social science research (Cone, Lang, Franklin, and Halbrook, 1994; Song, Lin, Ward, and Fine, 2013; Tabachnick & Fidell, 2007). In fact, Song et al., (2013) state that when deciding whether or not to combine several correlated outcome variables into a single composite variable, the researcher

should consider any pre-existing knowledge about the original variables themselves and the meaning derived from their combination. According to these authors, the theoretical goal in creating a composite variable is to capture the logical meaning in combining two or more variables that are correlated by analyzing the intent for considering the composite. In short, the composite variable must make sense.

It is important to understand the context, methods, and limitations for creating and using a composite variable in research design. The main advantage for creating the composite outcome variables in the current study was rooted in the decision to consider academic achievement as a general construct rather than exploring the relationship between specific, multiple IVs and DVs. In other words, exploring the linear combinations of different subject areas at different grade levels, or how these specific academic outcomes were related to program placement for students with LD, was not of primary interest. Instead, the main theoretical outcome variable of interest was academic achievement generally and its relationship with program placement.

Tabachnick and Fidell (2007) note that composite variables are typically created either by using averaging or grouping. Either simple arithmetic or weighted averaging is commonly used when the original variables are continuous in nature, while grouping is useful when the original variables are categorical (Song et al., 2013). The original outcome variables for the current study were continuous variables; therefore, simple arithmetic averaging was used in order to create the composites representing the achievement construct for each grade. To carry this out, a composite score was developed for both the primary and junior EQAO assessments based on average raw scores on the reading, writing and math sub-scales. The creation of the composite variables also controlled for the Type I error rate for small sample sizes in testing multiple comparisons in each subject area (Song et al., 2013).

One must also consider that there are several disadvantages to averaging related variables into a composite measure. Specifically: i) changing the relationship between the DV and multiple IVs in the research design; ii) a possible decrease in statistical power due to over-reduction or loss of information; and iii) the ambiguity in interpreting the relationships between the combined variable and the original variables (Song et al., 2013). The disadvantage of averaging related variables becomes evident in the context of the current study when considering that students with LD may manifest their learning difficulties in one subject area yet still perform normally in another. For example, a student might meet the provincial standard in EQAO math but fail to meet the standard in reading or writing. Averaging EQAO scores across subject areas risks concealing the unique differences in achievement for these particular students.

However, many students with LD also manifest learning difficulties in both literacy and numeracy skills. Therefore, it could be argued that the nature of the relationship between the multiple DVs and IVs in the current study was not altered by the creation of the composite variables. In fact, the intent of developing composite variables was to produce a more comprehensive picture of the relationship between the outcome variable of academic achievement and the predictor variable of prior achievement. Moreover, the study's main variable of interest was program placement, and when it comes to interpreting results, the creation of the composite outcome variables actually provided a more concise and meaningful relationship with the main IV. In sum, the rationale for creating the composite variables was grounded in both statistical and theoretical terms.

Creating the composite outcome variables. Given that the achievement scores in the data file were significantly correlated for both grades, and that there was both a theoretical and statistical rationale for so doing, it was decided to create a single composite variable of

achievement at both grades for use in the two-step analytical plan (see below). As such, a new variable was created for each grade in order to replace the three individual outcome variables of reading, writing and math, representing the multiple DVs in the research design. The resulting criterion variable for Grade 6 was called the 'SUPER6' variable, and was calculated using the average scores of the three original outcome variables in the original data file (i.e., G6READ, G6WRITE and G6MATH). A similar procedure was conducted with the predictor variables at Grade 3, and this new variable was called the 'SUPER3' variable.

To summarize, the MI procedure was employed in order to address the substantial amount of missing data in the original outcome data. The program placement variable was also recoded in order to allow for a two-step analytical procedure designed to highlight the nuanced role of program placement in conjunction with the other ecological variables in the conceptual research design, most notably the LAN variable. Finally, for both statistical and theoretical reasons, the original multiple DVs were combined in order to create two 'super' variables that reflected the general construct of academic achievement and rendered the secondary data file more suitable for subsequent hierarchical multiple regression analyses. The next section presents the rationale for a two-step analytical procedure using the multiply imputed data file, including the new PLACE variable, as well as the two newly created composite variables (SUPER3 and SUPER6).

Analytical plan. In applying a systems-level theory to the current study (Bronfenbrenner & Morris, 2006), the microsystems of school represented by the program placement and SES variables were hypothesized to influence achievement in students with LD. Analyzing the influence of these variables in addition to the other micro-level variables germane to the individual student (i.e., SEX, SUPER3), created a mesosystemic research problem. Given that

one of the study's three research questions sought to specifically explore the contribution of the LAN variable on the inclusion model, a two-step plan of analysis was conceived which could elucidate the nuanced effect of this crucial IV (see Table 4).

Table 4

Analytical Procedure

	Predictor Variables	Order of Entry
Analysis #1	1. Sociodemographic variables (SES; SEX);	Block #1
	2. Prior achievement (SUPER3); and	Block #2
	3. Program Placement (PLACE)*	Block #3
Analysis #2	1. Level of Academic Achievement (LAN);	Block #1
	2. Prior achievement (SUPER3); and	Block #2
	3. Program Placement (PROGRAM)**	Block #3

*PLACE = LDP vs. REGULAR + WAITLIST; **PROGRAM = LDP vs. WAITLIST

The first step of the analysis employed hierarchical multiple regression in order to explore the degree of the relationship between the criterion variable of academic achievement in Grade 6 students with LD, and the relative importance of four other predictor variables: i) program placement (regular vs. specialized); ii) student sex; iii) SES; and iv) prior achievement in Grade 3. As was noted earlier, Group 1 was compared against Groups 2 and 3 combined during this analytical step, where students from the waitlist group were included with all other students with LD in the regular program. This analytical approach is typical of most inclusive education studies which examine the influence of program placement in predicting achievement

without considering the severity of the students LD (i.e., LAN) placed in either program condition. Thus, in the first analytical step the main IV of program placement was examined in conjunction with the other predictor variables in the model over the entire sample.

A particular variable of interest in the follow-up analysis was the LAN variable which focused on the students in Group 1 (specialized) and Group 2 (waitlist) who represented a matched comparison group in terms of academic need. Thus, the second step of the analysis used hierarchical multiple regression in order to partial-out the unique variability in Grade 6 test scores associated with the LAN variable in addition to the other relevant variables identified in the first step of the analysis (i.e., prior achievement in Grade 3). In this way, the application of a two-step analytical procedure elucidated the mesosystemic relationship between the student with LD, and their instructional environment (i.e., program placement), as theorized in the bioecological model of human development (Bronfenbrenner & Morris, 2006).

Rationale for hierarchical multiple regression. Tabachnick and Fidell (2007) note that the hierarchical multiple regression technique is particularly useful when the IVs involved in the research design are correlated with one another and the researcher wishes to determine the relative importance of specific variables in predicting the criterion variable. In standard multiple regression variables are entered into the prediction model simultaneously with no consideration given to the order of priority specified by the researcher. By contrast, hierarchical multiple regression techniques are typically used when the order of entering variables into the prediction equation is based in theory or specified by the researcher (Osborne, 2017; Tabachnick & Fidell, 2007). This technique can be more effectively used to create an overall prediction model for a criterion variable and is also ideal for determining the relative importance of these predictors in contributing to the overall coefficient equation (Osborne, 2017). As noted above, the need for

assessing the influence of multiple variables in predicting academic achievement for students with LD is well-documented in the research literature. Thus, hierarchical multiple regression was selected as most the appropriate multivariate technique in order to examine the specific effect of the variables in this study.

Notably, other multilevel data analysis techniques such as Hierarchical Linear Modelling (HLM) may be used when data are ‘nested’ within schools (Albright & Marinova, 2010; Hox, 1995; Newton & Llosa, 2010; Tabachnick & Fidell, 2007). According to Hox (1995), a multilevel problem concerns the relationships between variables that are measured at a number of different levels. For example, students located within the same classroom or school can be similar to each other, but differ significantly from students in other classrooms or schools. This phenomenon is called ‘nesting’ and systematic differences among students nested in classrooms or schools can lead to ‘spurious significances’ using traditional statistical analyses (Tabachnick & Fidell, 2007). Under these circumstances, HLM may be considered appropriate.

Albright and Marinova (2010) note the main advantage of HLM is that it can generate more appropriate standard errors for estimates of program effect as well as other key parameters when compared to statistical techniques like linear regression that may ignore data clustering. Newton and Llosa (2010) contend that there are three main advantages to using HLM in K–12 educational settings. Specifically, HLM provides: i) a proper estimate of average program outcome and effect of other predictor variables; ii) an estimate of variation in student outcome; and iii) a framework for examining factors related to variation in program outcomes. In fact, several Canadian researchers have employed the HLM technique in order to address the hierarchical nature of large-scale achievement data (Klinger et al., 2009; Klinger et al., 2006;

Pang & Rogers, 2013). In sum, HLM is typically employed to account for the constrained variation and misleading potential of averages taken at the school level.

Although HLM has much credence as a statistical modelling technique, there are several reasons why this approach was not appropriate for the current study. First, the secondary data file was composed of information about four different cohorts of Grade 6 students with LD, gathered over a four-year time period (i.e., 2010-2014). Considering the smaller number of students in the LDP group ($n = 79$), the vast majority of these students would certainly not have shared the same classroom teacher, let alone the same school, in any given year in the sample. Moreover, the specialized program location sites varied over the four year timeframe and were scattered across several schools throughout the district. This variability in specialized program locations likely resulted in very few students in the LDP group receiving instruction in similar classrooms or schools. Unfortunately, the secondary data file contained no specific information confirming which schools, classrooms or teachers each student was assigned.

A second important consideration respecting the nestedness of the secondary data is the nature of classroom organization in schools respecting students with special education needs. This point is particularly applicable for students in the REGULAR and WAITLIST groups who may not necessarily have been nested in classrooms. Often schools are organized so that the proportion of students with special education needs is evenly distributed across the classrooms in the school. This ensures that no single teacher is responsible for more than their 'fair share' of higher need students, and results in a balance in the concentration of students with special education needs in any given classroom. Following this logic, the majority of students with LD in the REGULAR and WAITLIST groups would likely not have received instruction from the same classroom teacher, but rather, would have been spread across a number of classrooms in a

school. Keeping in mind that the secondary data file represented four different cohorts, any school or classroom-level nested effects would have been diluted due to the relatively small number of students with LD in each classroom.

One final consideration respecting the suitability of HLM for the current data file is the relatively large number of cases required for HLM analysis (Albright & Marinova, 2010; Hox, 1995; Newton & Llosa, 2010). Even if information about the students' school or teacher was available in the secondary data file, it would likely not have yielded a sufficient number of students required for HLM analysis. Adding other independent variables to the model (i.e., SEX and SES) would simply reduce the overall observed cell sizes and exacerbate this problem. Thus, any limitations of the current analysis respecting the 'nested' nature of students in schools, or potential concern that the outcome variables were constrained by classroom or school-level effects, should be mitigated by the longitudinal, secondary nature of the data.

Nevertheless, one of the goals of the present study was to assess the relationship between placement-type and academic outcomes in students with LD, as well as the relative importance of several other ecological variables in predicting these outcomes. Tabachnick & Fidell (2007) note that the general purpose of multiple regression is to assess the degree of relationship between a criterion variable and one or more independent variables. Further, the hierarchical multiple regression technique is also well-suited to Bronfenbrenner's systems level theory examining several micro-level factors influencing the individual student with LD and which combine to create a meso-systemic problem. The regression technique examines the influence of several variables at different blocks or layers; therefore, the rationale for using hierarchical multiple regression analysis was deemed sound for the study.

The hierarchical multiple regression technique is also well-suited to Bronfenbrenner's (1979) systems level theory examining several micro-level factors influencing the individual student with LD, and which may combine to create a mesosystemic problem. This regression technique examines the influence of several variables at different blocks or 'spheres'; therefore, the link between the bioecological model and the rationale for using hierarchical multiple regression analysis was a clear one. The standard method of entry in multiple regression is simultaneous, where all IVs are entered into the equation at the same time, whereas sequential regression allows the researcher greater control over the entry of variables in the model (Tabachnick & Fidell, 2007). In this study, the selection of variables, and their order of entry into the hierarchical regression model for each step, was decided based on logical and/or theoretical considerations, using the bioecological model for human development and/or the extant research base as a guide for the entry order of specific variables.

In the first analysis, the sociodemographic variables closest to the individual student were entered in the first block as they were logically related and grounded in previous research. In theoretical terms, the SES variable represented a separate microsystem which was explored in relation to the program placement microsystem. In accordance with previous research, the individual level IV of prior achievement was entered into a second, separate block, and explored for its unique relationship with the DV. Finally, the program placement IV, representing the school-level microsystem under the theoretical model was entered in the 3rd block. A similar logic was used for the order of entry of the IVs in the second regression analysis, particularly for the LAN variable. Thus, the decisions regarding which variables to include in each level of analysis, and their order of entry into either regression model, were connected with the theoretical model for the study.

Chapter Four: Results

Sample Characteristics

Demographics. The original secondary data file was comprised of a total of $N = 880$ Grade 6 students with LD, categorized into the following three sub-groups: Group 1 LDP (the specialized program group); Group 2 WAITLIST (the waitlist group); and Group 3 REGULAR (the regular program group). Table 5 demonstrates that the group sizes were unequal across program placement conditions, where students in the LDP and WAITLIST groups represented approximately 9% and 6% of the total sample, respectively. The remainder of the sample (i.e., 85%) was represented by the REGULAR group.

Table 5

Descriptive Statistics by Program and Sex

	<u>Group 1</u>		<u>Group 2</u>		<u>Group 3</u>		<u>OVERALL</u>	
	LDP		WAITLIST		REGULAR			
<i>Variable</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>n</i>	<i>%</i>
Males	54	68.4	38	76.0	489	65.1	581	66.0
Females	25	31.6	12	24.0	262	34.9	299	34.0
Total	79	100	50	100	751	100	880	100

The sample also consisted of 581 males and 299 females; thus, males comprised approximately two thirds of the overall sample. Notably, the unequal ratio in student sex was reflected across placement conditions as well, where the proportion of males was much larger

than females in each group (i.e., 68%, 75% and 65%, in the LDP; WAITLIST; and REGULAR groups, respectively).

Socioeconomic status. Descriptive statistics for the SES variable indicate that students in the LDP group had lower average SES scores than students in either the WAITLIST or REGULAR groups, whereas variability was highest in the LDP group when compared to either the WAITLIST or REGULAR groups (see Table 6). In order to determine whether or not the mean SES scores differed significantly between the groups, a one-way between-subject analysis of variance (ANOVA) was conducted on the SES variable using PROGRAM as the independent variable. Results from the ANOVA indicate that there was a statistically significant difference between the group means for the three groups, $F(2, 871) = 4.449, p < .05$. Post hoc comparisons (i.e., Scheffe) showed that the average SES score for the LDP group was significantly lower than that of the REGULAR group, but not the WAITLIST group. The mean SES scores in the REGULAR and WAITLIST groups also did not differ significantly from each other.

Table 6

Socioeconomic Status Descriptive Statistics by Program

Group	<i>n</i>	<i>M</i>	<i>SD</i>
LDP	79	.06	1.05
WAITLIST	50	- 0.25	0.71
REGULAR	751	- 0.24	0.83

Level of Academic Need. Descriptive statistics in Table 7 indicate that the mean LAN score was significantly higher in the LDP group than in the WAITLIST group, suggesting that

students in the specialized program were more likely to be rated higher in academic need than students who were on the waitlist for a specialized program. Given that the LAN scores were not calculated for students in the REGULAR group, an independent samples t-test for two independent groups was conducted on the mean LAN scores between the LDP and WAITLIST groups (see Table 7). Results show that the mean LAN scores were statistically significantly higher in the LDP group compared to the WAITLIST group.

Table 7

Independent Samples t-test of LAN by PROGRAM

Group	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t-test</i>	<i>df</i>
LDP	79	26.02	5.26	5.86*	127
WAITLIST	50	19.42	7.53		

* $p < .001$

In sum, descriptive statistics from the sample's demographic characteristics show that group sizes were unequal across program placement type, where the vast majority of students were placed in the REGULAR group. Males also comprised two thirds of the overall sample, and this disproportionality in SEX was clearly reflected across the PROGRAM IV. Results from the analysis of group differences on the SES variable showed that students in the LDP group had significantly lower SES than students in the REGULAR group, but not the WAITLIST group. Analysis of group differences for the LAN variable also indicated that students in the LDP group were significantly higher in academic need than students in the WAITLIST group.

Descriptive statistics of the composite outcome variables. Table 8 presents the means, standard deviations, skewness and kurtosis values by program for both the primary and junior

composite outcome variables. Results from the average group differences show that mean scores of the outcome data decreased over time for all three sub-groups of students. However, the decline in average test scores was not as drastic for the students in the REGULAR group. The academic trajectory for these students was fairly stable from Grade 3 to Grade 6. On the other hand, the decrease in Grade 6 test scores for the students in the LDP and WAITLIST groups was more pronounced. Indeed, the trajectories for these more severely LD students tended to appear more steeply negative, where their achievement was relatively low in Grade 3 when compared to students in the REGULAR group, and then became even lower in Grade 6, regardless of program placement.

Table 8

Means, Standard Deviations and Normality tests of the Composite Variables

		<u>REG (n = 751)</u>	<u>WAITLIST (n = 50)</u>	<u>LDP (n = 79)</u>
SUPER3	<i>M</i>	2.93	2.72	2.49
	<i>SD</i>	0.53	0.53	0.55
	skewness	-6.31*	- 0.02	-0.22
	kurtosis	3.01*	- 0.71	0.28
SUPER6	<i>M</i>	2.91	2.48	2.25
	<i>SD</i>	0.54	0.58	0.57
	skewness	-5.71*	-2.42*	0.63
	kurtosis	2.62*	1.72	0.92

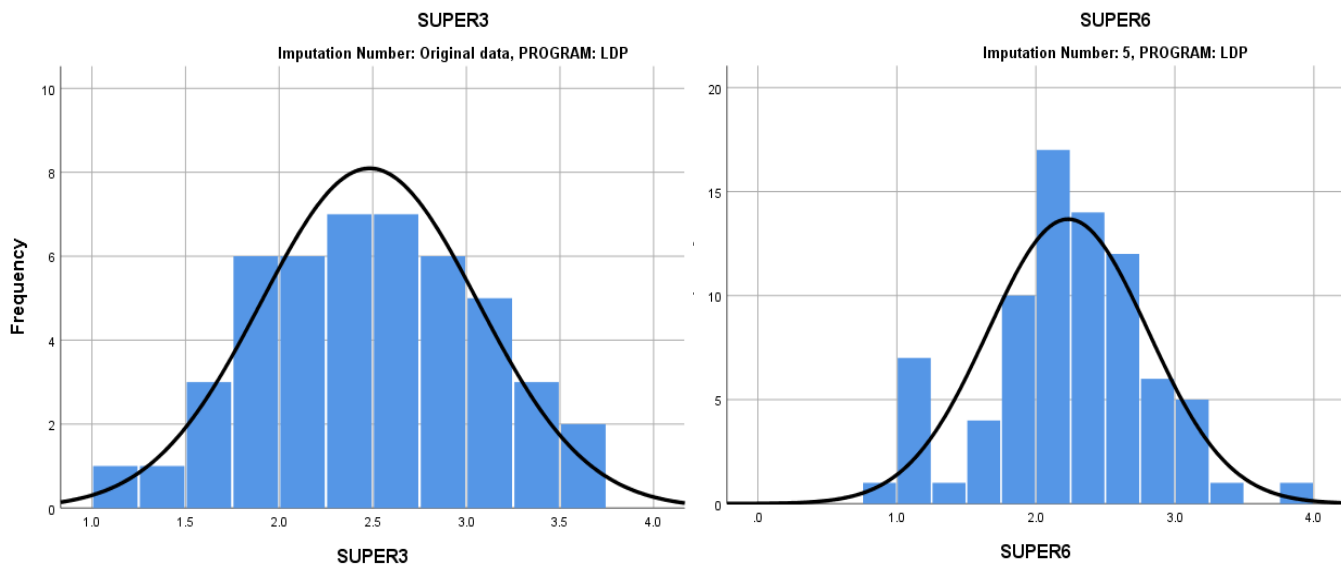
Note. * value exceeds the +/- 1.96 normality threshold

Normality tests confirmed that both the SUPER3 and the SUPER6 variables were skewed and leptokurtic for the REGULAR group. The distribution of the SUPER6 variable was also

moderately skewed in the WAITLIST group, but not for the LDP group. Visual inspections of the distributions for the SUPER3 by the PROGRAM variable suggest that these distributions were neither skewed nor kurtotic in both the WAITLIST or LDP groups (see Appendix C). In fact, visual inspections of the histograms for both composite outcome variables indicate that these variables represented approximately normal distributions across the level of PROGRAM, particularly for the LDP group (see Figure 5).

Figure 5

Distributions of the Composite Outcome Variables



The distributions for both composite variables were technically not normal in the REGULAR group. However, as was noted earlier, criterion-referenced tests are susceptible to the ceiling effect (Cramer & Howitt, 2004), and this was likely the case with the SUPER3 variable in the REGULAR group where the skew was more pronounced. One would expect the ceiling effect to transfer to the composite variables given that they were based on average scores of the original data. Moreover, Lumley et al., (2002) point out that normally distributed data are not necessary for multiple regression analysis in larger samples, and the normality assumption

applies only to the residuals of the DVs (Lumley et al., 2002; Tabachnick & Fidell, 2007). These two points are important to consider in assessing the remaining assumptions of the study's two-step plan using hierarchical multiple regression analyses.

Assumptions of the Hierarchical Multiple Regression Analyses

Ratio to cases assumption. According to Tabachnick and Fidell (2007), the rule of thumb for meeting the assumption of the required cases to IVs ratio in multiple regression analysis is, $N > 50 + 8(x)$, where (x) is the number of IVs for testing the multiple correlation. More cases are required when the DV is skewed; when smaller effect sizes or measurement error due to less reliable variables are expected; or when conducting a stepwise regression. In the first regression analysis, there were a total of four predictor IVs (i.e., SES; SEX; SUPER3 and PLACE); therefore a total of 82 cases were required in order to meet the cases to IVs ratio assumption. Given that the SUPER6 DV was skewed, and that the SES variable could be considered a less reliable indicator for socioeconomic status, a larger ratio of cases to IVs was deemed necessary to meet this particular assumption. There were nearly 900 cases included in the first step of the analysis; thus, the assumption of ratio of IVs to cases was clearly met.

