Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Jillian C. Gates

Entitled
Total School Cluster Grouping Model: An Investigation of Student Achievement and Identification and Teachers' Classroom Practices

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Marcia Gentry
Chair

Rebecca Mann

Aman Yadav

Yukiko Maeda

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Marcia Gentry

Approved by: Ala Samarapungavan 07/26/2011
Head of the Graduate Program Date

Graduate School Form 20
(Revised 9/10)

PURDUE UNIVERSITY
GRADUATE SCHOOL

Research Integrity and Copyright Disclaimer

Title of Thesis/Dissertation:

Total School Cluster Grouping Model: An Investigation of Student Achievement and Identification and Teachers' Classroom Practices

For the degree of Doctor of Philosophy

I certify that in the preparation of this thesis, I have observed the provisions of *Purdue University Executive Memorandum No. C-22,* September 6, 1991, *Policy on Integrity in Research.**

Further, I certify that this work is free of plagiarism and all materials appearing in this thesis/dissertation have been properly quoted and attributed.

I certify that all copyrighted material incorporated into this thesis/dissertation is in compliance with the United States' copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Purdue University from any and all claims that may be asserted or that may arise from any copyright violation.

Jillian C. Gates

Printed Name and Signature of Candidate

07/25/2011

Date (month/day/year)

*Located at http://www.purdue.edu/policies/pages/teach_res_outreach/c_22.html

TOTAL SCHOOL CLUSTER GROUPING MODEL: AN INVESTIGATION OF STUDENT ACHIEVEMENT AND IDENTIFICATION AND TEACHERS' CLASSROOM PRACTICES

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Jillian C. Gates

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2011

Purdue University

West Lafayette, Indiana

I would like to express my sincere thanks, gratitude and love to my husband, Russell Gates, and to my children, Amy, Joshua, Jonathan, Erin, Michael, and Zebedelia, for their unwavering love and support over the last five years.

ACKNOWLEDGEMENTS

I would like to thank Professors Yukiko Maeda, Rebecca Mann, and Aman Yadav for serving on my advisory committee and answering my questions over the years. They are all examples of excellent faculty mentorship.

I would like to thank my family, my friends, and my colleagues at Purdue University. Without you none of this would have been possible. I could not have completed this research without the support and mentorship you all provided.

Finally, I would like to express my gratitude to my major professor, Marcia Gentry, for her exemplary model, which I researched, her mentorship, support, and patience, during this process.

TABLE OF CONTENTS

Page

Page

Page

Page

Page

Page

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Gates, Jillian C. Ph.D., Purdue University, August, 2011. Total School Cluster Grouping Model: An Investigation of Student Achievement and Identification, and Teachers' Classroom Practices. Major Professor: Marcia Gentry.

This study involved the investigation of the effects of the Total Cluster Grouping Model on the achievement and identification of urban, elementary students and to learn about the classroom practices of and perceptions classroom teachers had of their students' ability levels over the course of the study. Proportionality of representation in clusters by ethnicity and socio-economic status was also investigated. Students from graduation years 2014 through 2017 were followed for three years from second through sixth grade. Achievement test scores in reading and mathematics were collected to assess changes in achievement between the treatment and comparison schools. Scores were analyzed by socio-economic status and ethnicity to evaluate whether this was a factor in achievement. Teachers of high-achieving clusters and other clusters were interviewed to assess their practice and perceptions of their students' ability. Students in the treatment school outperformed their comparison peers in both reading and mathematics by a small margin. Ethnicity and socio-economic status were not factors in student achievement, but were an issue when proportionality was investigated. However, the longer the students were in the program the less likely proportionality was an issue. Teacher practices influenced the

overall success of model implementation. High-achieving cluster teachers were the only teachers who received professional development in gifted education pedagogy. These teachers did not feel that the TSCG influenced their use of differentiation, but did make some changes in their practice. Other cluster teachers believed that the TSCG model enabled them to identify those who needed remediation and aided in the cross-grade grouping that occurred in reading. Identification became an issue during the study as high-achieving teachers believed that students were erroneously identified as high-achieving. This resulted in a change in the identification process, which in turn may have influenced the final wave of achievement scores.

CHAPTER 1: INTRODUCTION

Introduction

Total School Cluster Grouping (TSCG) is a school-wide programming option that provides educators with a practical, whole-school method for student placement within the classroom; educator training; and differentiation of instruction. This model differs from general cluster grouping found in much of the literature with regard to its implementation. This model, based on a talent-development model in the field of gifted education, aims to improve academic performance and achievement for all children regardless of ability level. The TSCG focuses on a child's achievement level and how educators can support and improve the innate skills and strengths of students using grouping and high-level instruction (Gentry & Mann, 2008). Total School Cluster Grouping has become popular in recent years due to education budget cutbacks that have resulted in heterogeneous grouping of students in classrooms and loss of funds for special programs such as those for gifted children (Purcell, 1994; Renzulli, 2005; State of the States, 2005).

General cluster grouping. General cluster grouping is defined as grouping high-achieving, gifted, or high-ability students in an elementary classroom with other students. Teachers who have received training to work with high-achieving students and who have the desire and ability to differentiate the curriculum teach these classes. Teachers differentiate instruction to meet the advanced academic needs of these students while also addressing the needs of other students in the classroom (Gentry, 1999; Gentry & Mann, 2008).

Total school cluster grouping. The TSCG model is a more refined, inclusive method of cluster grouping than those practices generally reported under the term cluster grouping (Gentry & Mann, 2008). The grouping process and delivery of services is designed to maximize the usefulness of the grouping of students. General cluster grouping targets high-achieving students for placement in a particular classroom; whereas TSCG considers the placement and academic performance of every child in the school and exposes all children to differentiation strategies and curricular materials commonly reserved for students who achieve at high levels. Identification of students occurs on an annual basis in order to accommodate student growth and development that may occur in different students at different times. This ongoing identification eliminates the one-time identification window that sometimes occurs in gifted education programs and implies an expectation that students will perform at higher levels over time when exposed to rigorous curriculum and teaching targeted at individual readiness levels.

Total School Cluster Grouping removes students who achieve at an above-average level from classrooms where high-achieving students are placed and distributes them among other classrooms in a grade level. This provides opportunities for the above-average achieving students to perform at high levels without having to compete with their higher-achieving peers (Gentry & Mann, 2008). In addition, some classrooms may receive clusters of students who achieve at low academic levels, and teachers are provided additional assistance within the classroom to meet the needs of these children. Groups of students who perform at below-average to average achievement levels are placed into classrooms across the grade level. The intention is to reduce the number of achievement groups placed in each classroom to three or four, while providing each teacher with a group of students who achieve at above-average levels. This conscious placement of students reduces the range of achievement levels commonly found in most heterogeneous classrooms today.

The TSCG model provides all teachers with professional development training in teaching strategies commonly used in gifted education. Educators differentiate and enrich the curriculum for all learners. All teachers in the TSCG model use similar differentiation strategies in their classrooms regardless of the achievement levels of their students. Pedagogy recommended for gifted education programs in the past is good practice for all teachers and raises student achievement at all levels (Gentry & Owen, 1999). All learners should receive enriched opportunities to extend their thinking and learning in order to increase their academic performance.

Evidence of the benefits of general cluster grouping and of TSCG models will be discussed in Chapter Two. Cluster grouping and TSCG offer high-achieving students opportunities to interact with their intellectual peers within the regular classroom and provides virtually no-cost, full-time, gifted education services; whereas, pull-out programs alone only offer part-time educational services (Purcell, 1994). The TSCG model provides continuous services to students who exhibit gifted behaviors, and has also been shown to work well in conjunction with pull-out enrichment programs as well as between-classroom and across-grade, flexible grouping strategies (Gentry & Mann, 2008).