For reasons of parsimony, the SES and SEX variables were not added to the regression model in the follow-up analysis which included three independent variables: PROGRAM; LAN; and SUPER3. This decision resulted in a lower required cases to IVs ratio (i.e., $N = 74$), and the total number of cases involved in the second step of the analysis was nearly $n = 150$ students. Therefore, the ratio to cases assumption was also met in the second regression model. Finally, the number of cases in both analytical steps also met the criteria for the overall multiple correlations, as well as for testing individual predictors for a medium-size relationship between the IVs and the DV (i.e., $N > 109$).

Outliers, multicollinearity/singularity, linearity and homoscedasticity. The secondary data file was screened during the initial stages of the analysis for univariate extremeness in order to uncover any outliers which might have an unnecessarily large impact on the regression models. It was determined that all values in the data file were within the appropriate range for the continuous variables being employed, SUPER3; SUPER6; LAN and SES. Also, no values out of range were present in the discrete variables being used in either regression model (i.e., PLACE and/or PROGRAM; and SEX). Therefore, the assumption for extreme outliers was met for both steps of the analysis. The absence of multicollinearity and singularity were also tested by examining the size of the standard errors for the regression coefficients relative to the scale of the variables (Tabachnick & Fidell, 2007). The multicollinearity assumption was tested for any value greater than $r = .7$ between the predictors themselves in the correlation matrix. This test also required that the predictor variables were reasonably correlated with the DV. Tolerance tests for collinearity statistics were also conducted in order to determine whether or not the tolerance measure (i.e., VIF) was greater than 0.1 (Hair, Anderson, Tatham, & Black, 1995). Thus, the assumptions of multicollinearity and singularity were also met.

An analysis of the residual scatterplots and normal P-P plots provided tests for the assumptions of normality, linearity and homoscedasticity (see Appendix D). Differences between the observed and expected residuals for the DV scores from both analytical steps were examined for normal distribution as well as inspected for a straight line relationship with the DV (Tabachnick & Fidell, 2007). Visual inspections of the normal probability plots for both regression analyses also indicated that the points generally followed the least squares line; whereas the scatterplot in the first regression model indicated a slight tendency toward heteroscedasticity. This was likely due to the fact that students in the REGULAR group were

included in the first regression model. Recall that the SUPER6 DV was positively skewed for this group.

Tabachnick and Fidell (2007) indicate that the failure of the heteroscedasticity assumption may weaken the analysis in multiple regression; however, the analysis can still be valid provided that the spread of the residuals in the overall scatterplot is rectangularly shaped. Such was the case in the scatterplots for both regression models (see Appendix D, Figures 6 and 7). Residual statistics also demonstrated that Cook's distance was not above .1, and that the minimum and maximum values of the standard residuals were within the acceptable range of -3 to 3 for both steps of the analysis. In sum, visual inspections of the normal P-P plots and scatter plots for the SUPER6 variable for both hierarchical multiple regression models indicated that the residuals for the DV met the assumptions of normality, linearity and homoscedasticity for both steps of the analysis.

Not Accounting for Level of Academic Need

A hierarchical multiple regression was performed in the first step of the analysis which included students from all three groups. This was done in order to explore the degree to which the influence of the program placement variable (PLACE) predicted the Grade 6 outcomes for students with LD, after controlling for the SES, SEX and prior achievement (SUPER3) variables. Per the analytical plan, the LAN variable was not included in the first step of the analysis. Results in Table 9 show that: i) the correlations amongst the IVs themselves were weak to moderate; and ii) the correlations between the IVs and the DV were weak to moderately strong. Three of the four predictor variables were significantly correlated with the criterion variable of achievement (SUPER6), which indicates that these variables were suitable for examination through hierarchical multiple regression analysis.

Table 9

Descriptive Statistics and Correlations from the 1st Regression Model

<u>Variables</u>	<u>SUPER6</u>	<u>SUPER3</u>	<u>SES</u>	<u>SEX</u>	<u>PLACE</u>
SUPER6	1				
SUPER3	0.46***	1			
SES	- 0.10**	- 0.07*	1		
SEX	0.03	- 0.06	.02	1	
PLACE	- 0.31***	- 0.22***	0.10**	- 0.02	1
<i>N</i>	880	880	874	880	880
<i>Means</i>	2.83	2.88	-0.21	1.34	0.09
<i>Standard Deviations</i>	0.58	0.55	0.85	0.47	0.29

Statistical Significance: * $p < .05$; ** $p < .01$; *** $p < .001$

In the first block of the model, the two sociodemographic variables of SES and SEX were entered. This model was statistically significant, $F(2, 871) = 5.09$; $p < .01$ and explained 1% of the total variance in Grade 6 achievement (see Table 10). Only one variable in the first block of the prediction model, SES, made a statistically significant contribution to the variance in the criterion variable. The addition of the prior achievement predictor (SUPER3) in the second block of the model was also statistically significant, $F(1, 870) = 220.94$; $p < .001$, and explained an additional 21% of the total variance in Grade 6 achievement. All three IVs made a statistically

significant contribution to the variance in the criterion variable in the second block of the prediction model.

After the entry of the main variable of interest, PLACE, into the third block of the model, the total variance explained by the combined IVs from all three blocks was 26%, where the predictive model as a whole was statistically significant, $F(4, 869) = 74.13, p < .001$. The introduction of the PLACE variable explained an additional 4% of unique variance in Grade 6 achievement, after controlling for the SES, SEX and SUPER3 variables, where R^2 change = .04; $F(1, 869) = 49.31, p < .001$. In the final adjusted model, two of the four predictor variables were statistically significant, with the SUPER3 variable recording a higher Beta value ($\beta = .41, p < .001$) than the PLACE variable ($\beta = -.21, p < .001$).

Table 10

Regression Model Not Accounting for Level of Academic Need

	<u>R</u>	<u>R²</u>	<u>R² Change</u>	<u>B</u>	<u>SE</u>	<u>β</u>	<u>t</u>
Block 1	.11	.01**					
SES				.05	.04	-.10*	-3.02
SEX				-.08	.02	.04	1.19
Block 2	.47	.22***	.21***				
SES				-.05	.02	-.07*	-2.31
SEX				.08	.04	.07*	2.07
SUPER3				.48	.03	.46***	14.46

Block 3	.51	.26***	.04***		
SES				- .04	.02
SEX				- .05	- 1.73
SUPER3				.07	.04
PLACE				.06	1.93
				.44	.03
				.41***	12.75
				- .44	.08
				- .21***	- 5.69

Statistical Significance: *** $p < .001$; ** $p < .01$; * $p < .05$

Adding Level of Academic Need to the Model

A follow-up analysis was performed in order to explore the relationship of the program placement variable with achievement after controlling for the unique influence of the LAN variable. For this reason, the follow-up analysis included only the students in the LDP and WAITLIST groups who were assigned LAN scores in the secondary data file. The prior achievement variable (SUPER3) was also included in the follow-up analysis because it made a statistically significant contribution to the first prediction model.

Tabachnick and Fidell (2007) emphasize that researchers should take a parsimonious approach respecting the number of variables to include in a regression model for the sake of the model's efficiency; therefore, the researcher should determine which variables will make the prediction model most efficient. Typically this is done by limiting the number of IVs included in the model. Due to the fact that neither the SEX nor the SES variables made a statistically significant contribution to the criterion variable in the first, final adjusted model, it was decided not to include these variables in the follow-up analysis. Finally, as per the analytical plan, students from the REGULAR group were not included in the follow-up analysis. Therefore, the

PROGRAM variable was used as the grouping IV in the second regression model rather than the PLACE variable. The correlation matrix from the follow-up analysis indicated that the correlations between the predictor variables and the criterion variable were weak to moderately strong, where both the SUPER3 and PROGRAM variables were significantly correlated with the SUPER6 variable (see Table 11). The correlations among the IVs were also weak to moderately strong where both the SUPER3 and LAN variables were significantly correlated with the PROGRAM variable. Results from the correlation matrix confirmed that the data was suitable for examination through hierarchical multiple regression analysis.

Table 11

Descriptive Statistics and Correlations from the 2nd Regression Model

<u>Variables</u>	<u>SUPER6</u>	<u>SUPER3</u>	<u>LAN</u>	<u>PROGRAM</u>
SUPER6	1.00			
SUPER3	0.19*	1		
LAN	0.04	0.02	1	
PROGRAM	0.19*	0.20*	- 0.46**	1
<i>N</i>	129	129	129	129
<i>Means</i>	2.34	2.58	23.46	1.39
<i>Standard Deviations</i>	0.59	0.56	7.00	0.49

Statistical Significance: * $p < .05$; ** $p < .001$

Table 12 shows that two predictor variables were entered in the first block of the model: LAN and SUPER3. This block explained 3.8% of the total variance in Grade 6 achievement and was not statistically significant, $F(2, 126) = 2.56; p = .14$. After the entry of the additional variable of PROGRAM into the second block of the model, the total variance explained by the combined IVs from both blocks was 7.8%, where the predictive model as a whole was also not statistically significant, $F(3, 125) = 3.23; p = .06$. The introduction of the PROGRAM variable into this model explained an additional 4% of unique variance in Grade 6 achievement, where R^2 Change = .04; $F(1, 125) = 5.00, p = .06$, or not statistically significant. Table 12, also demonstrates that the standardized regression coefficients for the contributions of the PROGRAM; SUPER3; and LAN variables were not statistically significant in the second regression model ($\beta = .228; \beta = .138; \text{ and } \beta = .134$, respectively). Thus, in the final adjusted model, none of the three predictor variables included in the model made a statistically significant contribution to the variance in Grade 6 test scores.

Table 12

Regression Model of Grade 6 Achievement Accounting for Level of Academic Need

	<u>R</u>	<u>R²</u>	<u>R² Change</u>	<u>B</u>	<u>SE</u>	<u>β</u>	<u>t</u>
Block 1	.19	.04					
SUPER3				.20	.10	.19	1.96
LAN				.00	.01	.04	.40
Block 2	.28	.08	.04				

SUPER3	.15	.11	.14	1.39
LAN	.01	.01	.14	1.33
PROGRAM	.28	.15	.23	1.88

In sum, descriptive statistics from the sample's demographic characteristics indicated that the vast majority of students were placed in the REGULAR group, with males comprising two thirds of the overall sample. The analysis of group differences on the SES variable showed that students in the LDP group had significantly lower SES scores than students in the REGULAR group, but not the WAITLIST group. The analysis of group differences for the LAN variable indicated that students in the LDP group were also significantly higher in academic need than students in the WAITLIST group. Results from first step of the hierarchical multiple regression analysis suggested that the inclusion model had a positive impact on achievement among Grade 6 students with LD, even after controlling for the influence of prior achievement in Grade 3. The first regression model also demonstrated that the sociodemographic variables of SES and SEX did not have a statistically significant influence on the criterion variable; therefore these variables were not included in the second regression model. Results from the follow-up analysis did not yield statistically significant results for any of the three predictor variables in the final adjusted model – including program placement. The implications of these findings are discussed in the next chapter.

Chapter Five: Discussion

The purpose of this study was to examine the impact of program placement on achievement for students with LD by developing a comprehensive research methodology that controlled for the influence of several ecological variables known to predict achievement in this group of students. As such, the academic trajectories of three groups of students with LD were analysed from Grade 3 to Grade 6: i) those in a regular class placement; ii) those in a specialized LD class placement; and iii) those waitlisted for the specialized LD class. In examining achievement for these three groups of students, four IVs were controlled for that are known to influence achievement test scores for students with LD: i) prior achievement; ii) student sex; iii) SES; and iv) the student's LAN. The study's DV was measured using the province of Ontario's provincial assessments that were developed by the EQAO. These tests were standardized, curriculum-based, criterion-referenced assessments administered annually in the primary and junior levels in all publicly funded elementary schools in Ontario.

The analysis was based on secondary data; therefore, the data file was screened in advance for suitability using multivariate statistical techniques. The missing values analysis uncovered a significant pattern of missing data at Grade 3 that was considered MAR for the purpose of employing the MI method. The resulting multiply imputed data file included original data for several variables, as well as imputed values for two newly created composite variables suitable for the subsequent multivariate analysis. The analytical plan was conducted over two steps. The selection of variables, and their order of entry into the hierarchical regression model for each step, was made based on theoretical and/or statistical considerations.

Students in the waitlist group were combined with students in the regular group before conducting the first step of the analysis. This was done in order to reflect the fact that students in

the waitlist group received their instruction in a regular classroom placement. With this in mind, the first step of the analysis sought to determine whether or not sociodemographic factors played a role in predicting Grade 6 performance over and above the main school-level factor of program placement. To explore this hypothesis, the variables of SEX, SES were entered in the first block of the regression model in order to examine their unique contribution in predicting academic outcomes beyond program placement. Then, in accordance with their theoretical fit in the research design, the SUPER3 and program placement (PLACE) variables were entered in successive blocks into the model. In this way, the IV of program placement was specifically explored for its unique contribution to the prediction model.

In the follow-up analysis, the specific influence of the LAN variable was examined for its unique contribution to the variance in the DV. As such, only students in the specialized and waitlist groups were included in the second regression model. Again, given that only the students from these two groups were assigned a LAN score, students in the regular group were not included in the follow-up analysis. These decisions regarding which variables to include in each analytical step, and their order of entry into either regression model, formed the theoretical basis for conducting two separate hierarchical multiple regression analyses. Ultimately, the purpose of the two-step analytical plan was to explore the differential impact of program placement on achievement for students with LD by employing a research design which took into account the LAN variable in each analytical step. The first step of the analysis combined students from the waitlist and regular groups before measuring the effect of program placement, while the follow-up analysis focused on matched comparison groups of students in an inclusive (waitlist) or specialized program placement.

There were four main findings from the study. First, the influence of the program placement variable depended on which students were included in each step of the analysis. The first regression model revealed that program placement provided a small but unique contribution to the variance in test scores for Grade 6 students with LD; however, this level of analysis included the students in the regular program who had a milder learning disability. The follow-up analysis included only the students in the waitlist and specialized groups who had a higher LAN, and showed that the program placement variable was not a significant predictor of achievement. The second main finding was that the influence of the prior achievement variable also depended on which group of students was involved in the analysis. In the first regression model, prior achievement was the largest statistically significant predictor of achievement in Grade 6 students with LD, regardless of whether or not the students were placed in a regular or specialized program. Similar to the program placement finding, the influence of this variable disappeared in the follow-up analysis. Third, results from the first regression model showed that neither of the sociodemographic variables (SES or SEX) included in the predication model had a statistically significant influence on achievement for Grade 6 students with LD, after controlling for program placement type. Finally, results from the follow-up analysis showed that the LAN variable was not a significant predictor of achievement for students placed either in the specialized or the waitlist group.

Interpretations of these findings are discussed below within the context of the study's theoretical framework (Bronfenbrenner & Morris, 2006), and three main research questions. Implications of the results are situated within the wider academic base and discussed in relation to both provincial and district level school leaders in Ontario who are charged with implementing programs and services for students with LD. Limitations and delimitations of the study are also

presented along with several methodological considerations for future research directions examining the nuanced influence of program placement and achievement for students with LD. Finally, concluding remarks based on the study's results are provided.

Main Findings

Nuanced influence of program placement. Despite several decades of research, the extant studies examining the impact of program placement on the academic achievement of students with LD can be described as equivocal at best. Due mainly to methodological limitations in the research designs themselves, the question remains as to whether or not an inclusive or general class placement, as compared to specialized class placement, will improve the academic outcomes for these students. This research topic was addressed in the current study not only in a more methodologically rigorous way through improvements to the study's research design, but also from an ecological perspective that accounted for the influence of several variables known to influence achievement for students with LD – variables which must be taken into account when considering the question: Does inclusion work for students with LD?

One of the intriguing findings from the first step of the analysis was that the program placement variable itself provided a small but unique contribution to the prediction model favouring the regular program. In fact, when students from the waitlist group were combined with those in the regular class during the first analytical step, the program placement variable added nearly 5% to the prediction of the variance in Grade 6 test scores. This unique contribution of program placement was over and above the influence associated with prior achievement in Grade 3, the only other variable which made a statistically significant contribution to the prediction model. These results suggest that students with LD placed in a regular program for instruction outperform students placed in a specialized class on standardized tests of

achievement, even after controlling for the influence of the other variables in the model. Thus, results from the first step in the analysis seemed to indicate that the inclusion model benefited students more than the specialized class model.

However, the program placement findings from the first analytical step did not hold in the follow-up analysis concentrated strictly on the students in the waitlist and specialized program groups. Using the metaphor of a microscope, the aperture of the lens was more focused on the academically ‘needy’ students during the follow-up analysis, and under this lens, program placement did not predict achievement. While it is true that the program placement variable was significantly correlated with the LAN variable, it did not make a statistically significant contribution to Grade 6 test scores in the follow-up analysis. In fact, it is likely that these two variables were highly correlated because they both reflected the difficulties experienced by the more severely LD students. In other words, higher LAN scores were correlated with a specialized program placement due to the fact that the more academically needy students were placed in the specialized program. Tabachnick and Fidell (2007) highlight this nuanced relationship in the correlation matrix between IVs and their ultimate contribution (or in this case non-contribution) to the prediction model. One variable may be significantly correlated to another in the correlation matrix but still make no statistically significant contribution to the prediction model in terms of its standardized regression coefficient.

One possible explanation for the lack of effect of the program placement variable in the second analysis may have been due to lower power. The technical definition of power is the probability of detecting a “true” effect when it exists (Creswell 2009a). A Type I error occurs when the researcher finds a statistically significant difference between groups when one does not exist, whereas a Type II error occurs when a difference actually does exist, but the statistical test

does not identify it. The question here is: Did the second regression analysis miss the program placement effect - was there a Type II error? In hierarchical multiple regression the researcher tries to avoid a Type II error through assumption testing before running the analysis. The applicable assumption respecting power analysis in this study would be the ratio of cases to IV assumption. As was noted earlier, the ratio to cases assumption was met in the second regression model. Therefore, the assumption testing process itself provides evidence that the lack of effect in the second analysis was not due to lower power.

What then do the different program placement findings from each analytical step mean? Other studies in the area suggest that a regular classroom instructional experience may be superior to non-inclusive service delivery models for students with LD (Andrews & Lupart, 2000; Bennett, 2009; Cole et al., 2004; Kalambouka et al., 2007). However, as was noted earlier, the LAN variable was not properly considered in these analyses. Results from the current study demonstrate that the program placement variable might not be the most critical factor in predicting achievement, as some authors have contended (Hocutt; 1996; Zigmond, 2003), yet this variable clearly requires closer scrutiny in order to understand its nuanced relationship with other variables in the prediction model. With this in mind it is important to consider more carefully what appeared in the microscope's lens during each step of the analysis.

The participants in this study were examined by looking back at two specific points along their academic trajectory. In this light, two separate slides were prepared for the stage of the microscope. At Grade 3, all students in the study received their instruction in the regular class. By Grade 6 these same students were all formally identified with LD, and one small sub-group was placed in a specialized program for instruction. The remainder of these students had a milder LD (a large majority) and were placed in a regular class for instruction. A third sub-group of

these students diagnosed with severe LD were waiting for placement in a specialized class, but still received their instruction in the regular program.

To illustrate, let us consider the experience of four different students with LD as they progress through elementary school using the two-step analytical plan outlined in this study. The first student is a boy called, Everett, who had an extremely high LAN, and received his instruction in a specialized program at some point during his junior years (i.e., Grades 4-6). Then, according to the analytical plan for step one, we compare Everett's performance on the Grade 6 EQAO test with another female student, Lily, who was formally identified with LD, but never met the criteria for a specialized program; therefore she received her instruction in a regular class throughout her elementary schooling.

Based on the findings from step one of the analysis, results would likely reveal that Lily, whose regular class experience had no real impact on her academic trajectory from Grade 3 to Grade 6, would outperform Everett on standardized measures of achievement. Results from the differences in group means on the dependent variable bore out this finding, where students in the regular group outperformed students in both the specialized and waitlist groups. Obviously, the root of Lily's superior performance on EQAO was that her LD was not as severe as Everett's, and therefore, she had an unfair advantage on the testing. In other words, Everett's LAN was not taken into account in the comparison. This is the typical level of analysis used in many inclusive education studies to suggest that the inclusion model is superior to a specialized program model. This method was essentially replicated in the first step of this study.

Contrast this method with exploring two other boys' academic trajectories, Diesel and Abraham, who both had a higher level of academic need as measured by the LAN variable. Diesel received his junior instruction in the specialized class. However, Abraham was placed on

the waitlist for this program because his parents felt that the LD program location was hosted by a school too far from their home. Thus, Abraham received his instruction in a regular class for Grades 4-6. All three boys, Everett, Diesel and Abraham would have likely struggled in the primary grades while learning to read and write and perform basic math skills. They would have also probably begun to experience certain socio-emotional problems directly related to their LD, such as test anxiety. Lily would have experienced similar academic difficulties but less so when compared with the boys - recall that the sex variable did not have a significant influence in the first regression model. Nevertheless, by the time they reached Grade 6 all four students would have been formally identified with LD, and placed in either an inclusive or specialized program setting.

Unlike the first comparison between Everett and Lily, results from the follow-up analysis showed that the performance on EQAO for students like Diesel and Abraham was not influenced by their program placement. Based on the findings, both boys would have performed equally poorly on the EQAO tests in both grades, regardless of whether they were placed in a regular or a specialized program. What is compelling about this finding is that Abraham was taught in the regular class, and he performed as well as Diesel who we might consider had the advantage of a specialized class. After all, the hallmarks of the specialized program are a lower student-teacher ratio, and the benefit of expert instructional support specifically designed to address the needs of the LD learner. Given this perceived advantage in instructional support, and all things being equal in terms of LAN, would we not anticipate Diesel to outperform Abraham on the EQAO testing? Clearly, this was not the case. When the analysis involved more closely matched groups of students the specialized class provided no advantage in academic terms.

This illustration serves to demonstrate how the higher resolution of the microscope's lens in the follow-up analysis helps to clarify the effect of program placement by properly accounting for the LAN variable. In other words, Abraham and Diesel were a better match for comparison purposes than were Everett and Lily to assess the influence of program placement on achievement. These results suggest that without including the LAN variable in the first step of the analysis, the placement variable - which was significant - was likely only masking the effect of the LAN variable on the test scores. Indeed, the rationale for designing the study along a two-step procedure was so that the blurriness of what appeared in the microscope's lens in step one (the obscurity of LAN) could become clearer in the follow-up analysis. This critical step revealed a less confounded relationship between the program placement variable and the DV. This finding directly addresses the main criticism of existing research which for many reasons has not adequately considered the influence of the LAN variable when making achievement comparisons across program placement type.

In sum, the relative importance of the program placement variable amongst a group of IVs in predicting achievement depended on the students' level of academic need involved in each step of the analysis. When the regression model combined students from the regular and waitlist groups in the first step, the program placement variable was a small but significant predictor of scholastic achievement. When the follow-up analysis did not include the students in the regular class, and focused only on students with higher academic need, the effect of program placement on academic achievement disappeared. In any case, the story regarding the effect of program placement was similar to another important variable in the analytical plan, the individual student's prior achievement in Grade 3.

Influence of prior achievement. The second main finding of the study was that prior achievement in Grade 3 was a significant predictor of achievement for students with LD; however, this finding was true only in the first analytical step when students from the regular program were included in the regression model. Similar to the program placement results, the influence of the prior achievement variable on later academic success depended on which groups of students were involved in each step of analysis. The prior achievement variable is important to discuss given that one of the study's main research questions sought to follow the academic trajectory of students throughout elementary school. The rationale for this methodological aspect of the study was to account for the unique variance in test scores attributable to the individual student across different program placements types over time. Methodologically speaking, this was one of the study's strengths. From a theoretical lens, the prior achievement variable's inclusion in the conceptual research model was also seen as important for understanding the influence of individual-level factors combined with system-level factors (i.e., program placement), creating a meso-systemic problem (Bronfenbrenner & Morris, 2006).

Results showed that the contribution of the prior achievement variable to the overall prediction model was much stronger in the first analytical step than in the follow-up analysis, as measured by the statistically significant standardized regression coefficient for this variable in the first regression model. This finding suggests that performance on the primary EQAO assessment was a more powerful predictor of scholastic success for the less academically needy students with LD receiving instruction in the regular classroom. Stated another way, previous academic history had the greatest influence for students with a less severe LD. Recall that the analysis in step one included the vast majority of students who had a milder form of LD and who were instructed in the regular program. Using the previous illustration, Lily's performance on the

Grade 3 EQAO assessment turned out to be a good predictor for her Grade 6 EQAO scores. This finding was obviated by the statistically significant and highly correlated test scores in both grades for students in the regular group.

Nevertheless, similar to the program placement results, the influence of prior achievement did not hold in the follow-up analysis when the aperture of the microscope's lens was more finely narrowed on the specialized and waitlist groups. At this level of granularity, prior achievement in Grade 3 was not predictive of later performance in Grade 6, regardless of program placement type. To be sure, the correlation between Grade 3 and Grade 6 scores in the follow-up analysis was also statistically significant; however, the r value was much larger in the first step than in the second (i.e., 0.46 vs. 0.19). This finding suggests that individual student factors such as prior achievement were no less important for predicting academic success than school-level factors (e.g., program placement) for students with higher academic need. In fact neither variable was a statistically significant predictor of academic success in Grade 6 in the follow-up analysis.

This finding is surprising considering that the relationship between prior achievement and later academic success has been established by previous research in the area. For example, Rogers et al., (2006b) identified the importance of prior performance, along with several other factors, in predicting later achievement in Grade 6 students using large-scale assessment data from Alberta. Results from the Rogers (2006b) study demonstrated that achievement patterns were stable over time for Grade 6 students in the general population as well as for students with mild/moderate disabling conditions. However, the study did not include the variable of program placement in partialling-out the individual and school factors influencing achievement. Nevertheless, stable achievement patterns from Grades 3 to 6 were also evident in the current

study when the large population of students with mild LD were included in the first step of the analysis. What about the follow-up analysis? Why was there a lack of stability in achievement over time for the more academically needy students with LD?

One possible explanation for these particular findings might be related to the nature of the LD and the types of accommodations which are typically used to facilitate participation in large, scale provincial assessments. It is generally understood that the psychological profiles of students with more severe LD present a much different clinical picture than students with milder learning difficulties. For example, deficits in certain cognitive processes such as working memory are more pronounced for students with serious learning difficulties, and this impairment would make performance on the EQAO tests much more challenging for students with this psychological profile. This is particularly true in Grade 6 when considering the nature of language-based assessments, like for example reading, which becomes more sophisticated as children develop cognitively. For instance, in the primary grades students are 'learning to read', whereas in junior grades, they are 'reading to learn'. Therefore, the types of cognitive processes being accessed for academic success in Grade 3 are less complex than in Grade 6. It is not difficult to imagine how cognitive impairment to a Grade 6 student's working memory would negatively impact a narrative recall task on the EQAO tests -- even with accommodations.

Typically, participation on the provincial assessments for students with moderate learning difficulties can be accommodated through the use of assistive technology which facilitates access to the assessments through aides to response format. For example, scribing test questions or having test questions read to the student through text to speech software. For students in the regular group these types of assistive technology accommodations would have enabled them to access the test more efficiently and helped to bridge the gap for their performance. This effect

was evident in the outcomes for students in the regular group whose EQAO scores were stable over time. Perhaps these types of accommodations were simply insufficient to overcome the more serious cognitive deficits which distinguish the students with higher LAN in the waitlist and specialized programs? These impairments would be particularly acute by the time these students reach Grade 6, and exacerbated by the nature of the assessments which become more cognitively challenging. It is a classic double-crush situation.

In any case, the results from both analytical steps respecting the prior achievement variable underscore the significance of the study's repeated-measures research design, which has been reported by several researchers as a better methodological approach for examining the impact of program placement on achievement in students with LD (Rogers et al., 2006b; Kauffman et al., 2008; Marston, 1996; Zigmond, 2003). Previous studies have shown that the repeated-measures designs can be a powerful methodological tool for separating the variance in outcome variables associated with the individual. However, there have been no reports to date using the repeated-measures design in order to examine the academic trajectories of students with LD in inclusive vs. non-inclusive settings. Thus, one of the current study's strengths was that achievement was measured within the same students over time and in inclusive vs. non-inclusive program settings. This important feature of the research design reduced the overall variability in the outcome data lending credibility to the study's methodology (Creswell, 2009a). A similar rationale was conceived respecting the utility of the LAN variable.

(Non) Contribution of LAN. Another important goal of the study was to employ a quasi-experimental research design that could reasonably measure achievement outcomes for students with LD across different program placement conditions, while accounting for the psychological profile of the students in each educational setting. The LAN score was

instrumental for this purpose. From a methodological perspective the contrast in the follow-up analysis focusing on students in the waitlist and specialized program groups addressed the study's main research question regarding program efficacy. The theoretical construct of LAN was also central to the study's conceptual framework which simultaneously assessed the influence of other important ecological factors at the individual and school-levels (Bronfenbrenner & Morris, 2006). Ultimately, the results from the follow-up analysis present evidence for a better way to measure the impact of the inclusion model for students with LD by accounting for the LAN variable in the prediction model.

Over the last few decades, several authors have identified the influence of LD severity as important in terms of its ability to explain variability in test scores for students with LD; however, the absence of this particular variable in the research design of most inclusion studies is considered a methodological weakness by many authors (Bennett, 2009; Carlberg & Kavale, 1980; Hocutt, 1996; McLeskey & Waldron, 2011). Clearly, there must be some allowance for controlling for the individual student's LAN before determining whether or not program placement is a critical factor in influencing their achievement. For this reason, the second phase of the analysis assessed the unique contribution of program placement on outcomes for the students in the LDP and WAITLIST groups while controlling for the LAN of the students in each group.

For the purposes of this study, the LAN variable served as a proxy measure for the severity of a student's learning disability. All students applying to the LD specialized program were assigned a LAN score by the committee charged with making program placement decisions at the school district under study. Based on the student's LAN score, this committee decided which students were recommended for a specialized class placement, and which students were

placed on the waitlist for this intervention. The LAN score assigned to each student was a composite of both psychometric and academic data, as well as other ecological information. For example, if the student was diagnosed with an additional mental health condition, socio-emotional, or behaviour problem, these additional diagnoses would be reflected in the student's overall LAN score.

The LAN variable was gathered strictly for students in the LDP and WAITLIST groups as students in the REGULAR group did not apply for a specialized class placement. Students from both the LDP and WAITLIST groups were considered higher in LAN, or had a more severe LD, than students in the REGULAR group. Lack of available spaces in the specialized classroom setting meant that students in the WAITLIST group received instruction in the regular class setting. In this light, students in the LDP and WAITLIST groups were theoretically a better 'match' than students placed in the REGULAR group. Given that students in the WAITLIST group had higher LAN but were instructed in the regular classroom, they provided a 'natural' comparison group for measuring the impact of program placement on achievement. In sum, the LAN variable served as a quasi- experimental grouping mechanism, which by its nature as a continuous variable, was ideal for the follow-up hierarchical multiple regression analysis.

Examination of the mean LAN scores between the waitlist and specialized groups indicated statistically significant differences, favoring the specialized program. Based on these results, it is clear that the two groups were not equally matched on LAN after all. At first glance this finding did not bode well for the study's conceptual design; however, it is also important to recall that the committee charged with making placement recommendations in the school district under study used a rank-order procedure based on LAN in order to prioritize offers for entry into

the specialized program. This fact is important to consider when interpreting the follow-up analysis results because the neediest students were not always placed into a specialized program.

Results showed that variability in LAN scores was actually higher in the waitlist group compared to the specialized program, whereas the minimum and maximum values for each group were nearly equal. These particular findings suggest that at least some of the students in each group were similarly ranked in terms of LAN, but were placed into different educational settings for their instruction. Why is this? The reason might be that the student's LAN score was not the only factor that was used by the committee to make placement decisions. Other factors such as parental choice, disruption to peer group, or even proximity to the school hosting the program might have influenced the decision to recommend different placements for academically similar students. Consideration of these factors might also explain why the variability in LAN scores was sporadic across the program placement variable.

The waitlist group contained some students with higher academic need while the specialized program group contained some students with lower academic need. Recall from the illustration that Abraham's parents decided not to accept the specialized class offer even though his LAN score was as high as Diesel's, for reasons of proximity to the school hosting the program. Consequently, Abraham was instructed in the regular program. The influence of this variability within the two groups must be weighed against the statistically significant mean differences, or when considering the fact that the groups were not technically matched in terms of LAN. It is also important to keep in mind that the second regression model was a more focused analysis as it did not include students with a milder disability who were taught in the regular class. Instead, only the high academic need students were included in the grouping variable used to compare achievement outcomes across program placement. Although not an

ideal match, it can be argued the waitlist group was a better comparator than the regular group for students in the specialised program. In fact, the absence of a technically matched sample is not a serious constraint given that the LAN variable is statistically controlled for by the second regression analysis. It renders the limitation moot.

Results from the follow-up analysis are also important because they indicate that the LAN variable was significantly correlated to the program placement variable but only weakly correlated with Grade 6 tests scores. This fact is not surprising considering that the results from the between groups analysis showed significant differences in mean LAN scores favouring the specialized group. Higher LAN scores meant a student was more likely to be placed in the specialized program. However, the LAN variable itself did not make a unique contribution to variance in Grade 6 test scores in the follow-up analysis. This finding is important because it means that the level of academic need for students in both the LDP and WAITLIST groups did not significantly influence their achievement outcomes in either program placement. In other words, Abraham's LAN did not influence his EQAO scores in the regular class. Nor did Diesel's LAN affect his achievement in a specialized class.

Other researchers have found differences in reading disability in students with LD irrespective of the disorder's severity (Flannery et al., 2000); however, there have been no academic reports measuring the LD severity variable in relation to outcomes in different program placements. Unpacking this finding further means that if a student with higher LAN were placed in a regular classroom, it is reasonable to expect that they would achieve equally to a student with similar academic need placed in a specialized class. Conversely, a specialized placement for an academically needy student would be equally beneficial in achievement terms to a regular classroom placement. This is precisely the illustration that was provided earlier when Diesel and

Abraham performed similarly on the EQAO tests in both grades. The policy implications of this finding, as well as the results respecting the influence of two other sociodemographic variables, are discussed below.

The (non) influence of sociodemographic variables. There were two sociodemographic variables included in the study's research design, SES and student SEX. The sociodemographic variables were included in the conceptual model because they were grounded in previous research in the area, and were well-aligned with the theoretical framework, the bioecological model of human development (Bronfenbrenner & Morris, 2006). These particular variables, one germane to the individual student and the other a school-level factor, were juxtaposed against the other micro-systemic variables in the model (i.e., program placement) to create a meso-systemic problem. There have been no research studies to date that account for sociodemographic variables relative to the program placement variable and in terms of their relationship with academic outcomes for students with LD. Therefore, one of the study's research questions sought to determine whether or not these variables were important predictors of achievement for this group of students placed either in a specialized or regular program for instruction.

Socioeconomic status. The initial data screening procedure revealed that students in the specialized program group had significantly lower mean SES scores than students in either the regular or waitlist groups. This finding confirmed previous research in the area which suggests that students from low income neighborhoods are more likely to be placed in a specialized program setting than a regular classroom setting (Morgan et al., 2009; Skiba et al., 2005; Whitley, et al., 2007). Results from the first step of the analysis also indicated that the SES variable was not a significant predictor of academic achievement for students with LD, while controlling for prior achievement and program placement. This distinction is considerable, as

SES was significantly correlated with the achievement variables but was not a significant predictor after controlling for SUPER3 and PLACE. Although students in the specialized program setting had lower average SES scores, this fact did not significantly influence their academic achievement. Conversely, higher SES students with LD who received instruction in the regular classroom achieved no better than their low SES peers. These results are consistent with Vandenberg and Emery (2009) who found that the SES variable was not a significant predictor of change over time on standardized achievement tests for students with LD.

In any case, the results from the present study are significant because they suggest that the placement of students with lower SES backgrounds into a specialized setting itself will not influence their academic outcomes. Instead, being placed in a regular classroom, these students may be exposed to other factors which research has shown could provide indirect benefit to their learning or well-being. For example, several authors have demonstrated that students with LD display problems with social and emotional determinants which contribute to their overall difficulties in school (Al-Yagon & Mikulincer, 2004; Kavale & Mostert, 2004; Wiener & Tardif, 2004). These studies suggest that the inclusion experience may help to mitigate these socially-based issues which these students face in addition to their disability. Results from the current study complement these findings in the sense that program placement does not necessarily mitigate the effects of low SES on achievement. Thus, poverty by itself may not be a primary consideration in making program placement decisions for students with LD.

Further, one of the categories on the LAN rating scale specifies 'family issues' as a factor other than the student's LD diagnosis causing the academic delay (see Figure 5). To continue with the illustration above, if one of Diesel's 'family issues' was low SES, and this factor was deemed to be influencing his learning problems, he would have received a lower LAN rating,

and therefore, would have been less likely to be recommended for placement in a specialized class. This postulation is the central idea of the exclusionary criteria used by the Ministry of Education and the LDAC, where both definitions attempt to eliminate the effect of SES from the LD construct. For argument's sake let us assume that Abraham came from a wealthy family. We know that both he and Diesel had similar learning problems as they scored similarly on the LAN. By this logic, Abraham would have been as likely to be recommended for a place in a specialized program as Diesel. Yet he was waitlisted instead, due to his parents' choice not to send him to the LD school for proximity reasons. Perhaps, though, SES also had an indirect influence on the committee's decision to place Abraham on the waitlist? Based on the study's results, we know that the LD program location sites contain student populations with lower average SES. Perhaps Abraham's parents knew this fact tacitly, and did not want their son to attend a school site with lower SES.

All this begs the question: Why the significant mean differences in SES scores between the specialized and waitlist groups? It seems that despite the selection committee's best efforts to mitigate the influence of SES on LAN scores, this variable still clearly appeared in the overrepresentation of poorer children in the specialized program. Unfortunately, it is not known how the committee making placement decisions in the school district under study interpreted the 'family issues' factor that caused the academic delay. For example, was the factor being referred to in 'family issues' specifically SES, or was it referring to something else? In any case, the study's demographic findings provide strong evidence to suggest that SES is related to program placement decisions, where low SES students are overrepresented in the non-inclusive educational setting. However, the regression results do not support the notion that SES predicts academic achievement for students with LD in various program placements.

Student sex. Descriptive data from the SEX variable confirmed previous research in the area showing that males are over-represented in the general population of students with LD (Devine et al., 2013; Wheldall & Limbrick, 2010). This disparity was also reflected across program placement conditions with males representing approximately two thirds of the sample in all three groups. Similar to the SES variable, results from the first step of the regression analysis showed that the SEX variable did not make a statistically significant contribution to the variance in Grade 6 test scores for students with LD in different program placement conditions. In fact, the SEX variable was the only predictor in the first regression model that was not significantly correlated with the dependent variable. These results are not consistent with the substantial body of evidence regarding sex differences in language and math achievement favoring girls with LD (Berninger et al., 2008; Flannery et al., 2000; Wei et al., 2013; Wheldall & Limbrick, 2010).

Notably, the findings on student SEX do not necessarily mean that achievement for females was significantly higher or lower than for males. A simple regression of the SEX variable on achievement was not conducted because the conceptual model sought to understand the nuanced contribution of several ecological variables combined. In other words, the current study was exploring the predictive ability of the sociodemographic variables combined on achievement, after accounting for prior achievement in Grade 3 and program placement. Results in the current study simply mean that the influence of the SEX variable might have been absorbed by the other variables included in the first regression model. Indeed, the SUPER3 variable was likely accounting for most of the impact on Grade 6 achievement in the first analytical step.

To summarize, the interpretation of the influence of the sociodemographic variables in the current study requires a nuanced approach. Many researchers have established the

relationship between the student's sex and/or socioeconomic status (SES) and achievement for students with LD. Therefore, the influence of these particular sociodemographic variables was included in the conceptual model in order to shed light on any differences they may exert on this population across program placement conditions. Descriptive statistics from the study show that males comprised two thirds of the overall sample, and this disproportionality in SEX was clearly reflected across the three groups. Results from the analysis of group differences on the SES variable also showed that students in the LDP group had significantly lower SES than students in the REGULAR group, but not the WAITLIST group. These results confirmed previous research in the area.

Results from the first regression analysis also showed that the correlations amongst the IVs themselves and between the IVs and the DV were weak to moderately strong. Three of the four predictor variables, including SES were significantly correlated with the criterion variable of achievement (SUPER6). Therefore, it seems that although the effect of these variables did not result in statistically significant predictors in the regression model, their relationship with the DV may still require a more nuanced consideration which emphasizes their moderately strong correlations and descriptive information which confirms the results from other studies.

Implications

The educational implications of the findings from the first step of the analysis for the SEX variable centre around the recent attention on the achievement disparity between boys and girls in Ontario's public school system and, in particular, literacy interventions that have been targeted at boys with LD (Council of Ministers of Education, Canada, 2018; Mullis, Martin, Kennedy, & Foy, 2007; Sokal, Thiem, Crampton & Katz, 2009). For example, from 2005-2008 the Ontario Ministry of Education engaged in a province-wide inquiry into the instructional and

assessment practices used by educators for improving boys' literacy achievement called, *The Road Ahead, Boys' Literacy Teacher Inquiry Project*. The final report of this inquiry presented several recommendations and resources designed to help teachers make a positive impact on the learning environment and reading experience of boys (Ontario Ministry of Education, 2009). Results from this study have clear implications for the continuation of such efforts specifically for students with LD. For instance, if we can expect no difference in achievement between male and female students with LD, regardless of program placement, then what justification is there for specific, gender-based literacy interventions for these students?

The implications of the SES findings are also clear respecting service delivery supports for students with LD. The new knowledge generated by the study in this area may be valuable for both district and provincial level policy-makers insofar as it can be used to refine the allocation procedures used for distributing financial or other human resources to low SES schools. Indeed, the index used for creating the SES variable was developed by the school district under study in order to identify schools with lower SES, ostensibly so that additional financial and human resources could be targeted at schools in most need. The fact that a disproportionate number of students with LD, having low SES backgrounds, are situated in specialized classes would justify the allocation of additional human resources, or other monetary investments at school sites offering a specialized LD program. Thus, policy-makers at the district and provincial level might use the novel information from this study to shape policy for addressing high needs populations located in specific schools, particularly those schools hosting specialized LD programs.

The main implication of the study's program placement findings for policy-makers at the provincial and district level relates to the provision of service delivery options for students with

LD. For example, results from the follow-up analysis could be construed as evidence that program placement does not influence academic outcomes for the largest group of students with special needs in the province. This argument has ramifications for the continuum of service delivery model for special education espoused by most districts across the province of Ontario. One feature of the continuum model is the specialized class which is typically reserved for students with the most severe LD diagnosis. Based on the study's results, the argument could be set forth that a specialized program placement will not yield positive educative gains for students with LD. This position contradicts the persistent notion amongst certain highly politicized parent and/or educator communities that a specialized class, featuring lower, more expensive staff-student ratios, will improve the academic outcomes for the most severely learning impaired students.

The vast majority of students with LD in Ontario schools receive their instruction in the regular classroom (Bennett, 2009; Ministry of Education, 2017). In fact, very few school districts in Ontario offer a specialized program placement option for students with LD, and of those that do, only a small subset of the total number of students identified with LD in any given district would receive such a placement. The student-teacher ratio for specialized LD programs is calculated according to provincial standards so that the ratio is currently set at 8:1, and the amount of instruction delivered in the specialized class ranges from at least 50 percent of the school day (Ontario Ministry of Education, 2017). These specific policy parameters make specialized classes a more expensive programming option for school districts to maintain due to staffing costs. Consequently, special class placement recommendations are typically reserved for students with the most severe academic need. A relatively recent philosophical and policy shift supporting more inclusive placements for all students with special education needs has also

resulted in fewer students being educated in segregated settings across the province (Bennett, 2009; Hutchinson & Martin, 2012; Ontario Ministry of Education, 2006). Thus, it is not surprising that a disproportionate number of students with LD was reflected across the sample's three placement types.

In any case, school districts in Ontario continually struggle with the logistical and cost ramifications of the provincial government's policy for maintaining low staffing ratios for specialized programs. From a financial perspective alone, the argument could be very compelling that precious resources currently being spent on maintaining low staffing ratios for specialized programs could be reallocated to hiring more educators to meet the needs of students with LD situated in the regular classroom. Evidence from this study could be used to support this argument on the premise that the more expensive, lower ratio specialized program had no influence on achievement. Clearly, the results from the follow-up analysis are important to consider in terms of policy implications for service delivery options.