Study Context

This study examined the effects of the TSCG model in a diverse, Midwestern, urban, public school district. School district administrators purposively selected one school as a treatment school and one as a comparison school. Educators at the comparison school were not aware of their school's status as a comparison school. School administrators chose the treatment school for the study as teachers and the administrator were implementing the TSCG model. The comparison school was matched to the treatment school based on similarities in size, demographics, and make-up based on administrator judgment and district data. Teacher perceptions and practices were explored as well as issues of administrator support of the model. The identification and achievement of students from the Classes of 2014 through 2017 over three years was examined, as were the practices and perceptions of classroom teachers in high-achieving cluster classrooms and other-cluster classrooms.

Statement of the Problem

Meeting the needs of students who achieve at high levels has been an issue for many years. Budgets and staffing for special programs are often limited (Gentry & Owen, 1999). Coupled with these issues, education has moved away from ability or achievement grouping towards total inclusion in an attempt to be fair to all students and to meet the demands of No Child Left Behind (21 USC § 1401a, 2001) regulations. These changes in policy have left parents and educators of high-achieving students concerned about the lack of appropriate educational services. Gifted education services have also been charged as elitist and discriminatory as students from low-income and/or culturally-diverse families tend to be underrepresented in these programs (Ford, 1998; Yoon & Gentry, 2009). Additionally, gifted education programs are often not culturally sensitive, which also results in attrition of culturally diverse students as they feel that their needs are not being met within the program (Ford).

The Total School Cluster Grouping model has become increasingly popular as the research has shown that achievement scores of all students in a cluster-grouped school increased significantly over those of their peers in non-clustered schools (Gentry & Owen, 1999). This finding has implications for meeting the requirements of No Child Left Behind (21 USC § 1401a, 2001); thus, school district personnel have adopted TSCG as a means of complying with this law.

Total School Cluster Grouping not only provides educational services to those students, who achieve at high levels, but also considers the achievement levels of all students and meets their educational needs. This model reduces the concern over elitism as all students are grouped and exposed to the same pedagogy – that commonly used in gifted education programs. Therefore, all students have an equal opportunity to learn at high levels and increase their academic performance. Total School Cluster Grouping also reduces the problems of under identification as research has shown that students from low-income and/or culturally diverse families are more likely to be identified as high-achieving, over time, when the TSCG model is used (Gentry, 2008).

Purpose of the Study

The purpose of this study was two-fold: first, to investigate the effects of TSCG on the achievement and identification of urban elementary students; and second, to learn about the perceptions classroom teachers had of their students' ability levels over the course of the study. In 2007 Total School Cluster Grouping was implemented in the treatment school selected for this study. Initial identification data were collected for placement in spring 2006. These data were also collected in spring 2007, 2008, and 2009 and were analyzed for trends in identification. Total School Cluster Grouping was implemented in fall 2006 and achievement measured three times per year starting in spring 2007. Final achievement data were collected in Spring 2009. The following research questions guided this study:

1. What effects does TSCG have on student achievement as measured by the NWEA**,** specifically:
    a. What differences exist between treatment and comparison student growth curves among four groups based on income and representation[1] status: (1) low SES/underrepresented; (2) low SES/represented; (3) not-low SES/underrepresented; (4) not-low SES/represented.
    b. Do treatment school students in a TSCG program out-perform matched comparison school students after three years in the program after controlling for SES and ethnic representation?

---

[1] To account for the correlation between ethnicity and SES, I collapsed ethnic groups into underrepresented (African American, Hispanic, Native American) and represented groups (White, Asian, Other), using categories of (1) low SES/underrepresented; (2) low SES/represented; (3) not-low SES/underrepresented; (4) not-low SES/represented.

  c. How do the learning trajectories differ between treatment and comparison students?
  d. For treatment school students only, based on initial identification categories, how do student achievement growth trajectories differ over three program years?
  e. For treatment school students only, using four groups of students based on income and representation status, how do student achievement growth trajectories differ over three program years?

2. What effects does TSCG have on identification categories?
  a. What changes occur in frequency of students in each of the five achievement groups?
    i. by income?
    ii. by ethnicity?

3. What factors exist within the classrooms and the school using TSCG that may influence student achievement?

Significance of the Study

This study adds to the research literature concerning cluster grouping and more specifically, the TSCG model, in five ways. First, this study extended the work of Gentry and Owen (1999) by examining the achievement gains for high-achieving students, something that Gentry and Owen were unable to do given the limitations of the available instruments at the time they conducted their study. This study used the NWEA (2005) to measure achievement, a computer-based, adaptive test with a high, test ceiling. Results from this test yielded more detailed results of student achievement as students could score above grade-level therefore growth trends were not muted by test limitations (NWEA). Second, this study was conducted in an urban setting, which differed from Gentry and Owen's rural setting. Third, teachers were interviewed using the same protocol used by Gentry and Owen; however, three additional questions were added regarding teacher perceptions of ability levels of student from low-income and/or culturally-diverse families. These questions explored perceptions regarding student performance in light of socio-economic status and ethnicity to inform the identification data. Fourth, identification trends of students according to ethnic groups and socio-economic levels were investigated quantitatively. Fifth, these questions were addressed in a longitudinal context over the course of three years and employed a multi-level model of change to examine student achievement. Therefore, this study examined the TSCG model in an urban setting in a school district with a diverse student population enabling modeling of the data to describe the TSCG model's effect on the racial and socio-economic groups of the sample.

I examined the quantitative change in achievement and identification based on the implementation of the TSCG model, qualitative factors that exist within the classrooms and school using TSCG, and teacher perceptions of students' ability levels since implementing this model. This mixed-methods approach to studying the TSCG model not only provided empirical data on achievement and identification of students, but also provided contextual information regarding factors that may have influenced these changes. The qualitative investigation provided additional information to support the empirical data. Delcourt et al., (1994) noted that programming without appropriate infrastructure results in no programming at all. A purpose of this study was to investigate what infrastructure is important and useful in supporting the TSCG model in a diverse, urban setting.

CHAPTER TWO: LITERATURE REVIEW

Grouping students within the classroom happens as a matter of course throughout the day. These groups can be student-initiated or teacher-initiated. Grouping occurs within the classroom or across several classrooms within a grade-level or across grade levels. Research has been done regarding the usefulness of grouping as well as concerning the academic and social outcomes associated with different kinds of grouping. Grouping can be differentiated by where they occur and how groups are composed.

Ability Grouping

The potential academic benefits of ability grouping have received attention in the research literature for over fifty years. Ability grouping refers to the process of grouping students according to specific criteria such as ability, achievement, or skill (Rogers, 1991). There are several forms of ability grouping such as within-class ability grouping and between-class ability grouping (Slavin, 1987). Within-class ability grouping occurs when teachers group students homogeneously within a heterogeneous classroom in order to adapt instruction to meet the specific needs of particular groups (Kulik, 1992; Slavin, 1987; Loveless, 1998). Ability grouping for reading instruction in the early grades is an almost universal phenomenon (Loveless).