Another policy implication of the program placement results from the study is related to the perceived benefits of inclusive educational settings for students with LD. This perspective is unpopular amongst Ontario's organized education labour unions whose members have long-decried the government's quasi-policy for inclusion, *Learning for All* (2013), as an encroachment on the continuum of service model tacitly targeting the elimination of special classes. Advocates for the preservation of specialized classes for students with LD might consider the results from the first analytical step as defunct because the regression model did not include the LAN variable. These activists rightly argue that this level of analysis ignores the fact that the high academic need students are placed in the specialized program while the less academically needy students receive their instruction in the regular program. This analysis represents the classic

‘apples-to-oranges’ comparison, which many advocates argue is an inadequate assessment of the inclusion model’s efficacy for students with LD. Technically, it is a biased comparison.

The new knowledge generated from this study respecting the non-contribution of the LAN variable to academic success in different program placements has possible implications for advocacy groups whose policy position does not support full inclusion for students with LD. For example, some advocates in Ontario’s education system oppose blanket policies mandating the arbitrary placement of all students with LD in any one educational setting, such as a regular class. Proponents from this perspective might argue that the specialized class option offers a ‘safe-haven’ for students with LD who suffer from socio-emotional deficits in addition to their academic difficulties. The contention here is that other factors beyond academic achievement are addressed by a special class placement. In this way, the results from the current study might be used by advocates to press the provincial government to maintain the continuum of program placement options in the most enabling instructional environment that support the various social, emotional and educational needs of students with LD.

From a philosophical perspective, the most enabling instructional environment under this policy umbrella could mean either a specialized or regular classroom placement. The findings from the follow-up analysis respecting the LAN variable suggest that the educational needs of more academically needy students are equally met in either instructional environment (inclusive vs. non-inclusive). This equivocality calls into question the moral imperative for school districts to provide a range of program placement options for students with LD. For example, what is the nature and purpose of specialized programming for students with LD? Do we expect the special class option to improve students’ academic achievement exclusively, or to provide a safe-haven for their socio-emotional and self-advocacy development? Thus, the study’s findings have

important philosophical implications for both advocacy groups and policy-makers alike with respect to the precise purpose of inclusive-oriented or specialized service-delivery models for students with LD.

The main philosophical implication of this study is additional support for the dismantling of the parallel systems view of special education. Radical or even moderate inclusionists in academia who challenge the operation of a parallel education system which marginalizes and discriminates against students with LD might find evidence in the study's results for a more inclusive education system (Fedler, 2019; Lindsay, 2003; Norwich, 2008). Full inclusion implies that all students, regardless of ability, will be educated in general education classrooms. The corollary of this position is that separate special education systems and their associated exclusionary practices of discrimination and marginalization would be abolished. The evidence from this study could be construed as supportive for combining all students with LD, regardless of academic need, into the regular classroom for instruction with their peers. After all, if one cannot expect any real scholastic gain from segregating students based mainly on academic criteria, then why subject these children to the negative stereotypes that often accompany the 'special class' moniker?

Limitations/Delimitations

Limitations inherent in the secondary data file. Limitations in social science research generally refer to the inclusion or exclusion criteria of variables in the research design, which for whatever reason, the researcher cannot control. There are three main inter-related limitations for the study which are rooted in the secondary nature of the data file. First, in order to be included in the secondary data file, the student required a formal diagnosis of LD from a registered psychologist in the province of Ontario. However, the specific date of the diagnosis was not

indicated in the data file. This means that the formal diagnosis could have been made at any point from when the student was registered in Grade 3 (at age 8yrs old), up to when they were registered in Grade 6 (at age 11yrs old). This time period spans several major developmental milestones for children in terms of their growth in the areas of cognition, or other social-emotional factors (Shaffer, 1993). Even with knowledge of the specific diagnosis date one cannot infer that the student was not suffering from the symptoms of their disability prior to being formally diagnosed. The time-lag between these two points (onset and diagnosis) is a murky area that has obvious service-delivery implications for students receiving instruction in a regular classroom setting. Concomitantly, the lack of knowledge regarding the date of diagnosis is an important limitation of the current study.

The variability in the psychological profiles of the students with LD included in the secondary data file was also unknown. It is generally understood that in psychological terms students with LD do not present with identical profiles, and thus, they can manifest a wide range of clinical characteristics ranging from cognitive processing delays to social-emotional deficits. The implication of this unknown variability in the nature of LD as a psychological construct is that the students may not have been matched across the placement conditions. At least some of the psychological dimensions of each student were summarized by the LAN scale (e.g., WISC-IV sores); however, the LAN variable itself did not provide the full clinical picture of the students included in the study. Therefore, critical information respecting the psychological profile of the individual students with LD was missing from the secondary data file. This shortfall in psychological information about the study's participants represents a second important limitation of the study.

A third related, and perhaps most important, limitation is that the precise period of time that students spent within an inclusive or non-inclusive educational setting was unavailable from the secondary data file. Program placement information was obtained from the district's student database at two specific, fixed in time points (i.e., Grade 6 and Grade 3). Thus, more clarity on placement is needed because all three groups were determined based on fixed program placements in both Grade 3 and 6. All three groups of students were in the regular program in Grade 3, while only students registered in the LD program at Grade 6 were considered for the specialized group. Given the possibilities, a student in the regular group might have actually spent more time in a specialized program (in Grades 4 and 5) than a student in the specialized group who may have only been in a specialized program for Grade 6. This situation is unlikely; however, such information about program placement would help to better interpret the findings, particularly those respecting program placement in the follow up analysis.

In a similar vein, students in the specialized group may have also spent equal amounts of time within the segregated setting. In fact, some of these students may have received the specialized program for up to three years, while others experienced only one year in a specialized setting. Nor can it be inferred that students in the regular group spent no time in a specialized class setting during their academic trajectory through elementary school. Although it would be rare for a student to transfer out of the specialized setting after receiving a placement offer, it is conceivable that some students in the regular program in Grade 6 may have received a specialized placement when they were in either Grade 4 or Grade 5. In short, the lack of knowledge regarding the treatment length in either the inclusive or non-inclusive settings is an important, related limitation of the study.

The fact that some of the study's participants may have had a formal LD diagnosis for a longer period than others, or that they may have spent more time in a specialized class, has definite implications regarding access to services in terms of remediation and support at the school and district levels. For example, the smaller special classes would mean that students in that setting received more personalized instruction from a special education teacher who was specifically trained to remediate the educative and social-emotional needs of children with the LD exceptionality. Students in the specialized program would have also received more indirect support from knowledgeable, school-based special education staff such as the school resource teacher or educational assistant(s). This additional school-level support may have been in the form of assistive technology training designed to help students with LD access the curriculum. Finally, students in the segregated setting would have received indirect district level support from expert special education staff, such as a psychologist or a master teaching consultant. Given the nature of tiered intervention approach espoused by school districts in Ontario, students with LD placed in a regular class simply would have not had access to this level of support.

The variability in both service and support across placement conditions has implications respecting the educational experience of these children over the three or so years between the snapshots of achievement represented by the study's outcome variables. With this limitation in mind while interpreting results, it could be argued that the non-significant achievement scores between the specialized and waitlist groups, after controlling for LAN, provides some evidence for the superior performance of the inclusion model. Students who received less support performed equally to those who received more support. In sum, three inter-related factors represent limitations of the study's research design using secondary data which must be considered in interpreting the results.

Selection bias between groups. Selection bias is introduced in a sample when proper randomization is not achieved, thereby risking that the sample obtained is not representative of its intended population (Creswell, 2009a). One of the study's research design strengths was that it considered a matched comparison group of students placed in inclusive and non-inclusive settings on LAN (i.e., waitlist vs. specialized) when comparing the groups on achievement. The inclusion of the LAN variable in the regression equation was seen as a methodological improvement over other studies in the area; however, as was noted earlier, the committee charged with making recommendations for the specialized program in the school district under study used a rank-order procedure based on the LAN scores in order to prioritize placement offers for entry into the program. Students with the highest LAN scores received offers first and those who did not get an offer were placed on the waitlist. This fact might represent a confounding bias in group composition affecting the validity of the study's findings.

Thus, it is possible that the selection bias built into the ranking procedure for program placement, as measured by the LAN, confounded the 'natural' matched sample in the quasi-experimental design. This bias was revealed by the results in the initial data screening procedure which found a statistically significant difference in the mean LAN scores favoring the non-inclusive setting. In other words, the waitlist and specialized groups were not equally matched in terms of LAN as was originally hypothesized. The difficulty in identifying a matching variable for LD severity which can serve effectively in a research model for examining program efficacy in inclusive vs. non-inclusive settings has been noted by several authors (Baker, et al., 1995; McLeskey and Waldron, 2011; Zigmond, 2003). Despite the attempt to address this problem using the LAN variable, the selection bias built into the group composition represents a possible methodological limitation of the study.

This so-called limitation had no real consequences on the results because LAN was statistically controlled for in the regression model. In other words, even though the groups were not precisely matched on academic need, they were still equated by controlling for LAN in the hierarchical regression. However, other forms of bias may have affected selection in the groups' composition. One possibility is that there may have existed in either parents or educators an *a priori* belief that either the specialized or the regular program placement was better for their child's particular learning or socio-emotional needs, and these beliefs may have blocked the nomination of certain students for either program at the school level. Thus, differences in the nomination 'path' that preceded the district-level committee's LAN ranking procedure may have been due, in part, to a self-fulfilling prophecy in group selection which biased the students who were nominated for either program. Indeed, parent and/or educator beliefs about the efficacy of inclusive education may be at the root of why the parallel systems model for special education delivery persists in most school Districts in Ontario today.

Assigning school level SES to individual students. This study was designed to avoid another potential problem respecting the misinterpretation of SES data, referred to as the ecological fallacy (Sirin, 2005). An example of an ecological fallacy would occur where individual-level inferences are made on the basis of group aggregated data. An endeavour was made to mitigate the indirect effects of the ecological fallacy by assigning a SES score from the home school of students in the specialized group who did not actually live in the LDP school's neighbourhood, but rather, were transported there in order to attend the program. In this way, at least the SES score assigned to the individual student represented their neighborhood school and not the school which hosted the LD program.

Nevertheless, the SES score was still an average, school-level score assigned at the individual student level. The problem with this procedure echoes the methodological limitations other researchers have identified in the operationalization of the SES variable and in constructing and applying an overall SES rating to individual students (Skiba et al., 2005; Vandenberg & Emery 2009). Indeed, Sirin (2005) warns researchers about the methodological complexities related to the aggregation of different combinations of SES-related variables and applying these to a single measure of SES. In other words, the application of the SES index in the current study does not account for the variation of individual scores within a school. For example, two families might have the same school SES score where one family contains highly educated immigrants working in low-earning employment, while the other family is simply struggling with the common symptoms of poverty, low stability or high transience. The current study attempted to mitigate the effects of the ecological fallacy; however, applying a group level SES score to individual students remains a potential methodological limitation of the study.

Further, the 'readiness to learn' variable contained in the SES index itself may also present a possible confound with the achievement data that was used as the DV in the study. The readiness to learn variable was developed on the basis of the six learning skills and work habits contained in the Ontario elementary report card: i) responsibility; ii) organization; iii) independent work; iv) initiative; v) collaboration; and vi) self-regulation. These readiness to learn factors are not actually achievement measures but are nonetheless related to scholastic success. For example, if a student scores highly on the six learning skills and work habits, they are more likely to achieve academic success. This presumed relationship between the readiness to learn factors and scholastic achievement is the rationale for including them in the elementary

provincial report card. For this reason, the readiness to learn variable contained in the SES index may represent a confounding variable for the achievement results.

Criterion-referenced tests as adequate measures of achievement. The limitations of large-scale, criterion-referenced tests such as the EQAO primary and junior provincial assessments have been documented by several researchers in the past (Kohn, 2001; Nichols et al., 2006; Rogers, 2014; Volante, 2006; Wolfe et al., 2004). While critics of the EQAO assessments emphasize the misuse of results for purposes beyond that which they were intended, the main limitation of their use in the current study lies in determining their adequacy as measures of academic achievement for students with LD. The EQAO is transparent about the fact that both the primary and junior assessments are a snap-shot of student achievement at a specific point in time, and therefore, should not be considered as wholly indicative of a given student's mastery of the Ontario curriculum (EQAO, 2019).

In fact, the focus of the EQAO tests is on assessing a student's explicit knowledge of specific curriculum strands, and this approach has been criticized as overly structured and not comprehensively reflective of either the Grade 3 or Grade 6 curriculum (Volante, 2006). Critics argue that as a function of their format, the EQAO tests are only a narrow measure of those areas of the curriculum which lend themselves to paper and pencil testing. Further, the absence of a revision process built into the assessment procedure itself, which is properly supported by advice from teachers, also denies students the opportunity for this critical and typical classroom assessment practice. The result is an artificial and stressful test environment for certain students, particularly those with special education needs. This argument is relevant for the current study because students with LD tend to manifest their academic delay in the very curriculum areas which the EQAO purports to measure - knowledge and/or skills in reading, writing, and math.

Therefore, the general limitations of the primary and junior EQAO assessments as adequate measures of achievement must be weighed against their validity as appropriate outcome measures for the purposes of the present study.

Epistemic delimitations of the research design. Despite the prevalence of several epistemological approaches in educational research, most theorists agree that empiricism and constructivism are the two major inquiry paradigms (e.g., Denzin, 2008; Usher, 1996). The current study employed a quasi-experimental, correlational research design examining the relative influence of several IVs on a criterion variable across different treatment conditions. In this respect it was neither a purely empirical study interested in determining causality, nor an exploratory constructivist research design with an aim to deepen insight into the educational experience of students with LD. Due to the fact that the epistemological approach of the study could be best described as post-positivist, there are certain philosophical delimitations which must be considered regarding the conclusions which can be drawn from the results.

The study was foremost delimited by its quasi-empiricist focus on outcomes, and for this reason, it did not address the actual program that was delivered in either the inclusive or non-inclusive settings. In other words, the treatment effect of the program itself was a major delimitation of the study. The absence of any information regarding the program activity architecture operating within either the inclusive or non-inclusive educational contexts represents a vast territory of unknown variables. Furthermore, many variables germane to the individual participants contained in the secondary data file were absent from the research design. For example, social-emotional factors which are known to influence the educative outcomes for students with LD could have effectively served as potential predictor variables but were not included as IVs for the study. Personal demographic information about the individual students'

family background was also not included in the secondary data file, such as their ethnic or racial heritage. Finally, the research was based on a secondary data file representing students with LD from a mid-sized, urban school district in Ontario; therefore, caution should be exercised respecting the generalizability of the results outside of this jurisdiction.

Connecting the study's results more directly to the bioecological model of human development would posit that the prediction model for achievement in students with LD should include the interaction of several ecological variables at various system levels. Main findings from the study show differential impacts of the program placement microsystem, depending on the interaction of this variable with important individual student characteristics such as LAN. Similar results were found in the interaction between the school level microsystem of program placement and the prior achievement variable, germane to the individual student. Finally, one of the sociodemographic variables was a microsystem (i.e., SES) while the other was an individual characteristic (SEX). Although neither of these variables were significant predictors of achievement, descriptive statistics from the study demonstrate that their relationship with this aspect of human development (i.e., achievement) remains strongly associated. Thus, the inclusion of these variables in the prediction model from a theoretical lens was supported.

In sum, there were several limitations related to the secondary nature of the data file, which resulted in an absence of critical information about the psychological profiles of the study's participants or the time at which they received their formal LD diagnosis. Moreover, the methodological improvements made to the research design addressed several operational weaknesses identified in previous research; however, these enhancements were not necessarily perfect solutions. The attempt to match the groups on the basis of LAN was prone to a selection bias which may have confounded the sample; the SES variable may have suffered from the

ecological fallacy due to the procedural limitations of assigning a group level score to individual students; the use of a standardized criterion referenced assessment (EQAO) may also have a limited scope as a valid measure of academic achievement. Finally, there were certain epistemological delimitations of the quasi-experimental, correlational research design which must be considered respecting the generalizability of the study's results, or when discussing future directions for research.

Directions for Future Research

Understanding program activity architecture. According to Teddlie and Tashakkori (2003), the mixed-methods research approach emerged as the major counterargument to the incompatibility thesis between the empiricist and constructivist research paradigms. Johnson and Onwuegbuzie (2004) agree, noting that the underlying philosophy of mixed methods research is a fusion of approaches which represents a new, third paradigm. Exponents of the mixed methods approach claim that it can simultaneously address both confirmatory and exploratory research questions (Creswell, 2009b), and this potential holds much promise for future research directions for informing program delivery options for students with LD.

One of the main limitations of the current research design noted earlier was that the specific program activity architecture operating within either educational setting was virtually unknown. For example, did all teachers in the inclusive or non-inclusive settings practice identical instructional strategies? What was the teachers' knowledge or beliefs about students with LD in either program placement? Were the Universal Design for Learning, Response to Intervention or Differentiated Instruction approaches espoused by the Ministry of Education's *Learning For All* (2013) document comparable in both inclusive and non-inclusive classrooms? Were teachers in both settings similarly trained in the specific remediation techniques required

for students with LD? Without knowledge of this critical program-level information, it is difficult to say whether or not the type and quality of instruction was homogenous across placement conditions. In other words, we have no knowledge of the actual program activity architecture that operated in either educational setting. This information surely has implications for the academic outcomes for students with LD placed in either inclusive or non-inclusive program placements.

In order to address this important knowledge gap, future research examining inclusive education could employ a mixed model, as outlined by Teddlie and Tashakkori (2003). This type of study could feature a sequential, confirmatory then exploratory design, as well as procedures for integrating both quantitative and qualitative methods at various steps of the research. In the confirmatory step, an empiricist methodology could be employed in order to measure achievement outcomes between two groups of students with LD. Similar to the current study, the results of standardized assessments could be effectively used as the outcome measure. An exploratory methodology could then be employed that features constructivist methods in order to examine teacher beliefs or instructional characteristics across program placement conditions as a measure of program fidelity. For example, a survey of teacher beliefs about inclusion exploring pedagogical practices related to Universal Design for Learning, Response to Intervention or Differentiated Instructions. This much constitutes mixing at the epistemological level.

Mixing at the methods level could involve integrative, collaborative processes for designing the teacher survey instruments, such as convening regular classroom teachers, special education teachers, parents and students together to discuss survey items or other key dimensions of interest for the survey. Other integration techniques in the data analysis step could involve combining data from the standardized assessments with the survey responses from teachers to

explore the tensions that may result from a mixed-methods approach to data analysis (Greene, 2009). Reports could also be integrated to feature statistical analyses and written narratives combined, exploring teacher practices or beliefs about inclusion together with quantitative outcome data disaggregated by program placement type. This way, a mixed model study investigating both exploratory and confirmatory questions related to the inclusion model could be useful for contributing to our understanding of outcomes for LD students placed in either specialized or regular classroom settings.

Developing an ecological referral process. Johnson and Onwuegbuzie (2004) point out that another goal of mixed methods designs is to expand the researcher's understanding of a particular research question. Thus, the mixed methods research approach may also be important for contributing to our future understanding of the process used to refer students with LD for services in specialized or inclusive settings. For instance, we know that a student who may have learning difficulties is frequently referred for a standardized, psychological assessment that typically employs the discrepancy model to determine a LD diagnosis. However, Klassen (2002) notes that there is no theoretical basis in the research which supports a definition of LD on the basis of the discrepancy model. The author contends that rather than defining the construct of LD in strictly empirical terms, (i.e., difference between intelligence tests and academic tests), other ecological variables such as social, cultural, and economic factors should be considered in order to develop a broader definition of LD. There is a connection here to Johnson and Onwuegbuzie's (2004) notion of 'mindful mixed methods', and Greene's (2002) perspective on the dialectical stance.

The dialectical typology outlined in Greene, Benjamin and Goodyear (2001) highlight a myriad of research design possibilities that would be useful in developing a more ecological

referral process for identifying a student with LD. For example, a coordinated parallel research design might include coordinating a structured interview with the parents, school staff, and the student, aimed at understanding how the social, cultural and other ecological factors contribute to the particular student's learning difficulties. This information could be used in parallel with information obtained from the standardized psychological assessments of intelligence and academic achievement tests in order to deepen understanding of the child's particular LD diagnosis. This type of information would be extremely useful in providing a more complete profile of the student with LD which moves beyond the discrepancy model because it includes other ecological information about the student that might contribute to a more fulsome understanding of their academic difficulties.

The province of Ontario revised its policy defining LD in 2014; however, the definition still centres on two key components of the discrepancy model (Ontario Ministry of Education, 2014). A more ecological referral process for students with LD would define the construct in terms of the learning processes occurring in school or at home rather than relying solely on idiosyncratic factors at the student level (Waber, 2010). Future research using mixed methods could instead provide this holistic, contextual information that would support a referral process based on a more ecological conceptual definition of LD. Moreover, the dialectical tensions resulting from the qualitative and quantitative information obtained from mixed approaches could be used to broaden the school system's understanding of the types of educational programs or services required to serve the wide range of psychological profiles of students with LD. In this way, a more ecological referral process could better inform service delivery for students with LD, regardless of placement in inclusive or non-inclusive educational settings.

Conclusion

What conclusion(s) can be drawn from the results of this study? Do students with LD educated in a regular classroom perform better academically when compared to their counterparts who receive instruction in a specialized program? Are achievement outcomes for these students impacted differently by their previous academic history, sex, or SES? What role does the student's LAN play in contributing to their achievement outcomes in inclusive or non-inclusive settings? The academic research literature abounds with studies related to these issues, commonly referred to as the 'inclusion' debate; however, the review of literature suggests that more current and methodologically rigorous research is required in order to better understand whether or not program placement is related to academic achievement in students with LD.

Most studies investigating the impact of placement-type on achievement are outdated, or were conducted within education systems that are much different than the present special education milieu in Ontario. Consequently, there is a lack of conclusive, large scale, and recent research examining the relationship between LD program placement and academic outcomes that is relevant for the current provincial context. Therefore, re-examining the issue of program placement for students with LD, using a more methodologically sound research design, was important for contributing to the ongoing academic inclusion debate.

Despite the limitations outlined above, one of the most important conclusions that can be drawn from this particular study is that its theoretical, conceptual and methodological framework was sound. For instance, this study addressed several methodological problems identified by previous researchers through improvements to the research design. This included: i) the use of a matched comparison groups on the variable of LAN prior to comparing outcomes; ii) a repeated-measures analysis in order to control for the variance in test scores attributable to the individual

student; iii) the use of standardized, curriculum-based criterion-referenced assessments as outcome data; and iv) the inclusion of other sociodemographic variables in the prediction model that were previously known to influence achievement for students with LD. In so doing, the study effectively applied the theory of bioecological human development (Bronfenbrenner & Morris, 2006) in order to explore the relationship between program placement and the educational outcomes of elementary students with LD. Thus, it can be concluded that the study's methodological, conceptual and theoretical approach presented a better way to examine the impact of inclusion on the academic outcomes for students with LD.