Between-class ability grouping occurs when teachers and/or administrators group students into separate classes based on perceived or demonstrated abilities and prior knowledge (Kulik, 1992). A common form of between-class ability grouping in secondary education is known as tracking (Ireson & Hallam, 1999; Oakes, 1985), although grouping is becoming more flexible as the research indicates that fluidity of movement between groups is necessary to allow student growth (Gentry & Owen, 1999; Gentry & Mann; 2008; Lou, Abrami, Spence, Poulsen, Chambers, & d'Apollonia, 1996; Lou, Abrami, & Spence, 2000). Other forms of between-class ability grouping include regrouping for a particular subject and cross-grade grouping for subject acceleration or remediation also known as the Joplin plan (Floyd, 1954). Research concerning cross-grade grouping in reading is abundant (Anastasiow, 1968; Kulik, 1992; Kulik & Kulik, 1992; Slavin, 1987; Tieso, 2003) and much of the research analyzed in meta-analyses occurred in the reading classroom (Kulik, 1992; Kulik & Kulik, 1984; 1992; Slavin, 1987). However, less research has been done on the effects of this same grouping in mathematics or other subject areas (Kulik, 1992; Leonard, 2001; McSherry & Ollerton, 2002; Tieso, 2005).

Opponents of ability grouping argue that it is elitist (Black, 1993; Oakes, 1985) and results in lower expectations for members of the low-ability groups, which results in lower achievement (Black; Oakes). Opponents also argue that ability grouping may deny students in low-ability groups the opportunity to access more challenging material, which widens the gap between high-ability and low-ability groups (Oakes). Gallagher (1993) noted that tracking and ability grouping were used as interchangeable terms in much of the early research, hampering analysis of study results. Another concern is the influence ability-grouping may have on student self-concept (Black). Findings regarding self-esteem and grouping will be discussed in the following meta-analyses section.

Evidence from Meta-Analytic Studies. Several meta-analyses on grouping have been written in the last 24 years. Some authors have concentrated on a particular level of schooling such as elementary or secondary school. Some have chosen to include the complete range of education from elementary to tertiary education. Six meta-analyses were chosen for review, because they either concentrated on elementary grades pertinent to this study (Elbaum, Vaughn, Hughes, & Moody, 1999; Kulik & Kulik, 1984; Slavin, 1987) or included studies done in either elementary and secondary grades, or tertiary education (Kulik, 1992; Kulik & Kulik, 1992; Lou, Abrami, Spence, Poulsen, Chambers, & d'Apollonia, 1996; Lou, Abrami, & Spence, 2000). Meta-analyses focusing only on secondary or tertiary education are excluded as grouping effects may differ for secondary students. The findings of meta-analyses discussed here are summarized in Table 1.

Tracking or XYZ grouping. Kulik and Kulik (1984) conducted a meta-analysis on the effects of ability grouping in the elementary school setting. They found thirty-one studies that investigated the effects of ability grouping on student achievement and self-concept. This meta-analysis investigated the effects on elementary students' achievement when they are grouped into separate classes by differing ability levels. This phenomenon is called tracking or XYZ grouping (Oakes, 1985; Rogers, 1991). Within-class grouping, ability-grouped schools, grouping for acceleration, and non-graded grouping were not investigated in this analysis.

Table 1

Meta-Analyses Findings

Year	Author(s)	# of Studies	Type of Grouping vs. Whole Class Instruction	Effect Size
1984	Kulik C-H.C., & Kulik, J. A.	31	XYZ Grouping (Ability-grouped classes)	+.19
			Self-contained gifted XYZ grouping	+.49
			XYZ grouping on self-esteem	-.06
1987	Slavin, R.E.	14	XYZ grouping	0.0
		7	Subject-specific XYZ grouping	unclear
		14	Cross-grade ability grouping for reading	+.45
		5	Cross-grade ability grouping for mathematics	+.32
1992	Kulik; Kulik, J. & Kulik(a)	51	Test-based ability grouping – gifted students	+.10
			Test-based ability grouping - general students	0.0
		14	Cross-grade ability grouping by subject	+.33
			Cross-grade ability grouping – gifted	+.12
			Cross-grade ability grouping – average	-.01
			Cross-grade ability grouping – low	+.29
			Within-class grouping	+.25
			Acceleration classes compared to age-peers	+.75
			Acceleration classes compared to grade-level peers	0.0

Year	Author(s)	# of Studies	Type of Grouping vs. Whole Class Instruction	Effect Size
			Enrichment classes	+.41
1996	Lou, et al.	51	Within-class grouping and differentiation	+.25
			Within-class grouping no differentiation	+.02
1999	Elbaum, Vaughn, Hughes, & Moody	20	LD students grouped or paired in reading	+.43
			LD students paired in reading	+.40
2000	Lou, Abrami, & Spence	51	within-class grouping	+.16

Overall results of analysis of studies investigating grouping and student achievement were positive. Kulik and Kulik (1984) noted that 20 out of 28 studies yielded positive effect sizes for grouping students by ability. Effect sizes in 13 of these studies were statistically significant. Eleven of those 13, favored homogenous grouping and two favored heterogeneous grouping. The average effect size for achievement was +.19. However, when results for classes of gifted students were examined, the effect size rose to +.49 with effect sizes for the general population at +.07.

Ability grouping in the elementary, general education setting. Slavin (1987) also conducted a meta-analysis of ability grouping on student achievement in elementary schools. The major difference between his analysis and that of Kulik and Kulik (1984) was that Slavin did not include any studies regarding grouping for gifted or special education students. His contention for excluding these studies was that both the pedagogy and curricula for students in these two groups differs fundamentally from that used for the general education population and would therefore account for much of the variance in achievement, rather than the grouping itself.

Slavin (1987) used his best-evidence synthesis method in his meta-analysis of ability grouping studies. Derived from his own theory, this method incorporates the elements of narrative analysis to provide the reader with detailed information on all studies chosen as well as empirical information such as effect sizes. Rigorous a priori criteria regarding inclusion of studies are used and adhered to in order to increase the rigor of the analysis (Slavin). Slavin noted that when effect sizes could not be computed, he characterized the effects as positive, negative, or zero rather than excluding them.

Slavin (1987) analyzed four components of ability grouping in elementary schools: ability-grouping by class; within-grade level grouping; Joplin Plan grouping, and within-class ability grouping. Fourteen studies investigating ability-grouped classes were included in this analysis. Of the fourteen studies, 13 were matched comparison studies with good evidence of equivalency. The remaining study was a randomized study. Slavin's analysis of ability grouping by class yielded no conclusive results. He reported an effect size of zero for student achievement overall and inconclusive results for studies in the content areas of mathematics and reading.

Joplin Plan studies were analyzed as this unique grouping plan had become very popular for reading instruction. Fourteen studies were included with positive effects (ES = +.45) on student achievement. Although Joplin Plan grouping is typically used for reading instruction, Slavin included one study of Joplin Plan grouping in mathematics with an effect size of +.46 SD for student achievement.

Finally, within-class grouping studies were analyzed. Research was limited to mathematics in the upper elementary grades so results from this analysis cannot be generalized to all elementary grade levels. However, of the five studies included, positive gains in student achievement were recorded for within-class grouping in mathematics (ES = +.32). Positive effects were larger if groups were kept small (three to five students).

In addition to reporting findings of studies, Slavin (1987) tentatively suggested some hallmarks of effective grouping: students should remain in heterogeneous groups for the majority of their education – ability grouping should be targeted for certain subjects; grouping should be flexible; differentiation of instruction and pace should occur for each ability level; and groups should be kept small for within-class grouping.

James and Chen-Lin Kulik (1992) wrote an historical and meta-analytic review of ability grouping. They revisited both their 1991 meta-analysis, and Slavin's findings (1987; 1990), which used a number of the same studies. Of the 143 studies that were used in the original meta-analyses (Kulik & Kulik, 1992; Slavin, 1987; 1990) only 127 were retained for this analysis. One of the studies rejected was used in both studies, while 9 had been used only by Kulik and Kulik (1992) and 16 only by Slavin (1987; 1990). These studies were rejected as their inclusion in the previous studies seemed to be due to idiosyncratic choices made by the authors and could not be justified at the time of the second analysis. Kulik and Kulik (1992) recoded the studies that remained using the previous coding as a guideline. However, given the subjective nature of coding and study choice in meta-analysis, the effect sizes from Slavin's and Kulik and Kulik's analyses were not perfectly correlated with each other (Glass, 1976). Effect sizes for the Slavin analysis were correlated .89 with the Kulik's findings, while effect sizes for the Kulik and Kulik were correlated .97 with their previous findings.

Kulik and Kulik's (1992) meta-analytic findings were different for different types of grouping. Overall, students placed in groups based on test scores and school reports with little curriculum change scored higher on achievement tests than their peers in mixed-ability groups. Of note was the finding that students in middle and lower groups under this condition did not achieve better than their peers in mixed-ability groups. However, students in the top ability group outperformed their intellectual and age peers in mixed-ability classrooms (ES=.10).

Kulik and Kulik's (1992) findings contradict Slavin's (1987) findings. None of the grouping criteria suggested by Slavin significantly influenced Kulik and Kulik's findings in any of the 51 studies analyzed for this type of grouping. In addition to the achievement findings, Kulik and Kulik found that self-concept was more positive for students in the low groups, whereas it decreased slightly for students in the upper group. What is not clear in this finding was how long after grouping the self-concept assessment was completed. Therefore, self-concept issues may be short-lived for students or might persist longer. Clearly, there may be marked differences in self-concept between students of different ability levels in heterogeneous classrooms (Kulik & Kulik, 1992); however, these differences seem to be mediated when students are grouped homogenously.

The clearest differences in student achievement can be seen in instances when students are grouped according to ability and the curriculum is adjusted to the particular ability level. Kulik and Kulik (1992) found that students in all levels of grouping outperformed equivalent comparison groups. Students grouped across-grade for a particular subject (i.e., Joplin Plan grouping) performed higher (ES = +.33) than students who were not grouped. Only 14 studies were found for this analysis with a range of effect sizes from -.03 to +.98. Two studies reported achievement results by ability. The effect size for high-ability students was +.12, average-ability -.01, and low-ability +.29. Students grouped within-class showed similar gains to those in cross-grade grouping. The average effect size for achievement was +.25.This finding is significantly different from zero and from ability-grouped classes, but not significantly different from cross-grade grouping.

Finally, Kulik and Kulik (1992) found that high-ability students in enrichment and accelerated classes far outperformed their comparison peers with acceleration yielding achievement gains of almost an entire grade level (ES=+.75) when compared to their age-peers and an effect size of zero when compared to peers a year ahead of them. This indicates that when students are accelerated appropriately they tend to perform just as well as their class peers.

Enrichment classes tended to deal with content outside the curriculum or in areas of student interest, rather than accelerating students. In this analysis, enrichment classes yielded achievement gains of +.41, a moderate, positive gain. Enrichment positively affected student achievement in the academic arena.

Kulik and Kulik's (1984, 1992) meta-analyses differed from Slavin's (1987) earlier meta-analysis in several ways. Slavin argued that four factors contributed to the usefulness of grouping practices: curricular differentiation; flexible grouping; method of group assignment; and extent of grouping. However, Kulik and Kulik (1992) found that effect sizes did not differ significantly for XYZ groupings with or without these features.

Slavin further contended that effect sizes were greater in less rigorous studies and that this affected results in meta-analyses. His assertion was refuted in later meta-analyses (Kulik & Kulik, 1992; Lou et al., 1996; Lou et al., 2000).

Slavin's work (1987, 1990) on both elementary and secondary grouping did not include studies in which students were designated gifted or special needs. His contention was that classes designed for these students are by nature different in content and teaching style than those of the regular classroom. However, subsequent researchers (Kulik & Kulik, 1992; Rogers, 1991) have argued that different classrooms inherently have different curricula and teachers have different teaching styles. These differences posited by Slavin do not occur only in special programs. Kulik and Kulik (1992) provided evidence to support these arguments in their later inclusion of studies of students identified as gifted or receiving special education services.

Small-size, within-class grouping. Lou, Abrami, Spence, Poulsen, Chambers, & d'Apollonia (1996) conducted a comprehensive meta-analysis on grouping. Their analysis investigated the overall effect size of small-size, within-class grouping versus no grouping. Studies were included from the elementary, secondary, and tertiary setting. They examined 51 studies that included all but six of the studies previously analyzed by Slavin (1987; 1990) and Kulik and Kulik (1984; 1992). Similar to the findings of Kulik and Kulik (1992), Lou et al. reported that greater effect sizes were found for achievement when curricular changes were made in addition to grouping (ES=+.25) versus simply grouping students (ES=+.02).

In a meta-analysis of this size, variability between study outcomes is inevitable. However, Prais (1998) in his rebuttal to Lou et al.'s (1996) meta-analysis used the variability in study results as a reason to invalidate their findings. Therefore it is important to note that the authors conducted an homogeneity test, the results of which showed that of 103 comparisons analyzed in 51 studies only 74 showed an effect size larger than zero when effect sizes were corrected for bias and weighted for sample size. Five studies had an effect size of zero, with an additional 24 studies with negative effect sizes. Given this information Lou et al. investigated factors that moderate the effects of within-class grouping on student achievement. Their inquiry yielded important factors described below that should be considered when grouping students and analyzing studies regarding grouping.

Teacher-made tests and researcher-made tests yielded higher effect sizes. These effects were increased when tests were written to address specific content outcomes. Along with this finding Lou et al. (1996) noted that there was no difference in effect sizes between heterogeneous and homogenous grouping. Teacher training, equivalence of materials, and reward equivalence between treatment and comparison groups also influenced effect sizes, although the difference in overall design quality did not influence results. Greater effect sizes were found in studies in which students were grouped and offered different teaching strategies; teachers were trained differently from their comparison counterparts; and different curricula materials were used. Another factor that influenced effect size in this analysis was rewards. When different rewards from comparison group rewards were used in the treatment groups, positive effect sizes were greater.

Lou et al. (1996) also found that there was a significant difference in outcomes for groups that worked cooperatively versus those that were outcome independent - that is students working in a group, but producing individual work. Independent-outcome grouping was also superior to traditional, teacher-directed learning or individualized learning-for-mastery, but was not superior to whole-class, experiential learning.

Lou, Abrami, and Spence (2000) increased the rigor and extended the findings of Lou et al.'s (1996) meta-analysis by re-examining their earlier analyses and conducting a multivariate analysis. This re-analysis was done to attempt to explain and account for the variability in the effect sizes of within-class grouping on student achievement in reaction to Prais' (1998; 1999) rebuttals to their previous findings. Lou et al. (1996) had coded studies according to three different types of study features: outcome, methodological, and substantive features. Forty-three features were initially identified and then collapsed into 26 features for analysis. Lou et al. conducted a series of 26 analyses of study features with the homogeneity approach. Significance of between-group chi-square tests by individual feature indicated that effect sizes were moderated by that particular feature. Lou et al. (2000) conducted a least-squares, multiple regression analysis using 24 of the 26 study features from Lou et al. (1996). Model testing was completed in two ways. Lou et al. produced an initial least-squares regression model using the 17 univariately significant features from the previous study. This analysis identified which features accounted for significant, unique variances in the findings. Additionally, Lou et al.'s modeling investigated if these 17 univariate features accounted for the non-random variance in the findings. The second model attempted to define a parsimonious model using a hierarchical multiple regression analysis using only 24 of the coded study features. Study features were then entered in blocks in a step-wise procedure in order to account for variance within the model. Only features that explained significant additional variance were retained in the model. After removing five outliers, the authors determined a weighted, mean effect size of +.16. This was a small, but significant positive effect on student achievement for within-class grouping over whole class instruction.