Results from this study also point to several other important conclusions. Foremost is that program placement itself does not predict achievement for students with LD. This important conclusion can be drawn because, after controlling for the influence of the LAN variable in the follow-up analysis, program placement did not influence achievement. The same is true for the effect of the prior achievement variable which disappeared after the LAN variable was accounted for in the second regression model. Thus, it can be concluded that the influence of these two variables on achievement really depended on which students were included in the frame of reference. A wider focus indicated the favourability of the inclusion model, whereas the benefits of this model vanished when the frame of reference narrowed onto more comparable groups of students. These findings have several potential implications for educators, advocacy groups, and policy makers on both sides of the inclusion debate. After all, a specialized program placement is still recognized by the Ontario Ministry of Education as a viable service-delivery option for students with LD, and is considered by some school districts as the ideal programming option for more severely learning disabled students. One can conclude from the results of the current study that this may not be the case.

The evidence from the study's sociodemographic variables also supports the twin conclusions that boys and students with lower SES are over-represented in specialized LD programs. However, one must also conclude from the sociodemographic results that these variables did not significantly influence achievement for students with LD in either inclusive or non-inclusive educational settings. The main scholarly conclusion which can be drawn from the study's findings is that they support the defensible concern in the academic community regarding the strength of the research supporting a policy of inclusion for students with LD. From an academic perspective, future research questions may not be as simple as examining whether or not students with LD perform better in inclusive or specialized settings, or what service delivery model will best serve the multitudinous academic, social and emotional needs of this particular student population. In the same vein, Ontario's provincial legislators, district policy-makers, human rights advocates and organized labour unions must move beyond their collective fixation on the organizational structure of program configurations. Stakeholders must instead concede that the questions of what and how programs are delivered to students with LD are likely more related to academic outcomes, than where the program takes place.

References

- Al-Yagon, M. & Mikulincer, M., (2004). Socioemotional and academic adjustment among children with Learning Disorders: The mediational role of attachment-based factors. *The Journal of Special Education*, 38 (2) 111–123.
- Andrews, J., & Lupart, J. L. (2000). *The inclusive classroom: Educating exceptional children* (2nd Ed.). Scarborough, ON: Nelson.
- Banerji, M., & Dailey, R. A. (1995). A study of the effects of an inclusion model on students with specific learning disabilities. *Journal of Learning Disabilities*, 28(8), 511–522.
- Baker, E. T., Wang, M. C., & Walberg, H. J. (1994-1995). The effects of inclusion on learning. *Educational Leadership*, 52, 33-35.
- Beitchman, J. H., Wilson, B., Douglas, L., Young, A., & Adlaf, E. (2001). Substance use disorders in young adults with and without LD: Predictive and concurrent relationships. *Journal of Learning Disabilities*, 34, 317-332.
- Bennett, S. (2009). *Including Students with Exceptionalities*, Ontario Ministry of Education, Research Monograph: 16. Retrieved from:
<http://www.edu.gov.on.ca/eng/literacynumeracy/inspire/research/bennett.pdf>.
- Berninger, V.W., Nielsen, K.H., Abbott, R.D., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology* 46 (2), 151–172.
- Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, 17(3), 399–417.

- Blair, C. & Scott, K. G. (2002). Proportion of LD placements associated with low socioeconomic status: Evidence for a gradient? *The Journal of Special Education, 36*, 14-22.
- Bronfenbrenner, U. (1979). *The ecology of human development*. 1st ed. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.). *Theoretical models of human development Volume 1 of Handbook of Child Psychology* (6th ed.). Editors -in-chief: W. Damon & R. M. Lerner. Hoboken, NJ: Wiley.
- Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada (2018), *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Retrieved from http://www.pre.ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html.
- Carlberg, C., & Kavale, K. A. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education, 14*, 295-309.
- Chard, D., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K. L. S., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly, 31*(1), 11-20.
- Cigman, R. (2007). A question of universality inclusive education and the principle of respect. *Journal of Philosophy of Education, 41* (4), 231.
- Cole, C. M., Waldron, N., & Majd, M. (2004). Academic progress of students across inclusive and traditional settings. *Mental Retardation, 42*(2), 136-144.
- Cousins, J. B., Ross, J. A., & Prentice, M. (1993). Teachers' evaluation of correlational reasoning skills. *Alberta Journal of Education, 39*(3), 297-317.

- Coutinho, M.J., Oswald, D. P., & Best, A.M., (2002). The influence of sociodemographics and gender on the disproportionate identification of minority students as having learning disabilities. *Remedial and Special Education, 23 (1)*, 49-59.
- Cramer, D., & Howitt, D. L. (2004). *The SAGE dictionary of statistics: A practical resource for students in social sciences*. London: Sage Publications.
- Creswell, J. (2009a). *Research design: Qualitative, quantitative and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.
- Creswell, J. (2009b). Editorial: Mapping the field of mixed methods research. *Journal of Mixed Methods Research, 3(2)*, 95-108.
- Crone, Linda & Lang, Michael & Franklin, Bobby & Halbrook, Arthur. (1994). Composite Versus Component Scores: Consistency of School Effectiveness Classification. *Applied Measurement in Education, 7*.
- Demeris, H., Childs, R., & Jordan, A. (2007). The influence of students with special needs included in Grade-3 classrooms on the large-scale achievement scores of students without special needs. *Canadian Journal of Education / Revue canadienne de l'éducation, 30*, 609-627.
- Denzin, N. (2008). The new paradigm dialogs and qualitative inquiry. *International Journal of Qualitative Studies in Education 21(4)*, 315–325.
- Devine, A., Soltesz, F., Nobes, A., Goswami, U., & Szucs, D., (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learning and Instruction, 27*, 31-39.
- D'Intino, J. S. (2017). Learning disabilities in Canada: Definitions and accommodations. *Canadian Psychology/Psychologie canadienne, 58(3)*, 228–237.

- Dyson, D., & Farrell, P T., (2007). But what about the others? Patterns of student achievement in inclusive schools." In *Included or Excluded? The challenge of the mainstream for some SEN children*, Ed. R., Cigman. London: Routledge.
- Ebsensen, F-A., Piper Deschenes, E., Vogel, R. E., West, J., Arboit, K., & Harris, L. (1996). Active parental consent in school-based research: An examination of ethical and methodological issues. *Evaluation Review*, 20(6), 737-753.
- Enders, C. K., (2016). Multiple imputation as a flexible tool for missing data handling in clinical research, *Behaviour Research and Therapy*. Retrieved from:
<http://dx.doi.org/10.1016/j.brat.2016.11.008>.
- Felder, F. (2018). The value of inclusion. *Journal of Philosophy of Education*, 52(1), 54-70.
- Flannery, K., Liederman, J., Daly, L., & Schultz, J. (2000). Male prevalence for reading disability is found in large sample of Black and White children free from ascertainment bias. *Journal of the International Neuropsychological Society*, 6, 433-442.
- Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligence testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice*, 13, 186-203.
- Florian, L.D., & McLaughlin, M. J., (2008). *Disability classification in education: Issues and perspectives*. Thousand Oaks, CA: Corwin Press.
- Fuchs, D., & Fuchs, L. S. (1995). Special education can work. *Issues in educational placement: Students with emotional and behavioral disorders*, 363-377.
- Gandhi, A. G., (2007). Context Matters: Exploring relations between inclusion and reading achievement of students without disabilities. *International Journal of Disability, Development and Education* 54 (1), 91-112.

- Gorbunova A.A., Lemeshko, S., & Lemeshko, B. (2010). *Classical tests of variances homogeneity for non-normal distributions*. Conference: Proceedings Third International Conference on Accelerated Life Testing, Reliability-based Analysis and Design. 19-21 May 2010, Clermont-Ferrand, France. Retrieved From:
https://www.researchgate.net/publication/236174159_Classical_tests_of_variances_homogeneity_for_non-normal_distributions
- Government of Ontario (2019). O. Reg. 181/98, s. 16 (7). Retrieved From:
<https://www.ontario.ca/laws/regulation/980181>.
- Greene, J. C., Benjamin, L., & Goodyear, L. (2001). The merits of mixing methods in evaluation. *Evaluation, 7*, 25-44.
- Greene, J. C. (2002). With a splash of soda please: Towards active engagement with difference. *Evaluation, 8*, 259-266.
- Greene, J. C. (2009). Meaningfully engaging with difference through mixed methods educational evaluation. In K. E. Ryan & J. B. Cousins (Eds.) *The sage international handbook of educational evaluation* (p. 323-329). Thousand Oaks, CA: Sage.
- Grice, J. & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. *Applied Multivariate Research, 12*, 199-226.
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. (1995) *Multivariate data analysis*, 3rd ed. New York: Macmillan Publishing Company.
- Harrington, S.A., (1997). Full inclusion for students with learning disabilities: a review of the evidence. *The School Community Journal, 7*, 63–71.

- He, Y., Zaslavsky, A., Landrum, M., Harrington, D., & Catalano, P. (2010). Multiple imputation in a large-scale complex survey: a practical guide. *Statistical Methods in Medical Research*, 19(6), 653–670.
- Hocutt, A.M. (1996). Effectiveness of special education: Is placement the critical factor? *Special Education for Students with Disabilities*, 6(1), 77-102.
- Holmes-Finch (2016). Missing data and multiple imputation in the context of multivariate Analysis of Variance. *The Journal of Experimental Education*, 84 (2), 356-372.
- Hchokr, (2012). *Bronfenbrenner's Ecological Theory of Development*. Retrieved from: http://en.wikipedia.org/wiki/File:Bronfenbrenner%27s_Ecological_Theory_of_Development.jpg.
- Hutchinson, N. L., & Martin, A. K. (2012). *Inclusive classrooms in Ontario schools*. Toronto, ON: Pearson Education Canada.
- Johnson, R., & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14 36.
- Judge, S., & Bell, S. M. (2011). reading achievement trajectories for students with learning disabilities during the elementary school years. *reading & writing Quarterly: Overcoming Learning Difficulties*, 27(1-2), 153–178.
- Kalambouka, A., Farrell, P T., Dyson, D., I & Kaplan (2007). "The impact of placing pupils with special educational needs in mainstream schools on the achievements of their peers." *Educational Research* 49 (4) 365-382.
- Kang, H., (2013). The prevention and handling of the missing data, *Korean Journal of Anesthesiology* 64(5), 402–406. Retrieved from: doi: 10.4097/kjae.2013.64.5.402.

- Kauffman, J. M., Mock, D. R., Tankersley, M., & Landrum, T. J. (2008). Effective service delivery models. In R. J. Morris & N. Mather (Eds.), *Evidence-based interventions for students with learning and behavioral challenges* (pp. 359-378). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kavale, K.A., & Forness, S., (2000). History, rhetoric and reality: Analysis of the inclusion debate. *Remedial and Special Education, 21* (5), 279-296.
- Kavale, K.A., (2005). Identifying specific learning disability: Is responsiveness to intervention the answer? *Journal of Learning Disabilities, 38*, 553–562.
- Kavale, K.A., & Mostert, M.P., (2004). Social skills interventions for individuals with learning disabilities. *Learning Disability Quarterly, 27*, 31-43.
- Keogh, B. K., & Speece, D., (1996). *Research on classroom ecologies*. Hillsdale, NJ: Erlbaum.
- Klassen, R. M., (2002). The changing landscape of learning disabilities in Canada: Definitions and practice from 1989-2000. *School Psychology International, 23*, 199-219.
- Klinger, D. A., Rogers, W., Anderson, J., Poth, C. A., & Calman, R., (2006). Contextual and School Factors Associated with Achievement on a High-Stakes Examination. *Canadian Journal of Education, 29*, 771-797.
- Klinger, D. A., Shulha, L. M., & Wade-Woolley, L., (2009). *Towards an Understanding of Gender Differences in Literacy Achievement*. Academic research report for the Education Quality and Accountability Office, Toronto, ON.
- Knief, U., & Forstmeier, W. (2018). *Violating the normality assumption may be the lesser of two evils*. Retrieved From:
https://www.researchgate.net/publication/329839272_Violating_the_normality_assumption_may_be_the_lesser_of_two_evils

- Kohn, A., (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi delta Kappan*.
retrieved June 7, 2008, from <http://www.alfiekohn.org/teaching/ftt.htm>.
- Kozey, M. & Siegel, L.S. (2008). Definitions of learning disabilities in Canadian provinces and territories. *Canadian Psychology*, 49, 162-171.
- Krawec, J., Huang, J., Montague, M., Kressler, B., & de Alba, A. M. (2013). The effects of cognitive strategy instruction on knowledge of math problem-solving processes of middle school students with learning disabilities. *Journal of Learning Disabilities Quarterly*, 36(2), 80-92.
- Learning Disabilities Association of Canada (2019). Retrieved from: <http://www.ldac-acta.ca/learn-more/ld-defined/official-definition-of-learning-disabilities>
- Learning Disabilities Association of Canada (2002). Retrieved May 2014 from:
<http://www.ldao.ca/introduction-to-ldsadhd/ldsadhs-in-depth/articles/about-lds/learning-disabilities-statistics/>
- Lee J. H., & Huber, J., (2011). Multiple imputation with large proportions of missing data: How much is too much? *United Kingdom Stata Users' Group meetings 2011 23, Stata Users Group*. Retrieved from: <https://ideas.repec.org/p/boc/usug11/23.html>.
- Lewthwaite, B.E., & Wiebe, R. (2011). Chemistry teacher development towards a tetrahedral orientation in the teaching of chemistry. In B.E. Lewthwaite (Ed.) University of Manitoba Centre for Research in Youth, Science Teaching and Learning: *Applications and Utility of Urie Bronfenbrenner's Bio-ecological Theory* (8-23). Winnipeg, MB. MERN.
- Liddell, G. A., & Rasmussen, C. (2005). Memory profile of children with nonverbal learning disability, *Learning Disabilities Research and Practice*, 20, 137-141.

- Lindsay, G. (2003). Inclusive education: A critical perspective. *British Journal of Special Education, (30)* 3-12.
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health, 23*, 151-69.
- Lupart, J., & Webber, C. (2012) Canadian schools in transition: Moving from dual education systems to inclusive schools. *Exceptionality Education International, 22*, 8-37. Retrieved from <https://ir.lib.uwo.ca/eei/vol22/iss2/4>
- Lufi, D., Okasha, S., & Cohen, A. (2004). Test anxiety and its effect on the personality of students with learning disabilities. *Learning Disability Quarterly, 27*(3), 176-184.
- McDonnell, J., Thorson, N., Disher, S., mathot-Buckner, C., Menel, J., & Ray, L. (2003). The achievement of students with developmental disabilities and their peers without disabilities in inclusive settings: An exploratory study. *Education and Treatment of Children, 26*, 224-236.
- Manly, C. A., & Wells, R. S. (2015). Reporting the use of multiple imputation for missing data in higher education research. *Research in Higher Education, 56*(4), 397–409.
- Manset, G., & Semmel, M. (1997). Are inclusive programs for students with mild disabilities effective? A comparative review of model programs. *Journal of Special Education, 31*(2), 155–181.
- Marshall, A., Altman, D., Holder, R., & Royston, P., (2009). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Medical Research Methodology, 10* (7) Retrieved December 2019 from: https://www.researchgate.net/publication/41087746_Comparison_of_techniques_for_handling_missing_covariate_data_within_prognostic_modelling_studies_A_simulation_study

- Marston, D., (1987). The effectiveness of special education: A time-series analysis of reading performance in regular and special education settings. *Journal of Special Education* 27, 466–80.
- Marston, D. (1996). A comparison of inclusion only, pull-out only, and combined service models for students with mild disabilities. *Journal of Special Education*, 30, 121-132.
- McLeskey, J., & Waldron, N. (2011). *Full inclusion programs for elementary students with learning disabilities: Can they meet student needs in an era of high stakes accountability?* Paper Presentation at the Council for Exceptional Children Convention, National Harbor, MD, April 2011. Retrieved from:
<http://files.eric.ed.gov/fulltext/ED529797.pdf>
- McKinney, J. D. (1989). Longitudinal research on the behavioral characteristics of children with learning disabilities. *Journal of Learning Disabilities*, 22, 141-165.
- Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, 42, 306–321.
- Mullis, I.V.S., Martin, M.O., Kennedy, A.M., & Foy, P. (2007). *PIRLS 2006 International Report: IEA's Progress in International reading Literacy Study in Primary Schools in 40 Countries*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College. Retrieved from:
https://nces.ed.gov/pubs2008/2008017_2.pdf
- Murphy, M. M., Mazzocco, M. M.M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function

of the cut off criterion used to define MLD. *Journal of Learning Disabilities*, 40(5), 458–478.

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14 (1). Retrieved May 2015 from <http://epaa.asu.edu/epaa/v14n1/>

Norwich, B. (2008). Dilemmas of difference, inclusion and disability: international perspectives on placement. *European Journal of Special Needs Education*, (23):4, 287-304.

Ontario Human Rights Commission, (2019). Public inquiry, *Right to Read*. Retrieved From: <http://www.ohrc.on.ca/en>.

Ontario Ministry of Education, (2020). An Introduction to Special Education in Ontario. Retrieved From: <http://www.edu.gov.on.ca/eng/general/elemsec/speced/ontario.html>

Ontario Ministry of Education, (2019). *Realizing the Promise of Diversity: Ontario's Equity and Inclusive Education Strategy*. Retrieved from: <https://www.edu.gov.on.ca/eng/policyfunding/equity.pdf>

Ontario Ministry of Education, (2016). Special Education Update, p 5. Retrieved From: www.edu.gov.on.ca/eng/general/elemsec/speced/special_ed_update.html.

Ontario Ministry of Education, (2014). *Policy/Program Memorandum No. 8: Identification of and Program Planning for Students with Learning Disabilities*, August 26, 2014. Retrieved from: <http://www.edu.gov.on.ca/extra/eng/ppm/ppm8.pdf>.

Ontario Ministry of Education, (2013). *Learning for All: A Guide to Effective Assessment and Instruction for All Students Kindergarten to Grade 12*. Retrieved from: <http://www.edu.gov.on.ca/eng/general/elemsec/speced/LearningforAll2013.pdf>

- Ontario Ministry of Education, (2013). *Policy/Program Memorandum No. 119: Developing and implementing equity and inclusive education policies in Ontario Schools*. Retrieved from: <http://www.edu.gov.on.ca/extra/eng/ppm/119.pdf>
- Ontario Ministry of Education, (2006). Memo: *Special Education Transformation Initiatives*.
- Osborne, Jason., (2017). *Regression and Linear Modeling: Best Practices and Modern Methods*. Thousand Oaks: CA, Sage.
- Pampaka, M., Hutcheson, G. and Williams, J., (2016). Handling missing data: Analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, 39 (1), 19-37.
- Pang, X., & Rogers, W. T., (2013). *Comparative Examination of the Influence of Selected Factors on Achievement in Grade 9 Academic and Applied mathematics Courses in English-Language Schools in Ontario*. Report prepared for the Education Quality and Accountability Office (EQAO). Retrieved from:
http://www.eqao.com/Research/pdf/E/Comparative_Examination_Influence_Factors_Achievement_G9math_en.pdf
- Patrikakou, E. N. (1996). Investigating the academic achievement of adolescents with learning disabilities: A structural modeling approach. *Journal of Educational Psychology*, 88, 435–450.
- Pavri, S., & Monda-Amaya, L. (2000). Loneliness and students with learning disabilities in inclusive classrooms: Self-perceptions, coping strategies, and preferred interventions. *Learning Disabilities Research and Practice*, 15, 23-34.
- Popham, W.J. (1999). Where large scale educational assessment is heading and why it shouldn't. *Educational Measurement: Issues and Practice* 18 (3), 13–17.

- Rea, P. J., McLaughlin, V. L., & Walther-Thomas, C. S. (2002). A comparison of outcomes for middle school students with learning disabilities in inclusive and pullout programs. *Exceptional Children, 68*, 203-222.
- Ring, M., Reetz, L., (2000), Modification effects on attributions of middle school students with learning disabilities. *Learning Disabilities Research and Practice, 15(1)*, 35-43.
- Robinson, T. T., & Cousins, J. B. (2004). Internal participatory evaluation as an organizational learning system: A longitudinal case study. *Studies Educational Evaluation. 30*, 1-22.
- Rogers, W. (2014). Improving the utility of large-scale assessments in Canada. *Canadian Journal of Education/Revue canadienne de l'éducation. 37*, 1-22.
- Rogers, W., Anderson, J., Klinger, D., & Dawber, T. (2006a). *Pitfalls and Potential of Secondary Data Analysis of the Council of Ministers of Education, Canada, National Assessment*. Canadian Journal of Education / Revue Canadienne De L'éducation, 29(3), 757-770.
- Rogers, W., Ma, X., Klinger, D., Dawber, T., Hellsten, L., Nowicki, D., & Tomkowicz, J. (2006b). Examination of the Influence of Selected Factors on Performance on Alberta Learning Achievement Tests. Canadian Journal of Education / Revue canadienne de l'éducation. 29, 731-756.
- Rubin, D. B., (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91 (434)*, 473-489.
- Ruijs, N. M., Van der Veen, I., & Peetsma, T. D. (2010). Inclusive education and students without special educational needs. *Educational Research, 52*, 351-390.
- Shaffer, D. R., (1993). *Developmental psychology: Childhood and adolescence* (3rd ed.). Pacific Grove CA: Brooks/Cole.

- Shaywitz, S. E., & Shaywitz, B. A., (2006). *reading disability and the brain. Educating exceptional children*. Dubuque, IA: McGraw Hill.
- Simpson, C. G. Rose, C. A., Bakken, J. P. (2013). Placement of students with learning disabilities. In J. P. Bakken (Ed.), *Learning Disabilities: Identification, assessment, and instruction of students with learning disabilities (75-91)*. Bingley, UK: Emerald Group Publishing Limited.
- Sirin, S.R., (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research* 75 (3), 417–453.
- Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Feggins-Azziz, R., & Chung, C-G. (2005). Unproven links: Can poverty explain ethnic disproportionality in special education? *The Journal of Special Education*, 39, 130-144.
- Sokal, L., Thiem, C., Crampton, A., & Katz, H. (2009). Differential effects of male and female reading tutors based on boys' gendered views of reading. *Canadian Journal of Education*, 32(2), 245–270.
- Song, M. K., Lin, F. C., Ward, S. E., & Fine, J. P. (2013). Composite variables: when and how. *Nursing research*, 62 (1), 45–49.
- Sontag, J. C. (1996). Toward a comprehensive theoretical framework for disability research: Bronfenbrenner revisited. *Journal of Special Education*, 30, 319-44.
- Sousa, D. (2006). *How the brain learns*. 3rd ed. Thousand Oaks, CA: Corwin.
- Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (Fifth Edition). Boston: Pearson/Allyn and Bacon.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In Tashakkori, A. & Teddlie, C. (Eds.)