Since this analysis was performed in a step-wise manner, Lou et al. (2000) were able to specify which study features accounted for the most variance in the model. Forty-eight percent of the variance was accounted for by: the source of the outcome measure; equivalence of teacher training between treatment and comparison groups; specificity of grouping; type of instruction given to small groups; grade; and ability of students relative to each other. From these data Lou et al. stated that although small-group ability grouping is beneficial to all students, it is most beneficial to those of higher ability, and much more so for elementary students than high school or university students. In addition, they concluded that effect sizes were significantly greater for outcome measures that were locally developed than for standardized tests. Finally, Lou et al. found teacher training a significant factor in the success of grouping. Teacher training resulted in teachers using different grading and reward strategies from their peers teaching whole group classes, especially when cooperative teaching strategies were used. This training significantly increased the effect of grouping for students. Although training might indicate non-equivalence between treatment and comparison groups, it does highlight the importance of teacher training in the implementation of ability grouping.

Nomi (2010) conducted a study of within-class ability grouping on achievement in the early elementary years. Data from the Early Childhood Longitudinal Study were used and a propensity school analysis conducted to determine whether ability grouping increases student achievement, student achievement inequalities, and whether these effects differ by school context. Her findings indicated that ability grouping either negatively affects or does not affect student achievement in low-performing public schools with students from low-socioeconomic, minority families. Ability grouping positively affects the achievement for students in schools with students from less diverse and higher-economic families, particularly private schools. In addition, achievement gaps are narrowed in this type of school as ability grouping most benefits lower-ability students.

Grouping as a socio-cognitive process. Grouping is both a social interaction between peers as well as an academic endeavor. If one takes a socio-cognitive perspective on grouping it could be said that learning in groups is a cognitive process that takes place in a social context (Rogers, 1987; Tierney & Rogers, 1989). There are both academic and social outcomes associated with grouping students for learning. Using this perspective, grouping becomes a much more complex process than merely grouping students together. Students become acculturated to their groups and the group members' learning styles. They learn their roles within the groups as well as the teacher practices associated with membership of that group. The authors concluded that evaluating grouping in isolation from the context and process occurring in and around those groups results in a misunderstanding of the importance of grouping as a classroom practice.

The authors of the TSCG model suggest that all students, no matter their group association, should experience instructional practices similar to those in the high achieving cluster (Gentry & Owen, 1999; Gentry & Mann, 2008). Only when students are exposed to the same pedagogy and classroom practices can the efficacy of clustering be adequately evaluated. This is not to say that students will all be exposed to the same materials, but the model does stipulate that all children should be exposed to high-level questioning, opportunities to problem-solve in a real world context, and opportunities to explore information in a personally meaningful manner, whatever their academic level of performance (Gentry & Owen; Gentry & Mann).

Summary. As with any data analysis, results depend upon the researcher's application of the method. Kulik and Kulik (1992) and Slavin (1987) used many of the same studies, but reported different results and interpretations of them. These differences stem from a difference in the breadth of studies used. Kulik and Kulik (1984; 1992) used studies from all levels of ability; whereas, Slavin (1987; 1990) only used studies that involved average students.

The consistent results of meta-analyses of grouping support the conclusion that ability grouping has some effect on student achievement. Effect sizes ranged from +.75 for acceleration of gifted students to -.01 for cross-grade grouping of average students. This negative effect is trivial; therefore, students who are placed in some sort of ability group achieve as well as or better than their peers who are instructed in heterogeneous classrooms. In addition, when curricula and pedagogical changes are added to grouping, positive effects are increased.

Students who exhibit gifted behaviors tend to achieve better than their age peers, who also benefit, when grouped by ability (Kulik & Kulik, 1984, 1992). Average- and low-ability students achieve higher than their age-peers in all grouping types, except for XYZ grouping or tracked classes (Kulik & Kulik; Lou et al., 1996; 2000; Slavin, 1987). In this type of grouping, their performance is indistinguishable from that of their peers in heterogeneous classes.

When results regarding self-concept were analyzed, these meta-analyses showed that although XYZ grouping may not benefit students academically, it does not have the negative emotional effects that Oakes (1985) suggested (Kulik & Kulik; Lou et al., 1996; 2000; Slavin, 1987). Students of high ability in XYZ-grouped classes tended to show a small decrease in self-concept compared to average- and low-ability students who showed an increase in self-concept (Kulik & Kulik, 1992). These data come from a small number of studies as little research was found concerning affective effects of grouping.

Overall, while XYZ grouping may not be the most effective, it is clear that most forms of ability grouping led to increased student achievement. Grouping provided better academic gains for students than heterogeneous, whole-class, instruction for students of any ability level. Therefore, whole class instruction mediated by some form of grouping and targeted instruction for at least part of the instructional day is productive. This does not run counter to the inclusive classroom (Mastropieri & Scruggs, 2000), which includes students of all levels of ability as well as those with significant learning challenges, but offers a means of providing responsive instruction to students with different learning needs.

Rationale for and realities of grouping. Davies, Hallam, and Ireson (2003) explored the rationale given by schools for the adoption of grouping practices and investigated how such practices were managed, resourced, and how they have changed over time. Students were interviewed about their perceptions of how they were placed in groups. Six schools, varied in population and geographic location in England were chosen in order to minimize regional bias. Davies et al. found that allocation to groups was done based on both locally-made and standardized tests. Some schools also used supplemental data. They found some difficulties associated with group allocation by test score. For instance, behavior must be considered in group composition to minimize difficulties. Smaller group size for lower-ability students was also important. Teachers found that keeping group size small was difficult due to resources and physical space. Parental pressure was also an issue as parents did not always agree with group placements.

Davies et al. (2003) found that monitoring of group arrangements was done using testing and analysis of performance - both formal and informal. This was done more frequently in mixed-ability classes, but record keeping of who was in what group in the mixed-ability classes was nonexistent. Difficulties in moving students were due to lack of space in groups and minimized the flexibility of grouping. This proved to be the biggest obstacle to grouping for teachers. The sole use of testing for grouping resulted in students from low-income and culturally/or families consistently over-represented in low-ability groups (Davies).

Grouping in the mathematics content area. Grouping has been a much-debated topic in the field of mathematics education. Central issues for those who support grouping in mathematics are the linear nature of mathematics content and the differences in ability and rate of learning often found in students who learn mathematics (Ruthven, 1987). Kulik and Kulik (1992) noted that there were not many studies on grouping in specific content areas such as mathematics and reading. However, since their meta-analysis several studies have been conducted in both of these areas in elementary, middle, and high school settings. A search of multiple databases found 33 studies dealing with mathematics ability grouping in middle and secondary schools since Kulik and Kulik's

(1992) meta-analysis. Studies dealing only with homogenous grouping in elementary mathematics classes were not found, but comparison studies by McSherry and Ollerton (2002) and Leonard (2001) were located. A second search using studies cited in the National Mathematics Advisory Panel final report yielded three articles that investigated cooperative or collaborative learning versus individual learning in the elementary setting (Hurley, Boykin, & Allen, 2005; Peklaj & Vodopivec, 1999; Stevens & Slavin, 1995). Although these researchers did not directly investigate grouping by ability or achievement they did find results pertaining to the benefits or problems associated with students working together or alone. In addition, Tieso (2005) looked at both grouping and curricular modifications in elementary mathematics classrooms in New England. She found that grouping alone did not yield significant increases in achievement. Curricular modification was necessary to increase achievement in addition to grouping.