- Handbook of mixed methods in social & behavioural research* (p. 3-50). Thousand Oaks, CA: Sage.
- Test, D. W., Fowler, C. H., Brewer, D. M., & Wood, W. M. (2005). A content and methodological review of self-advocacy intervention studies. *Exceptional Children, 72*, 101-125.
- Thorndike, R.M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*, 8th Ed. Boston, MA: Pearson Education Inc.
- Usher, R. (1996). *A critique of neglected epistemological assumptions of educational research*. In D. Scott & R. Usher (Eds.) *Understanding educational research* p. 10-32. New York: Routledge.
- Vandenberg, B., & Emery, D., (2009). A longitudinal examination of the remediation of learning disabilities: IQ, age at diagnosis, school SES and voluntary transfer. *International Journal of Special Education, 24 (1)* 45-52.
- Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33*, 223-238.
- Vogel, S.A. (1990). Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities: A review of the literature. *Journal of Learning Disabilities, 23 (1)*, 44-52.
- Volante, L. (2006). An alternative vision for large-scale assessment in Canada. *Journal of Teaching and Learning, 4 (1)*, 1-13.

- Waber, D. P. (2010). *Rethinking learning disabilities: Understanding children who struggle in school*. New York: The Guilford Press.
- Waldron, N. L., & McLeskey, J. (1998). The effects of an inclusive school program on students with mild and severe learning disabilities. *Exceptional children*, 64(3), 395-405.
- Wang, J., and Johnson, D. E., (2018). An Examination of Discrepancies in Multiple Imputation Procedures Between SAS® and SPSS®, *The American Statistician*, DOI: 10.1080/00031305.2018.1437078 Retrieved from: <https://doi.org/10.1080/00031305.2018.1437078>
- Wiener, J. and Tardif, C. Y., (2004) Social and Emotional Functioning of Children with Learning Disabilities: Does Special Education Placement Make a Difference? *Learning Disabilities Research & Practice* 19 (1), 57-70.
- Wei, X., Lenz, K. B., & Blackorby, J. (2013). math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, 34 (3), 154-165.
- Wheldall, K., & Limbrick, L. (2010). Do more boys than girls have reading problems? *Journal of Learning Disabilities*, 43 (5), 418-429.
- Whitley, J., Lupart, J. L., & Beran, T. (2007). The characteristics and experiences of Canadian students receiving special education services for a Learning Disability. *Exceptionality Education International*, 17, 85-110. Retrieved from <http://ir.lib.uwo.ca/eei/vol17/iss3/5>.
- Whitley, J. (2008). A model of general self concept for students with learning disabilities: Does class placement play a role? *Developmental Disabilities Bulletin*, 36 (1/2), 106-134.

- Whitley, J., (2010). Modelling the influence of teacher characteristics on student achievement for Canadian students with and without Learning Disabilities, *International Journal of Special Education*, 25(3), 88-97.
- Winzer, M. (2008). *Children with Exceptionalities in Canadian Classrooms* (8th ed.). Toronto, ON: Pearson Prentice Hall.
- Wolfe, P., & Nevills, P. (2004). *Building the reading brain*. Thousand Oaks, CA: Corwin Press.
- Wolfe, R., Childs, R., & Elgie, S. (2004). *Ensuring Quality Assessments: A Project to Refine and Affirm Assessment Processes*. Ontario Institute for Studies in Education of the University of Toronto. Retrieved from: <file:///C:/Users/C38341/Downloads/EQAORReview.pdf>
- Young, G., Kim, H. J., & Gerber, P.J., (1999). Gender bias and learning disabilities: School age and long term consequences for females. *Learning Disabilities: A Multidisciplinary Journal*, 9 (3), 107-14.
- Zigmond, N. (2003). Where should students with disabilities receive special education services? Is one place better than another? *Journal of Special Education*, 37(3), 193-199.

Appendix A

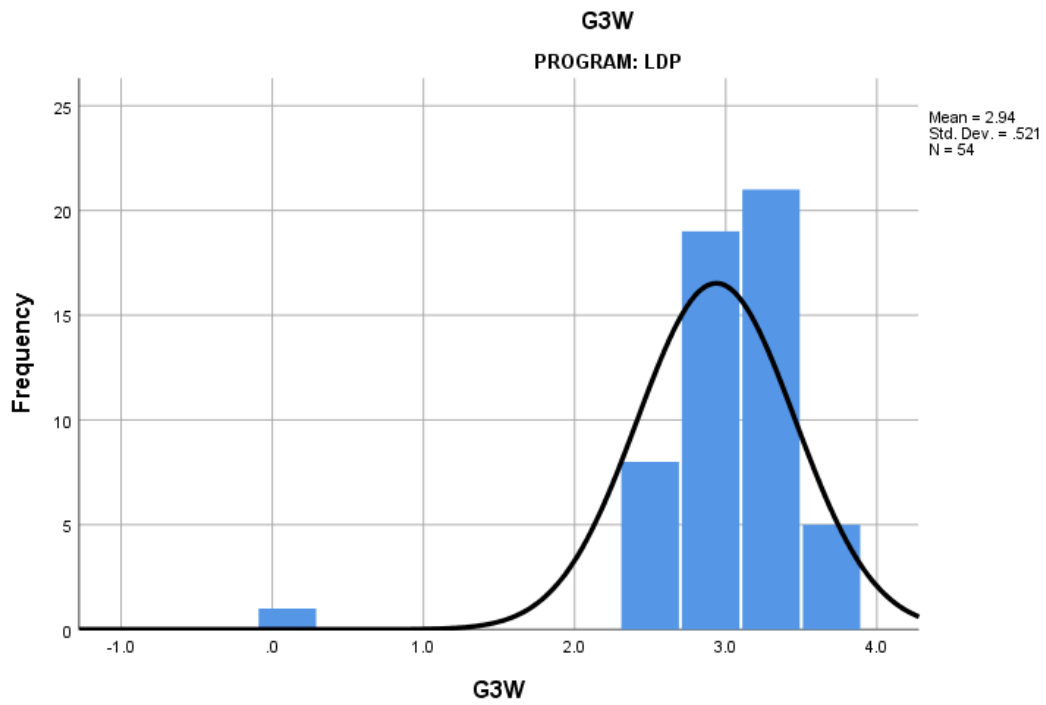
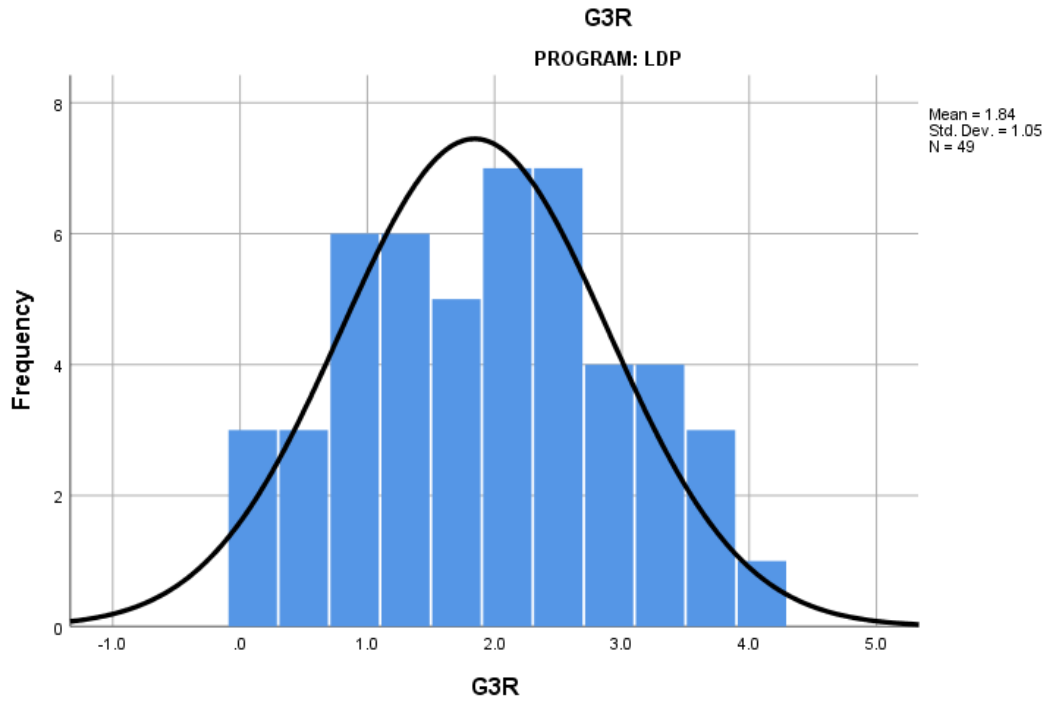
Normality Tests of the Original Data by Variable and Program

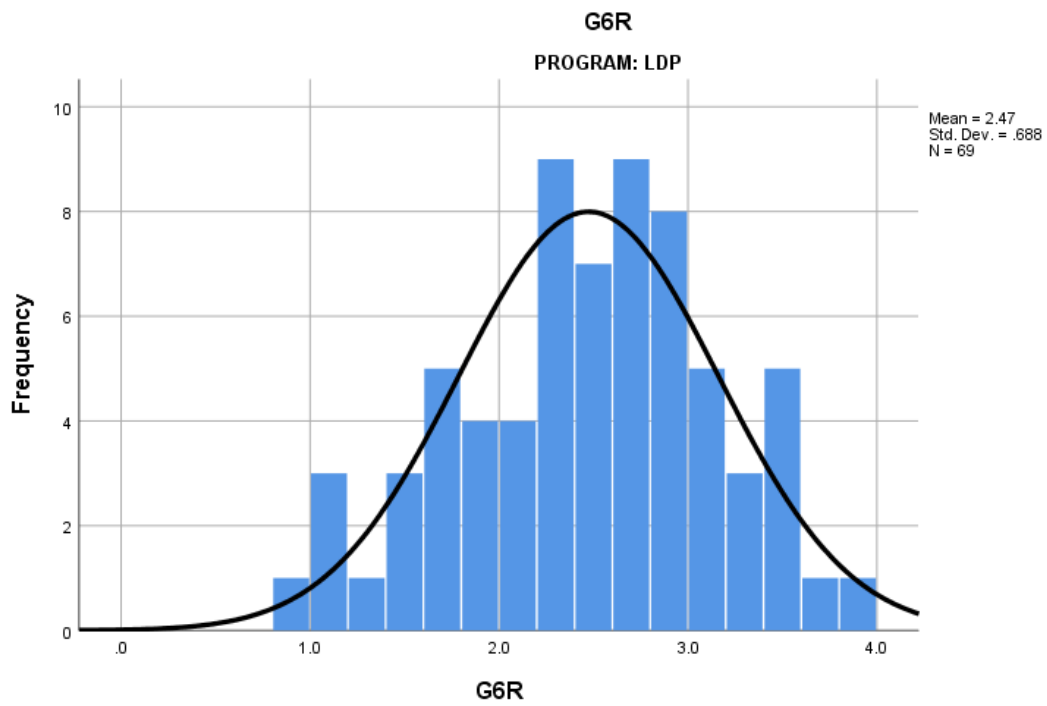
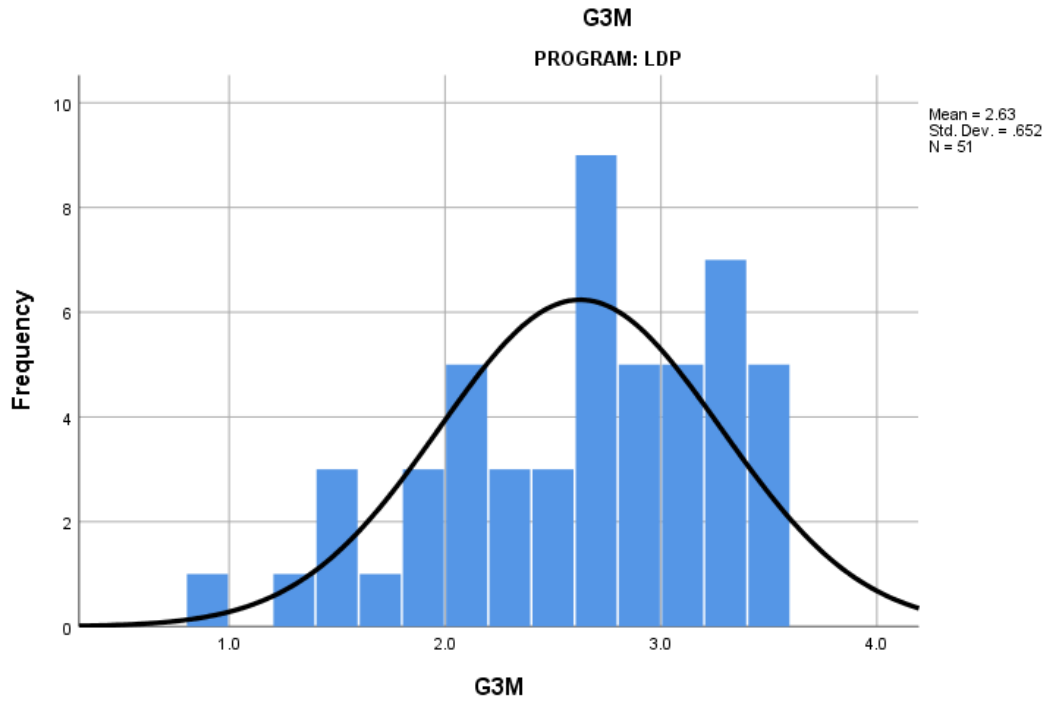
Program		Skewness			Kurtosis		
		<i>Statistic</i>	<i>SE</i>	<i>Value</i>	<i>Statistic</i>	<i>SE</i>	<i>Value</i>
LDP	G3R	0.163	0.34	0.48	-0.846	0.668	-1.27
	G3W	-3.031	0.325	-9.33*	15.856	0.639	24.81*
	G3M	-0.629	0.333	-1.89	-0.275	0.656	-0.42
	G6R	-0.26	0.289	-0.90	-0.411	0.57	-0.72
	G6W	-1.495	0.289	-5.17*	2.858	0.57	5.01*
	G6M	0.101	0.297	0.34	3.092	0.586	5.28*
WAITLIST	G3R	-0.385	0.414	-0.93	-0.201	0.809	-0.25
	G3W	-0.079	0.403	-0.20	0.292	0.788	0.37
	G3M	0.402	0.409	0.98	-0.969	0.798	-1.21
	G6R	-0.673	0.35	-1.92	-0.376	0.688	-0.55
	G6W	-2.169	0.35	-6.20*	10.574	0.688	15.37*
	G6M	-0.141	0.347	-0.41	1.499	0.681	2.20*
REGULAR	G3R	-0.641	0.096	-6.68*	0.032	0.191	0.17
	G3W	-1.313	0.095	-13.82*	7.406	0.19	38.98*
	G3M	-0.612	0.095	-6.44*	0.737	0.19	3.88*
	G6R	-0.99	0.091	-10.88*	1.371	0.182	7.53*
	G6W	-0.937	0.091	-10.30*	3.237	0.182	17.79*
	G6M	-0.112	0.091	-1.23	-0.406	0.182	-2.23*

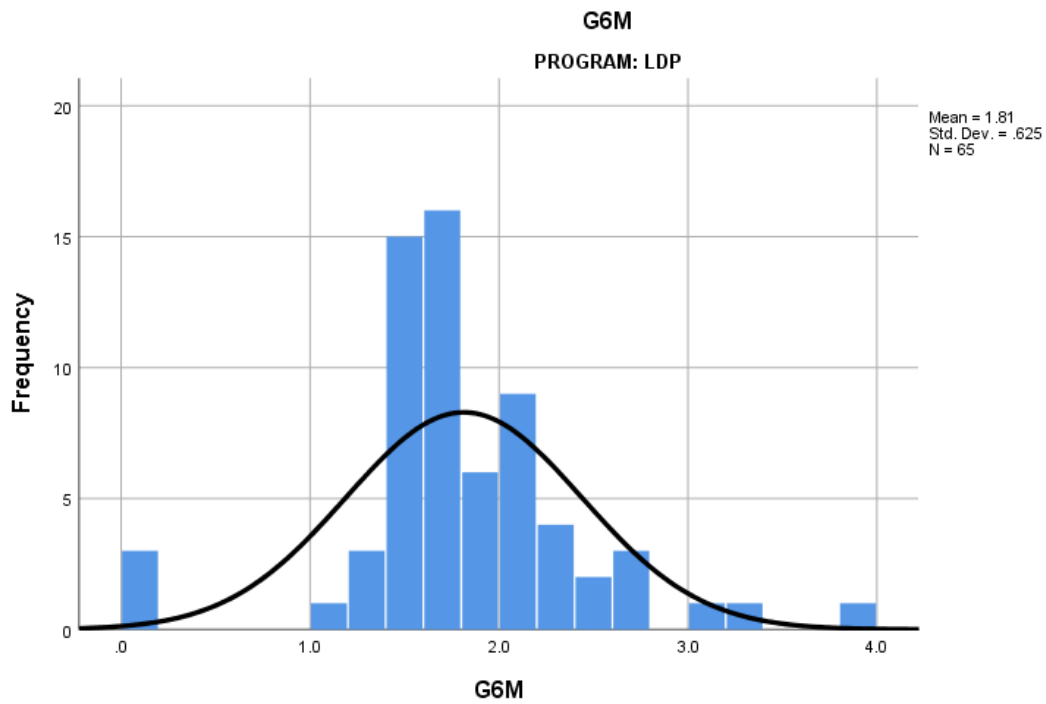
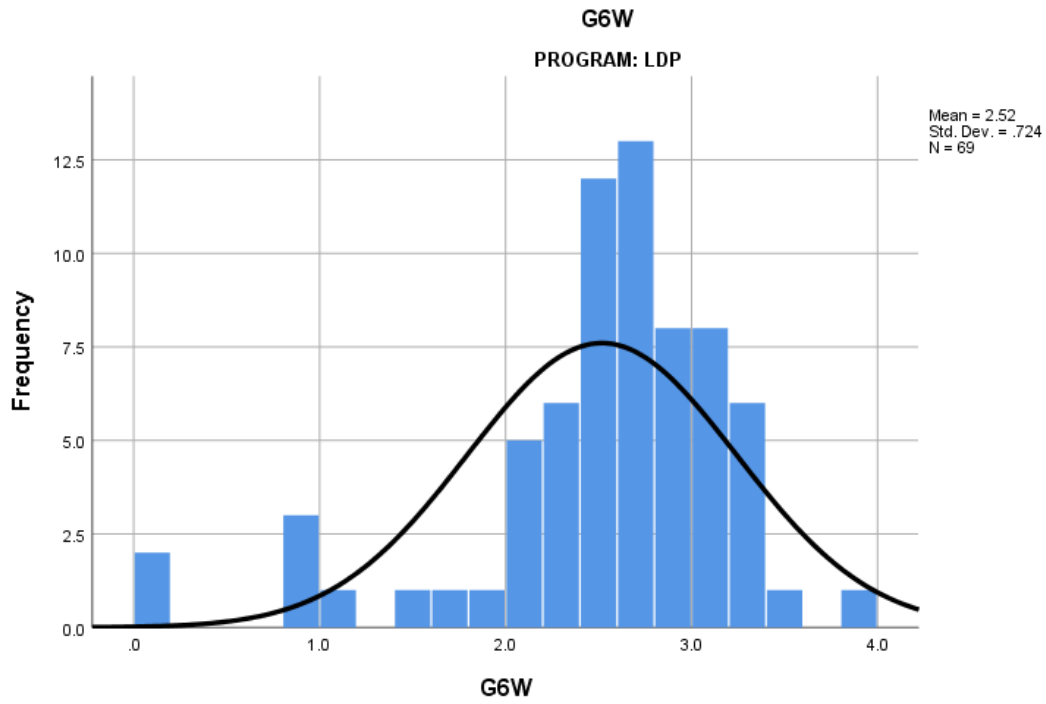
Note. * value exceeds the +/- 1.96 normality threshold for the measure.