Homogenous grouping versus heterogeneous grouping. Leonard (2001) conducted an action research study in her sixth-grade mathematics classroom over a two-year period. During the first year, students were grouped heterogeneously for all mathematics instruction. During the second year students were grouped homogenously according to pre-test results. Although the year 1 and year 2 classes were different, they were relatively similar in composition over the two years of the study, reducing possible bias.

Leonard (2001) found that students in low- and middle-achievement groups performed better on state mathematics assessments when grouped heterogeneously. High-achieving students had no significant differences in gains between heterogeneous and homogenous grouping. Leonard analyzed post-test scores by gender and ethnicity and found that African American girls were most affected by grouping type. They scored lowest in the homogenous grouping condition. However, the author also noted that at least one of the African American girls was very shy and did not interact with her group much, and almost all of the African American girls were more withdrawn from the group than other members, which may have influenced their retention of materials and understanding of concepts

Although this study seems to support previous research that heterogeneous grouping raises student achievement more than homogenous grouping (Oakes, 1985) these data must be viewed with caution. Leonard (2001) did not report the use of differentiated activities for different ability groups, which may account for some of the variation in scores. In addition, two different groups of students were compared over a period of a year each. Also, results reported in this study are only for one geometry unit although the study ran for the full academic year. These data lend further credence to Delcourt et al.'s (1994) contention that grouping without curricular modifications may not have the results intended by educators.

McSherry and Ollerton (2002) conducted a survey of 350 schools in the United Kingdom. Their intent was to learn whether schools were ability grouping for mathematics, and if so, the rationale for this grouping. This study was done in response to a national initiative to de-track schools. One hundred and ninety-two schools responded for a response rate of 55%. Ninety percent of students in this study were grouped by ability for all or part of the day. This ability grouping was either within-class grouping (55%) or across-grade grouping (20%) for all mathematics lessons with an additional 15% of schools grouping intermittently for mathematics instruction. Teachers explained their rationale for choosing to group by achievement or “measured ‘ability’” (p. 6) in terms of child-focused, teacher-focused, or administrative-focused reasons. Most teachers’ child-focused reasons centered on their belief that grouping facilitates differentiation. Teacher-focused reasons dealt with efficiency and reduction of ability ranges. Administrative-focused reasons concentrated on the national mandate to de-track schools, but these reasons were the least prevalent.

This study addressed grouping and not the practices associated with grouping. However, the authors’ discussion focused almost solely on the detrimental effects of grouping students in low-achievement groups and the resultant negative self-perceptions they might develop because of placement in a low-achievement group. The authors did not report data concerning differentiation or pedagogy in support of their conclusions.

Grouping and curricular adjustments. Tieso (2005) conducted one of the most comprehensive grouping and curricular differentiation studies to date. She conducted this study based on Kulik and Kulik (1992), Slavin (1987), and Davenport et al.’s (1994) work. These authors noted that grouping alone does not yield the same achievement gains as grouping with curricular adjustment. Tieso grouped 31 classes of fourth and fifth grade students in New England into three different conditions: comparison; treatment 1 – curriculum revision in a whole group setting; and treatment 2 – curriculum differentiation. She further divided the treatment 2 group into within-class groups or between-class groups. In addition, all groups were classified into low-, middle-, and high-achievement subgroups based on prior knowledge assessed through a curriculum-based pre-test. Students participated in a three-week mathematics statistics unit. The comparison group received instruction directly from the school’s adopted textbook.

Students in the curriculum revision group participated in lessons from a researcher-designed unit that removed unchallenging and repetitious material from the curriculum and enhanced learning activities with advance organizers, higher-order questioning strategies, and critical thinking activities. Students in the differentiated treatment group learned material through a series of tiered lessons based on their prior knowledge with additional activities such as interest centers and anchor activities.

Students in the revised curriculum condition made greater gains in all achievement subgroups than their peers in the comparison group. Effect sizes were +.18, +.25, and +.81 standard deviations (SD) for the low-, medium-, and high-achievement subgroups, respectively. Effect sizes for the differentiated instruction condition were also in the same range. The within-class-low subgroup yielded an effect size of +.28 SD, with an effect size of -.13 SD for the between-class-low subgroup. Within-class-middle and between-class-middle groups yielded effects sizes of +.42 SD and +.10 SD respectively. Finally, within-class-high yielded an effect size of +.83 SD and +.30 SD for between-class-high.

Tieso (2005) noted that student achievement increased without grouping students into small groups if curriculum was appropriately adjusted to student needs. These results support the findings of Lou et al. (1996). Overall, within-class instruction was superior to between-class grouping. This may be due to time lost in changing classrooms or other extraneous variables. Tieso designed this study to gather data in as many different grouping conditions as possible to provide comprehensive comparisons between groupings in a controlled manner. Therefore, these results pertaining to the superiority of within-class learning (grouped or otherwise), along with differentiation, are robust and should be viewed as important.

Notably, Tieso (2005) showed that ability-grouped students who received differentiated instruction achieved significantly higher in mathematics than those taught from the regular textbook. This has implications for educators who choose curriculum materials. No matter what textbook they choose, additional materials and curricular accommodations are necessary to ensure that all students receive adequate instruction in a curriculum area. Teachers taught this unit to fourth and fifth graders using middle school-level materials. Therefore, all students learned using challenging materials and yet, the textbook material still proved inferior to modified curricular materials and pedagogy.

Cooperative or collaborative learning versus individual learning. Hurley, Boykin, and Allen (2005) investigated communal/cooperative versus individual learning of a math-estimation task. This study was completed with 78 African American fifth-grade students in two urban, public schools. The duration of this study was very short, one 20-minute instructional session, but the rigor of the study was high. Group assignment by gender was random. Students were assigned to the two conditions: communal/cooperative or individual. Students participated in a twenty-minute study session either working in a triad with one set of instructions and materials or individually at different desks with individual instructions and materials. A pre- and post-test design was followed using the rotated splithalves of a thirty-item, researcher-designed test. Students were assessed immediately before and immediately after the study session. Pretest scores were similar across both conditions and across gender. A 2 X 2 analysis of covariance was conducted with condition and gender as factors and post-test as the dependent variable. The main effect of the experimental condition based on post-test performance was significant. Students in the communal/cooperative condition outperformed those in the individual condition by a magnitude of 17%. No gender effects were found.

Peklaj and Vodopivec (1999) investigated cooperative versus individual work in their study of 373 fifth-grade students. Students were in 15 classes in nine elementary schools in Slovenia. One hundred seventy students were assigned to the cooperative learning condition, while 203 students were assigned to the comparison condition – that is they were taught mathematics in the traditional format used in their classes. Students in the cooperative and comparison conditions were evaluated based on one lesson per week over seven months. Basic mathematical concepts covered during this time were problem solving, various calculation methods, and transforming measurement units. Results from the teacher-developed test at the conclusion of the study were used to determine that students in the cooperative learning condition did not achieve at a higher level than their comparison peers. Curriculum materials and pedagogy were not changed in this study. Again, this highlights the fact that grouping alone does not change the academic outcomes for students. Grouping must be done in conjunction with curricula and pedagogical adjustments to ensure student achievement rises.

A two-year, longitudinal study on the cooperative elementary school model was completed by Stevens and Slavin (1995). The cooperative elementary school model is a practical and philosophical approach to changing not only what happens in the classroom, but the school structure as well (Stevens & Slavin). The model is comprised of six factors: (1) comprehensive use of cooperative learning in academic areas; (2)

mainstreaming learning-disabled students; (3) peer coaching; (4) teacher collaboration in lesson planning; (5) collaboration between teachers and administrators for school planning; and (6) encouragement of parent involvement by both teachers and administrators.

One thousand twelve students in grades two through six in five schools comprised the sample. Twenty-one classes in two schools were designated as treatment classes, while the other 24 classes in three schools comprised the matched, comparison classes. Matching was done using the California Achievement Test (CAT) in the areas of reading, language, and mathematics. In addition, schools were chosen from neighborhoods of similar socioeconomic status and ethnic composition. During the first year of the study only 60 percent of students with a learning disability were mainstreamed. However, by year two pull-out, special-education programs were discontinued and students received remedial services within the regular classroom. In addition, students with learning disabilities were incorporated into heterogeneous learning teams along with their general education peers (Stevens & Slavin, 1995). Although data for students identified as gifted were reported, effects for this group were not initially hypothesized. These data were analyzed post hoc due to an increase in interest in gifted students' academic outcomes expressed in the literature at that time (Stevens & Slavin).

Pre-test results showed no significant difference in language and reading, but a significant difference was found in mathematics scores in favor of the comparison schools for students of all achievement levels. After the first year, students with learning disabilities (LD) and those identified as gifted (GT) showed no significant difference in achievement from the comparison group. However, after the second year of the study, students from both these achievement groups showed statistically significant achievement gains in all academic areas measured.

Stevens and Slavin (1995) provided academic achievement and social relations data for students who experienced a learning disability. Pre-test results for this sample were not significantly different from the comparison group for both achievement and social relations. After year one of the study, achievement scores for students with a learning disability were not significantly different from their comparison peers. However, after year two of treatment, students with a learning disability outperformed their comparison peers on measures of reading comprehension, language expression, math computation, and math application. Effect sizes were larger than those for students without a learning disability and ranged from +.35 to +.85. Social relations scores also increased significantly over the two years with an overall effect of +.86.

Students who were designated as gifted were also followed in this study. Pretest data in language arts showed no significant differences between students labeled as gifted in the treatment and comparison schools. However, students labeled as gifted in the comparison school had significantly higher math pre-test sores than their treatment group peers. Analysis of covariance was used to control for pre-test differences. Year one results showed no significant differences between treatment and control groups. However, post-test results at the end of year two showed significant gains for treatment students labeled as gifted in the areas of reading vocabulary, reading comprehension, language expression, and math computation. Effects sizes ranged from +.48 to +.68. No significant differences were found in language mechanics or math applications.

Stevens and Slavin (1995) noted that when students work cooperatively and teachers change their pedagogy to support this grouping student achievement is improved. This systematic change occurs within classrooms and schools as a whole and makes a difference in the achievement of students. The finding, that grouping, pedagogy, administrative policies alone cannot have the same effects as a concerted effort by all stakeholders is also supported in the literature (Delcourt & Evans, 1994; Gentry & Owen, 1999).

Kenny, Archambault, and Hallmark (1995) conducted a study of 786 fourth-grade students to investigate the effects of cooperative group composition on elementary gifted and non-gifted students. The study investigated achievement, self-concept, school attitudes, and perceptions of peers. Students were placed in either heterogeneous or homogeneous groups of gifted and non-gifted students. The study was designed to determine whether cooperative group structures used in different content areas yielded similar results.

The results of this study suggest that cooperative grouping whether homogeneous or heterogeneous does not have an appreciable effect on student academic performance (Kenny, Archambault, & Hallmark, 1995). Further, the presence of a gifted student in the group does not improve the learning of other students. Gifted students are noted as being more helpful than their non-gifted peers, but this helpfulness does not translate into increased performance.

The researchers further noted that inclusion of a gifted student in a cooperative learning group increased the negative perceptions that non-gifted students had of one another. The presence of a gifted student in the group increased non-gifted students' perceptions that their non-gifted peers are not as intelligent or helpful to the group, and are less able to lead than the gifted student. Therefore, although gifted students did not experience any adverse effects from cooperative grouping there do not appear to be any benefits to this type of grouping if the desire is to improve academic performance.

Robinson (1991) conducted a meta-analysis of cooperative grouping studies and the relevance of results to the field of gifted education. She found that most studies were not relevant to gifted education even if this population was included in the sample. The reason she gave was that although gifted students might be included in the sample they were rarely the group of interest and comparison groups were not well matched to treatment groups so causal comparisons were not possible.

Based on her findings, Robinson (1991) made five recommendations regarding cooperative grouping. These recommendations were based both on extant literature regarding educational practice for gifted students as well as the literature on cooperative grouping. Robinson recommended cooperative grouping should not replace specialized programming for gifted students. Further, if cooperative grouping is used, it should include advanced materials for academically advanced students to use. In addition to advanced materials, flexible pacing should also be considered when grouping academically gifted students. Although cooperative groupings are by nature heterogeneous, Robinson also contended that the disparity in achievement levels should be minimized and these groups should only be used part of the day. Gifted students should still be given time to work in homogeneous groups or individually at their own level.

Summary. Studies such as the ones discussed above and the earlier meta-analyses have been done in order to inform instruction in the mathematics content area. The question of which type of grouping best promotes student achievement is far from answered, as results of studies are varied and often contradictory in nature. Researchers who studied middle and high school students reported the difficulties associated with tracking students (Gamoran & Berends, 1987; Kerckhoff, 1986). These same researchers argued that students who exhibit gifted behaviors or high achievement perform the same in either a homogenous or heterogeneous environment, seemingly justifying the use of heterogeneous grouping. What remains to be investigated in these studies and much of the research is how much better these high-achieving students might do in homogenous mathematics groupings. Tieso (2005) noted that high-achieving students made large gains when placed in within-class ability groups with curricular modifications over three weeks. This is significant in that students of the same achievement levels made gains, but at a lower level in revised curriculum and across-grade conditions with effect sizes of +.81 and +.30 standard deviations, respectively.

Hurley, Boykin, & Allen (2005), Peklaj & Vodopivec (1999), and Stevens & Slavin (1995) noted that working together on a task did not detract from students' academic achievement. Further, Leonard (2001) contended that students who achieve at low levels fair better in heterogeneous groups, which is supported by Stevens and Slavin's (1995) cooperative grouping findings. What Leonard's and Stevens and Slavin's studies lack is a comparison between heterogeneous or cooperative grouping and homogenous grouping. Therefore, it is impossible to say that their findings offer evidence of maximum gains in achievement. Further research is needed to answer this question.

Grouping for reading instruction. Research studies investigating the effect of grouping used for reading instruction are often naturalistic in nature and describe, rather than evaluate, reading instructional practices in the classroom (Barr & Dreeben, 1991; Davies, Hallam, & Ireson, 2003; Grant & Rothernberg, 1981; Hallam, Ireson, & Davies, 2004; Wilkinson & Fung, 2002). These studies are often observational or ethnographic and describe, rather than impose, a treatment. However, data from naturalistic studies add to the evidence presented in the meta-analyses described above.

Grant and Rothenberg (1981) conducted a qualitative investigation of the social interactions of students and teachers in reading ability groups in first and second grade classrooms. They reported the quantity and quality of interactions in these groups by ability. Teachers seem to allocate less instructional time to the low-ability groups than high-ability groups when grouping is homogenous (Barr, 1992; Collins, 1986; Grant & Rothenberg, 1981). This finding does not indicate that less time is spent with low-ability groups, but that more time is used correcting students' reading rather than instructing them in reading strategies (Lou et al., 1996; Barr, 1992; Collins, 1986; Grant & Rothenberg, 1981). Teachers interrupt poor readers more than they do high-ability readers. Teachers also give more corrective feedback to low-level readers rather than prompting; whereas, when they interrupt high-ability readers, teachers tend to prompt rather than correct errors. Teachers often direct lower-level questions to low-level readers rather than using lower-level texts and keeping the questioning level high (Barr, 1992; Collins, 1986; Grant & Rothenberg, 1981).