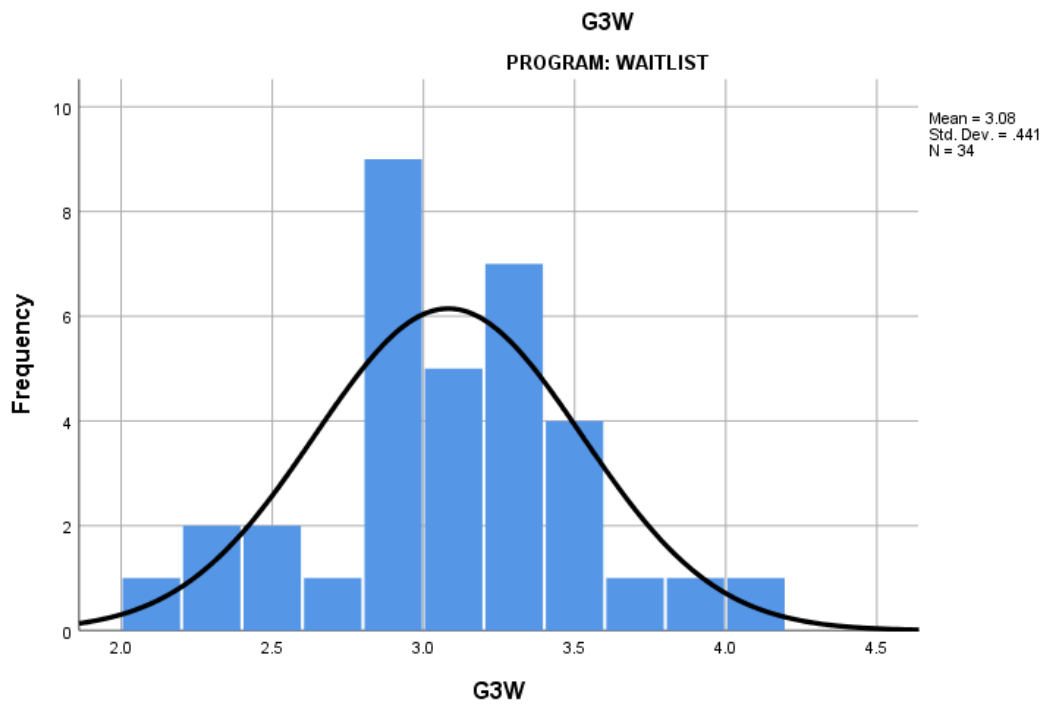
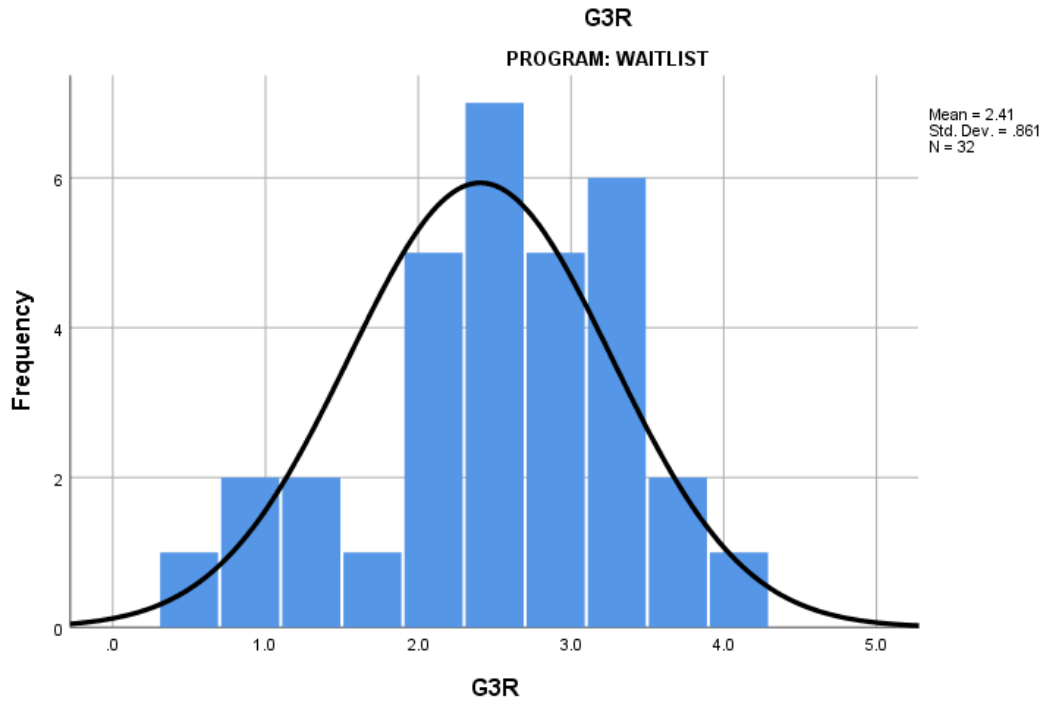
Appendix B

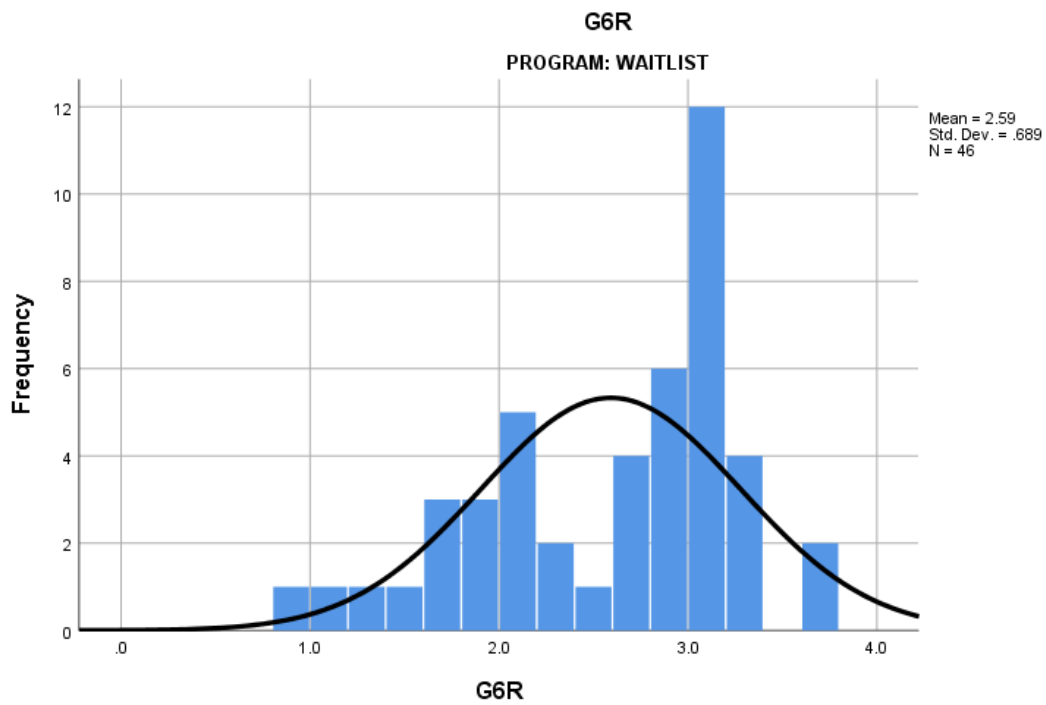
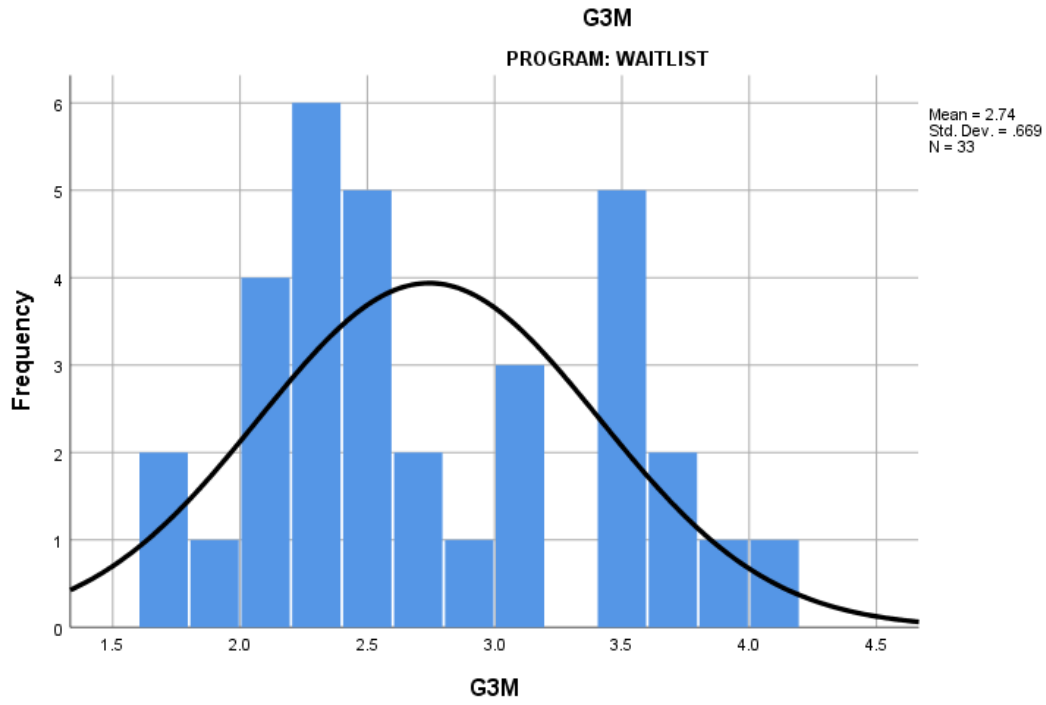
Histograms of the Original Outcome Data by Subject Area and Program

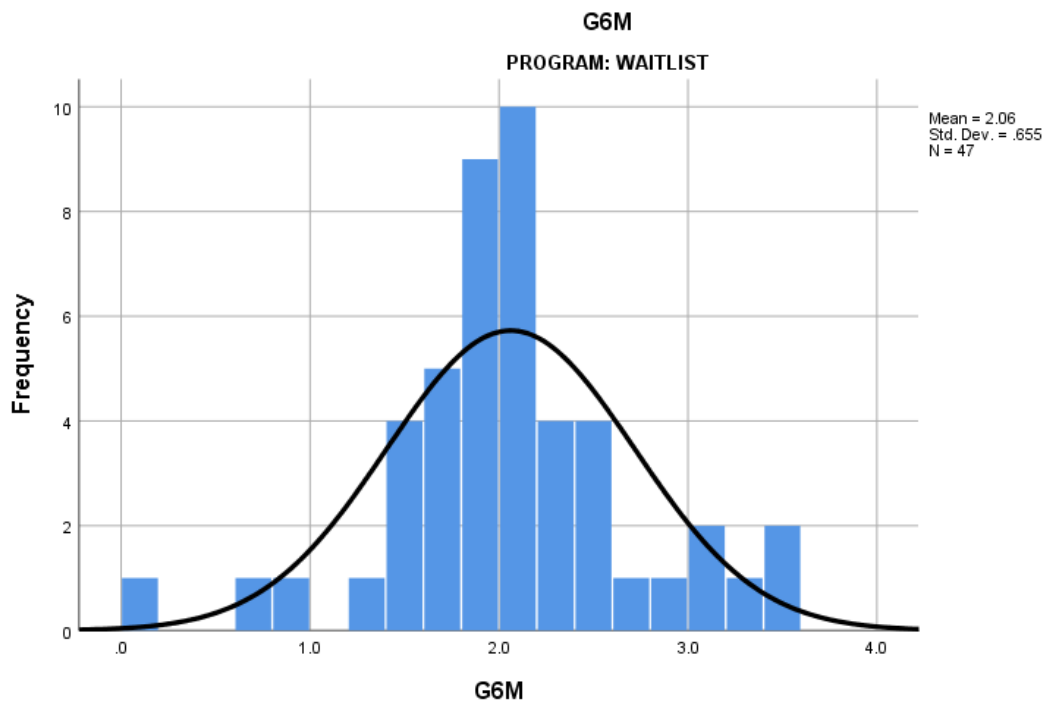
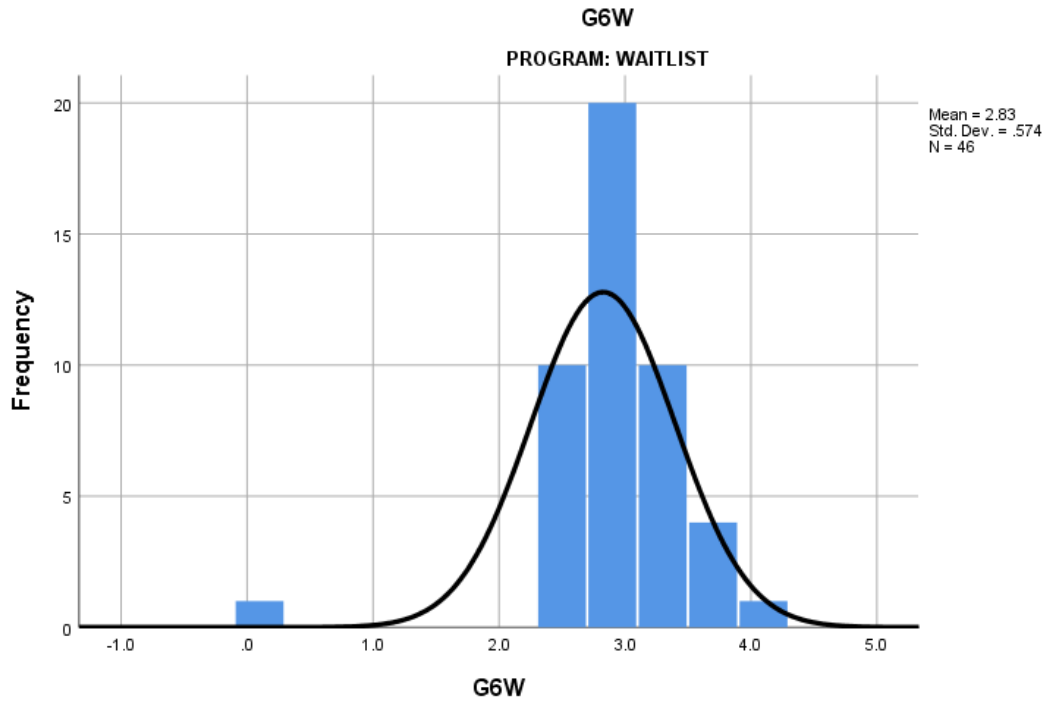


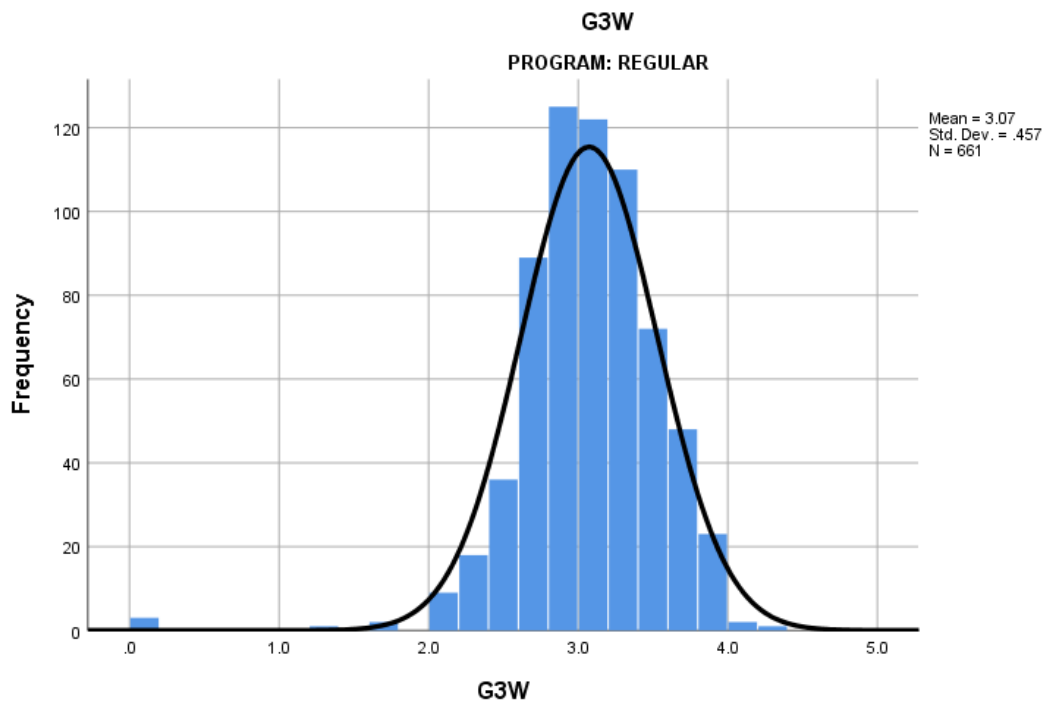
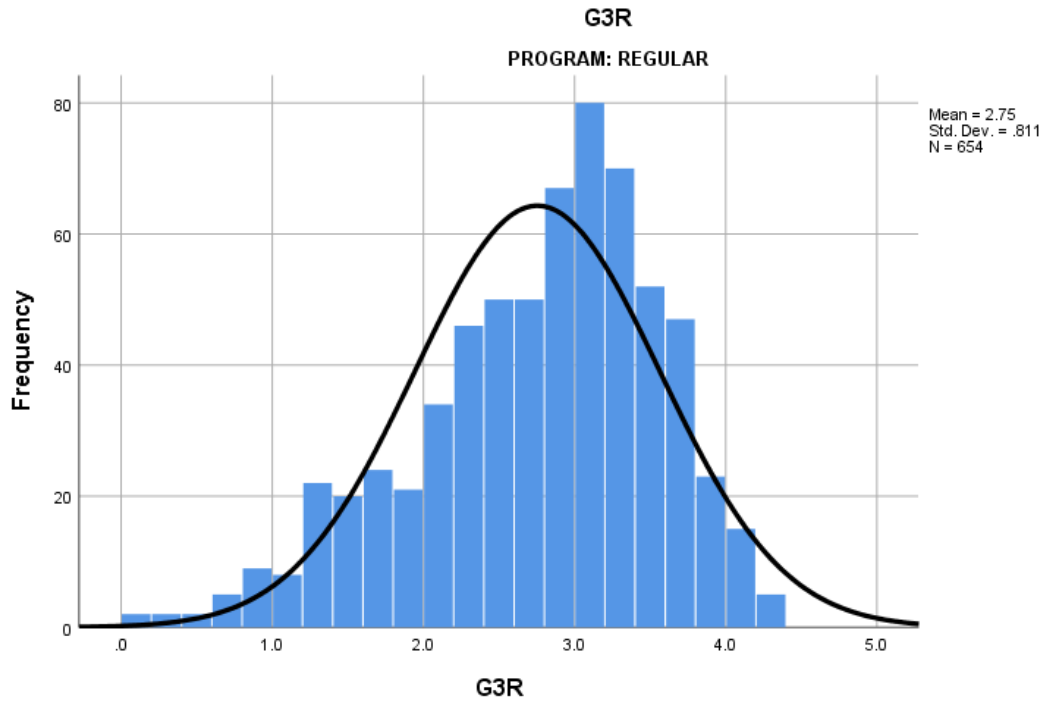


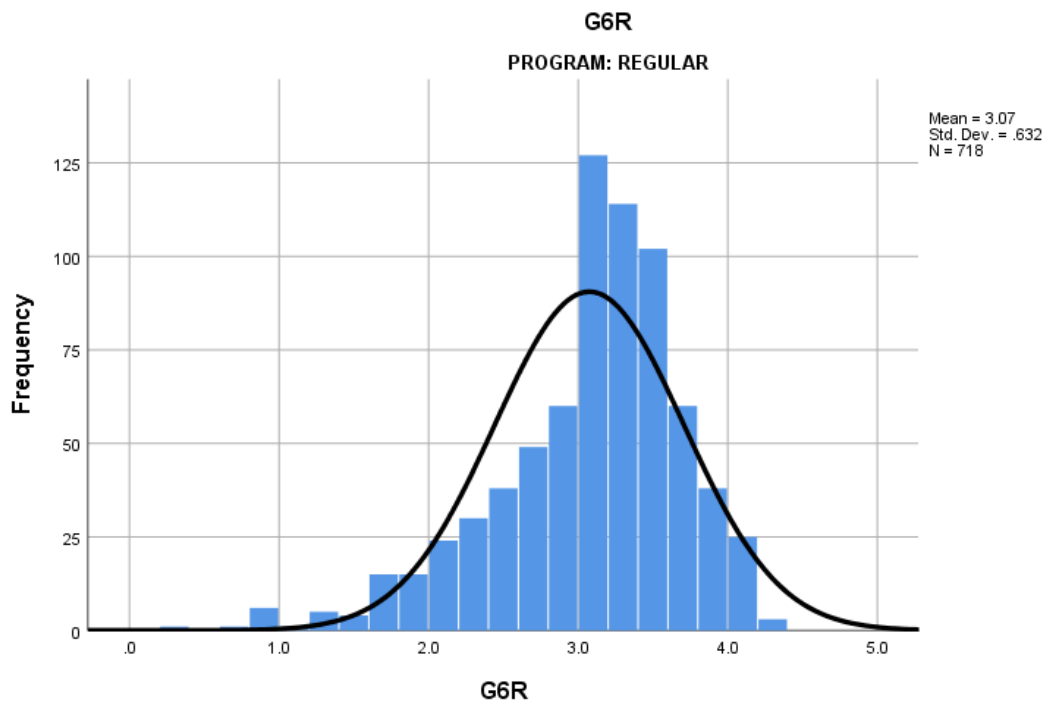
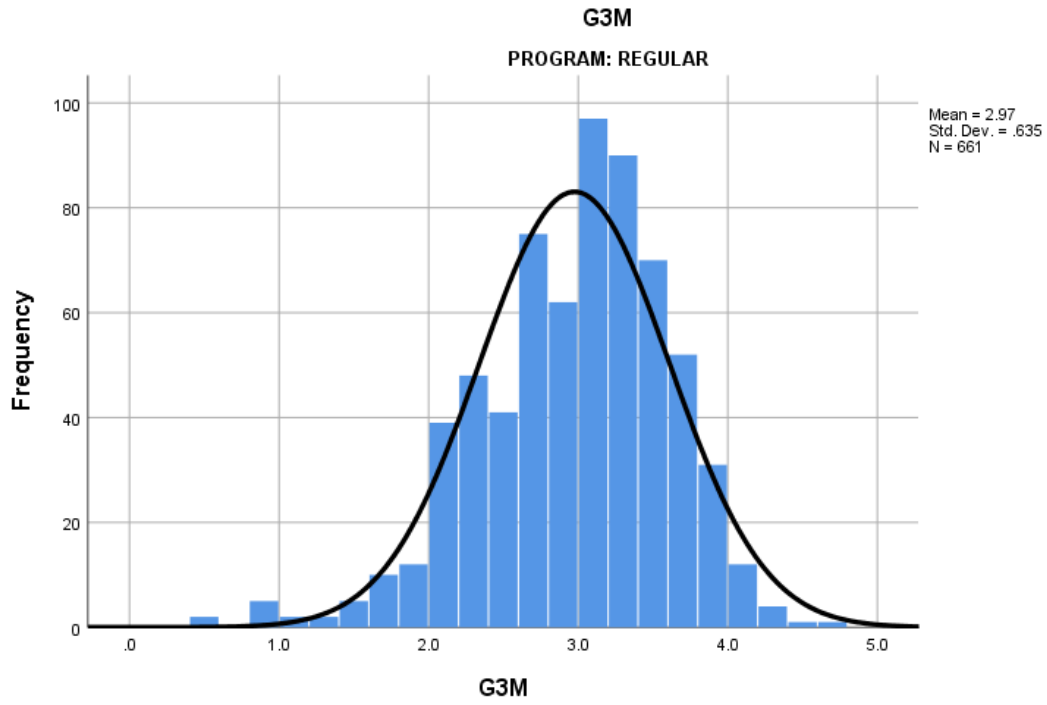


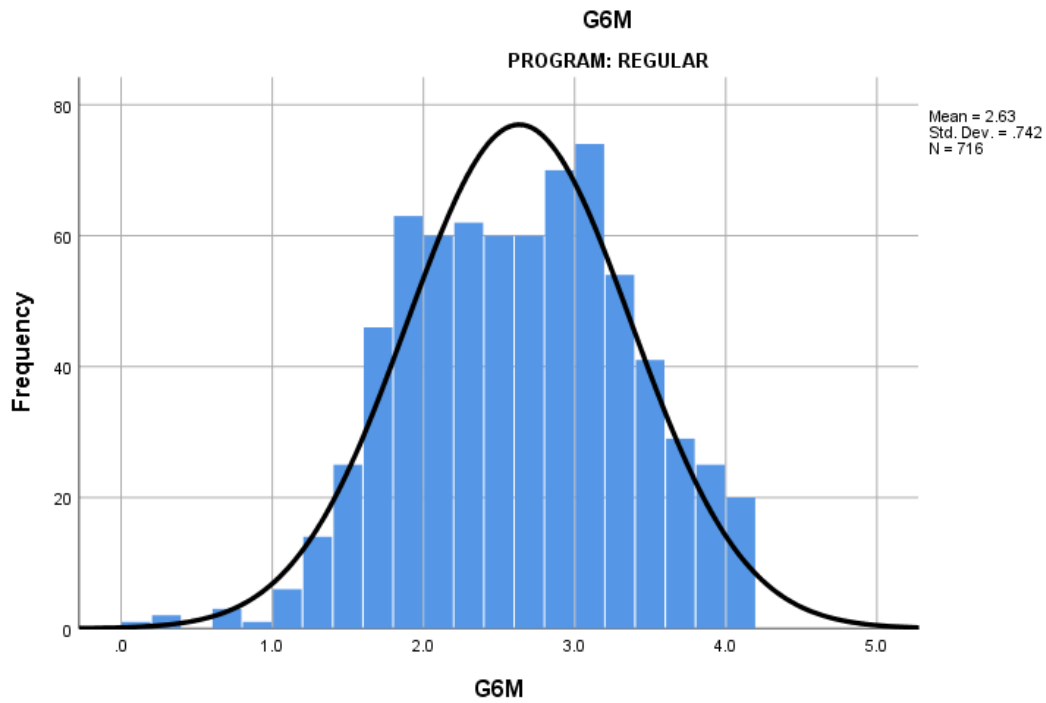
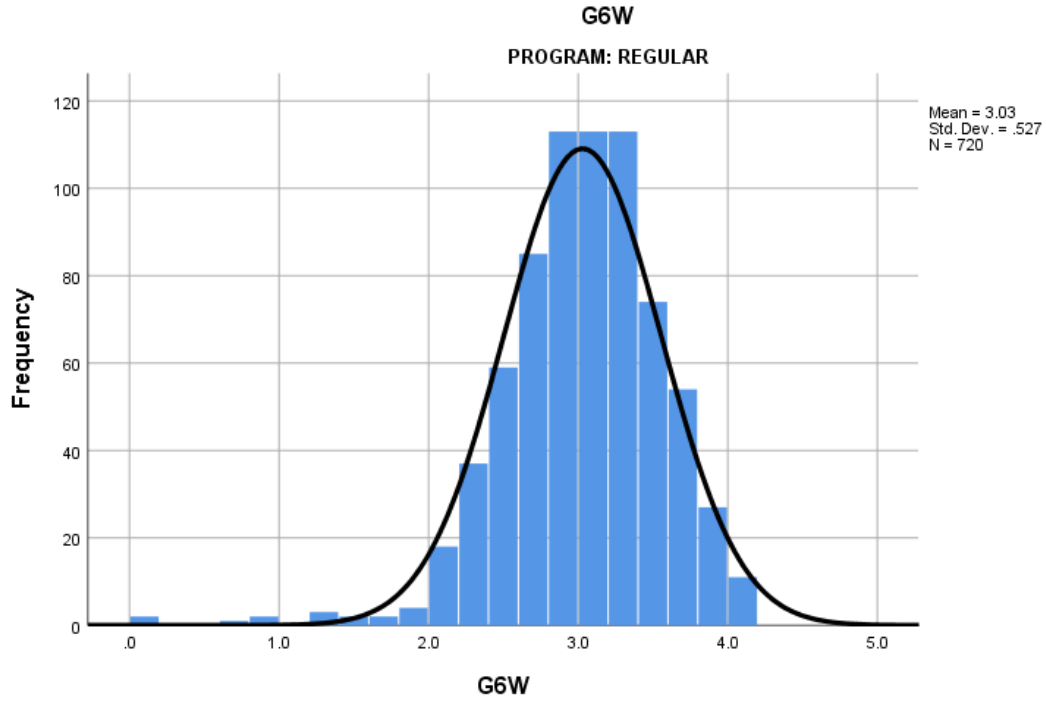






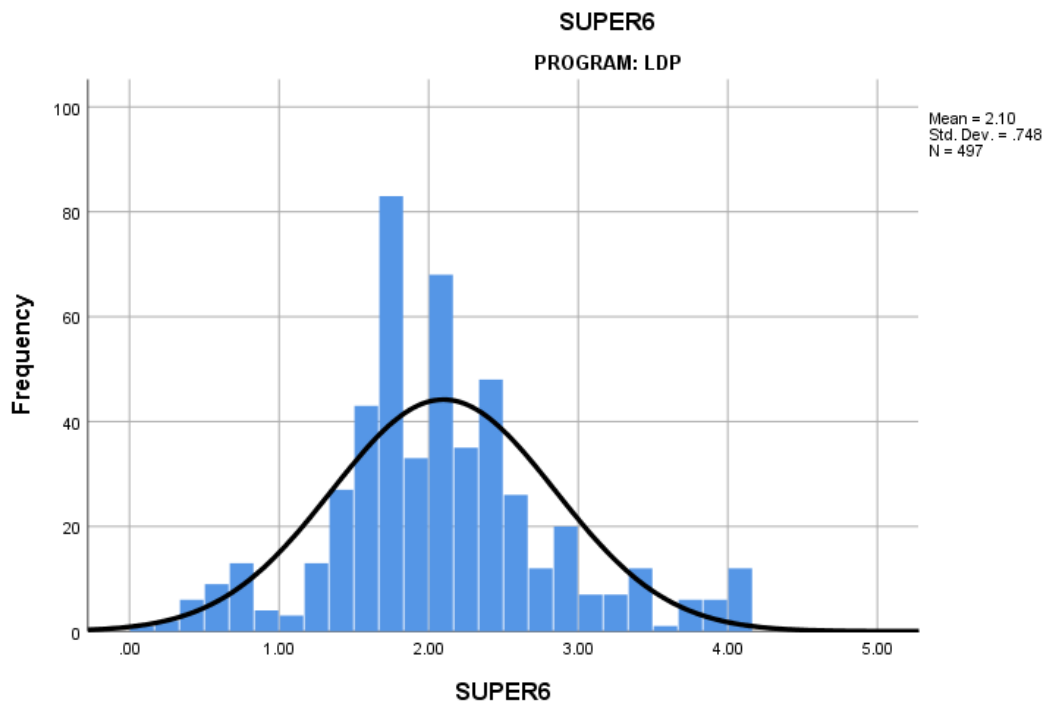
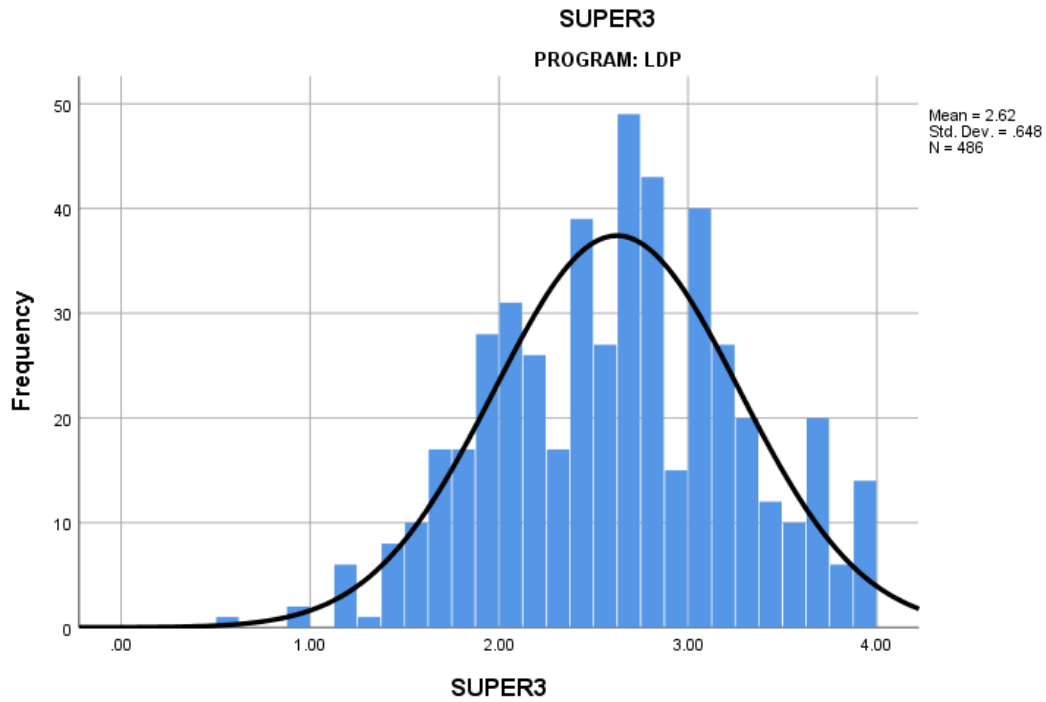


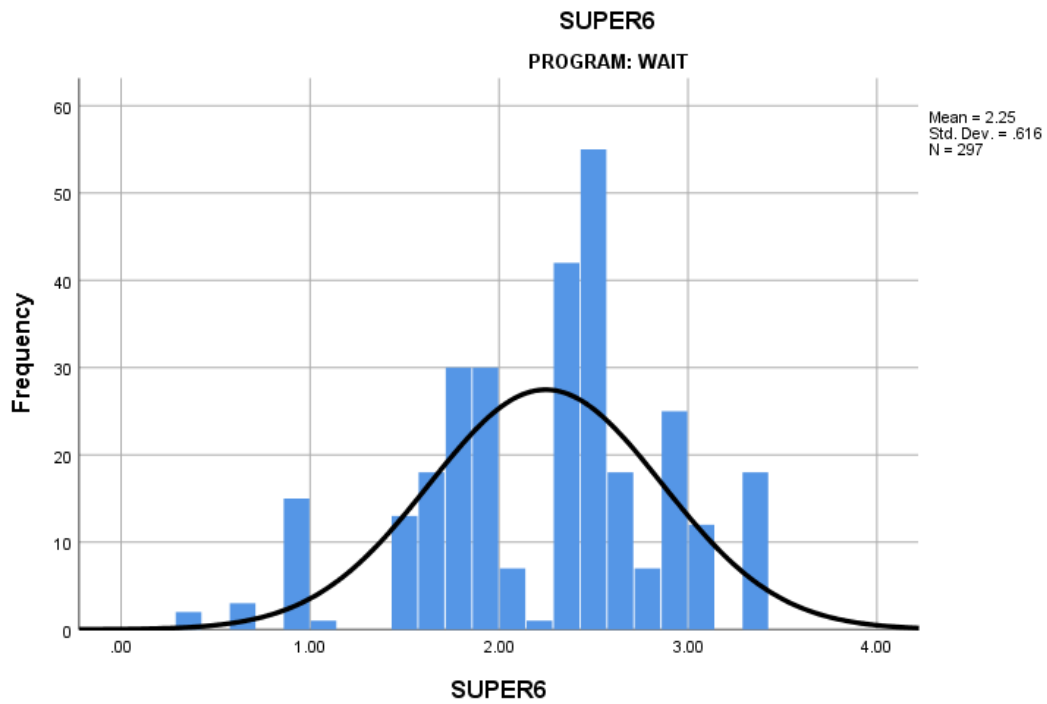
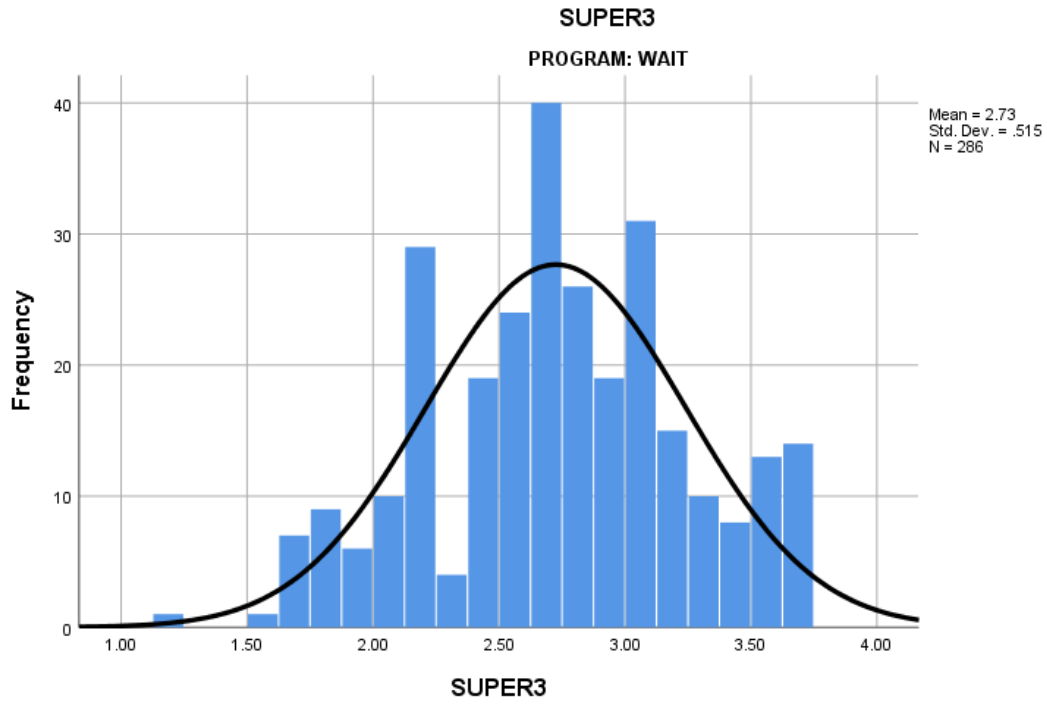


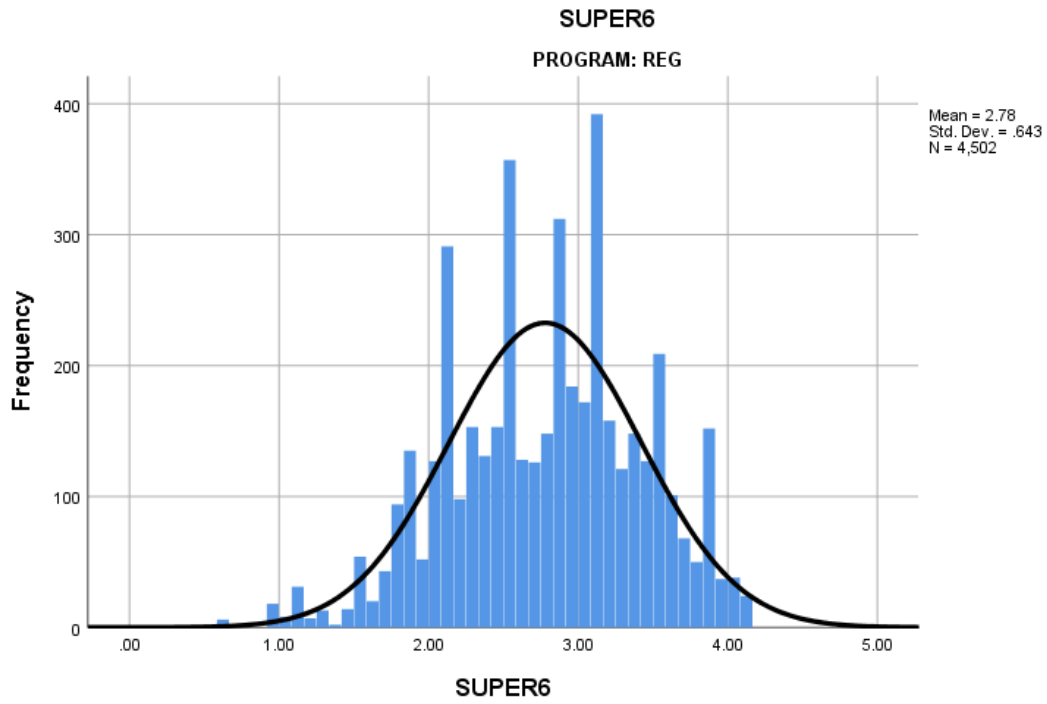
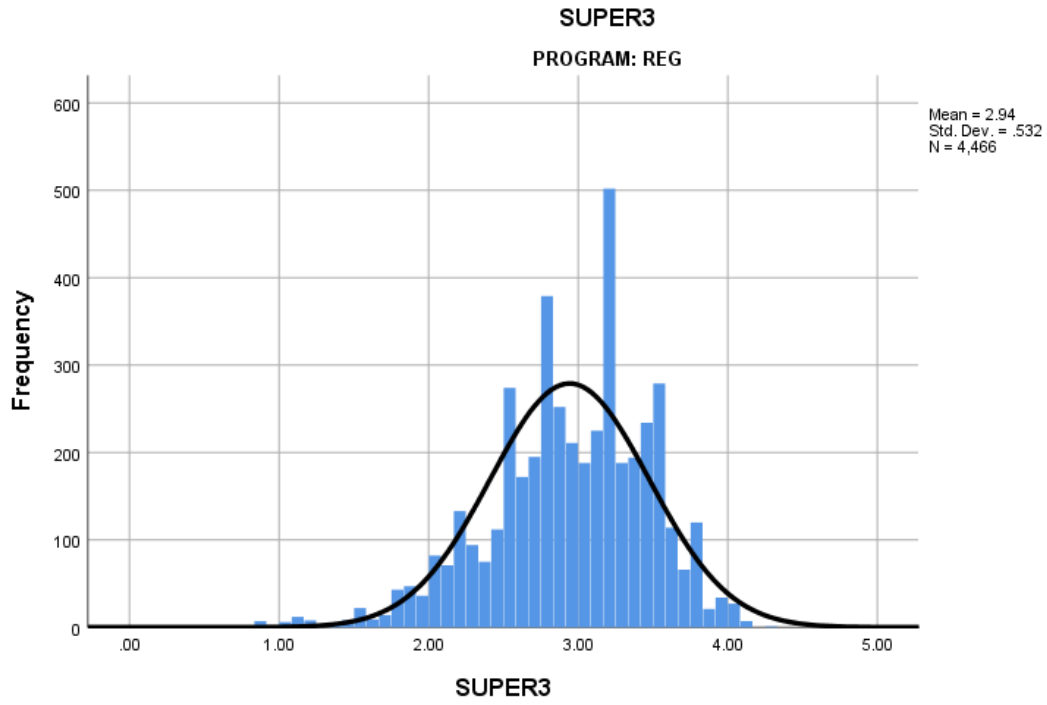


Appendix C

Histograms of the Composite Outcome Variables by Grade and Program



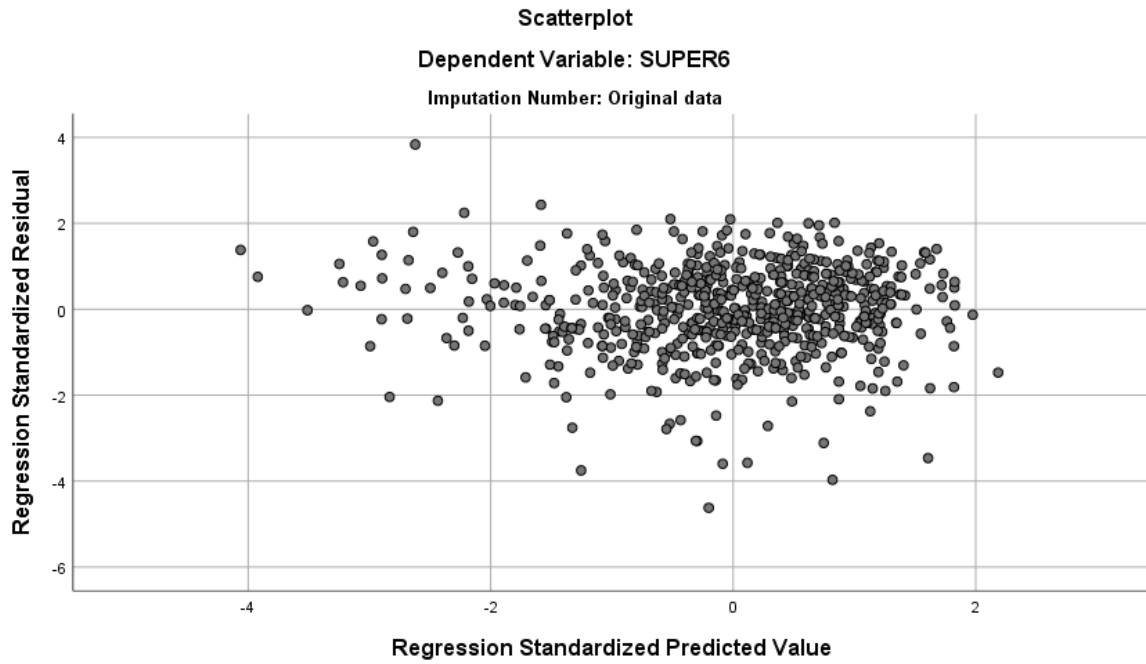




Appendix D

Figure 6

Scatterplot and P-P plot of the SUPER6 Variable in the 1st Regression Model



Normal P-P Plot of Regression Standardized Residual

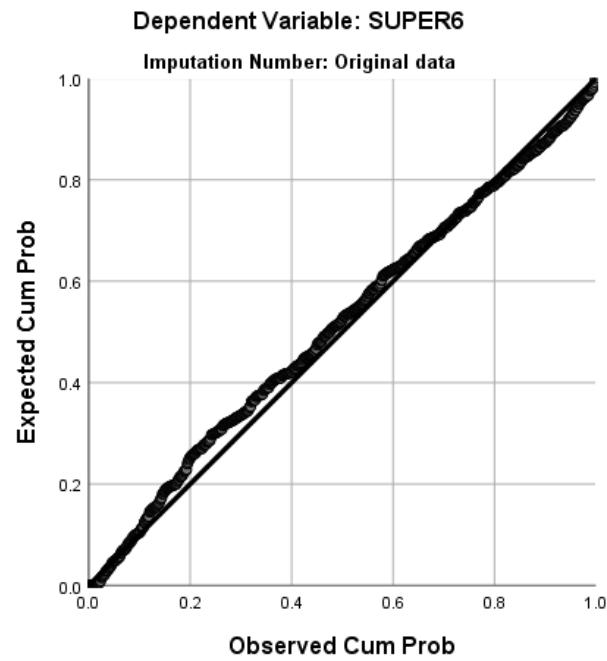


Figure 7

Scatterplot and P-P plot of the SUPER6 Variable in the 2nd Regression Model