These results may only seem to be pertinent to reading instruction, but they are important to note as TSCG is designed to group students and to train teachers to use the same pedagogy regardless of the ability levels of their students. Furthermore, the efficacy of the model and has been studied in both reading and math content areas (Gentry & Owen, 1999). Differentiation strategies were used in all cluster classrooms regardless of content area and students' achievement levels. Gentry & Mann (2008) indicated that if the TSCG model is correctly implemented it should be indiscernible what achievement levels of students the teacher is teaching when pedagogy is observed.

Further analysis of interactions in groups according to ability showed that interactions between teachers and students are more than just pedagogical in nature (Eder, 1981; Eder & Felmlee, 1984; McDermott, 1976). For instance, teachers and students responded to one another and built group norms based on the expectations of both the teacher and students. Therefore, it could be said that merely shifting the pedagogy to a more rigorous level might not be enough to change the performance of low-ability students. A cultural shift in student-perceived self-efficacy as well as pedagogical change may need to occur in order to raise student achievement.

Researchers observing teacher practices have found that grouping is not as prevalent as reported by teachers in the field (Vaughn, Moody, & Schumm, & Klingner, 1998). This may be due to the fact that grouping requires prior planning and differentiation. This supports Lou et al.'s (1996) and Kulik's (1992) contention that grouping is not useful unless teachers make modifications to the curriculum and materials used by each group. The TSCG model intentionally places students into achievement groups and differentiates content based on the needs of students in each group – a practice supported not only in the general education literature, but also as shown here, in the special education literature, in the area of reading (Barr, 1992; Collins, 1986; Grant & Rothenberg, 1981).

Chorzempa and Graham (2006) investigated the types of grouping used and the instructional practices of elementary school reading teachers. Seventy-eight percent of teachers in this study used within-class ability grouping to meet the instructional needs of students. This was done either due to teacher beliefs in the usefulness of this type of grouping or because the reading curriculum lent itself to this type of grouping. The researchers found that lower-ability readers spent less time reading silently, spent more time reading orally with the teacher, answering literal questions about the text, and completing worksheets than their higher-ability peers. These findings are consistent with Grant and Rothenberg (1981).

Poole (2008) investigated the practice of heterogeneous grouping in reading. She observed the teacher-student interactions in groups to assess the usefulness of this type of grouping on reading skills acquisition. Poole's findings mirror those of Grant and Rothenberg's (1981) findings that teachers allocate less instructional time to lower-ability readers and interrupt their reading more than their higher-ability peers. What is pertinent to these findings is that groups were heterogeneous while Grant and Rothenberg's research was done with homogeneous groups. Results from both studies were identical, which indicates that it is not the grouping that affects students' reading proficiency, but the nature of teacher interactions and pedagogy. Lleras and Rangel (2009) completed a similar study of African-American and Hispanic students and found similar results. When ethnicity was controlled for, primary students in homogeneous reading groups learned less than their peers in classrooms where no grouping was used.

Schoolwide enrichment model - reading. The Schoolwide Enrichment Model- Reading (SEM-R) has been studied for a number of years. Researchers have shown its effectiveness in motivating students of all reading levels and increases reading skills acquisition. Students are exposed to high-level literature and challenged through engaging, hand-on activities. Those who choose to go further and complete level two and level three enrichment activities may do so, while other students may move on to new texts or learning experiences. In this way students are grouped, but some student choice is offered in the grouping as they advance through the levels of activities, while all students are engaged in high quality learning.

Reis et al. (2007) conducted a randomized study to investigate the usefulness of SEM-R on urban elementary school students' reading comprehension, fluency, and engagement in reading. Treatment students spent twelve weeks participating in the SEM-R program that exposes them to literature-based activities that are engaging, challenging, and high-interest. Control students remained in the extant, remedial reading program that provided reading instruction and test preparation. At the end of the study, treatment students outperformed the control students in tests of reading fluency. In addition, attitudes towards reading were more positive for treatment students. Variability in reading fluency and attitude can be explained by the content and pedagogy used with students. This indicates the importance of challenging all students with appropriately leveled materials and requiring them to think at high levels regardless of reading ability.

Cluster Grouping

Cluster grouping students in the regular classroom has received attention in the literature for more than 20 years. Clustering places students who are identified as high-achieving or high-ability in one classroom in a grade level together with students of other ability levels (Gentry & Owen, 1999; Gentry & Mann, 2008; Hoover, Sayler, & Feldhusen, 1993). This is an attractive model as it is low-cost, meets the needs of gifted students on a full-time basis within the regular classroom (Hoover, Sayler, & Feldhusen, 1993; Purcell, 1994), and allows gifted students to interact with their intellectual peers on a regular basis (Delcourt & Evans, 1994; Rogers, 1991; Slavin, 1987). Clustering gifted students into homogenous groups with teachers who are interested in and qualified to teach them facilitates differentiation and results in more widespread use of differentiation (Bryant, 1987; Gentry & Owen, 1999; Kennedy, 1995; Kulik, 1992; Rogers, 1991). This removal of gifted students from the other classrooms allows other students to emerge as high achievers, thus potentially increasing the number of students identified for gifted education programming (Gentry & Owen, 1999; Kennedy, 1989). Many of the benefits of ability grouping outlined by Kulik (1992) and Kulik and Kulik (1992) also pertain to cluster grouping. These benefits include time for high achievers to work with their academic and intellectual peers; grouping students by achievement levels allows teachers to target learning activities at particular groups of students; and provides academically advanced students to be grouped with teachers who are trained to meet their educational needs.

Cluster Grouping Research

Clustering high-achieving students together in one classroom has potential benefits, but also raises some concerns, which are the same as those outlined for ability grouping in general. Removing the high-achieving students from the classroom concerns educators. They question the effects on self-efficacy and self-esteem of the other learners (Hoover, et al., 1993; Lou et al., 1996; Slavin, 1987). Another issue is how well cluster grouping provides differentiation for high-achieving students (Delcourt & Evans, 1994; Rogers, 1991; Gentry & Owen, 1999).

Gentry & Owen (1999) were the first to show the usefulness of cluster grouping quasi-experimentally using longitudinal achievement data, although several studies have also been done that include analysis of student achievement for high-achieving students who are clustered (Bear, 1998; Brulles, 2005; Delcourt, Lloyd, Cornell, & Evans, 1994). Benefits include the opportunity for high-achieving students to interact with their intellectual peers on a full-time basis (Gentry & Owen; Hoover, Sayler, & Feldhusen, 1993). Research has shown student achievement increases when students learn in cluster-grouped classes (Brulles, 2005; Gentry & Owen); and over time, more students are identified as high-achievers and fewer as low-achievers (Gentry, 1999; Gentry & Owen).

Hoover, Sayler, & Feldhusen (1993) conducted a survey of Indiana educators to ascertain what forms of gifted education programs were available in the state. They found that only 28% of Indiana school districts were using cluster grouping in their schools, and that this form of gifted education was relatively new, one to three years since inception. Survey findings from 96 cluster teachers indicated that although parent support was high (65%), cluster teachers reported that regular classroom teachers either felt neutral (49%) or negative (10%) about cluster grouping as a program. Most of the negative perceptions were based on teacher perceptions that all the good students had been removed from their classrooms. Hoover et al. also indicated that teachers felt that cluster grouping benefited not only gifted students, but also students of other ability levels. However, they noted that empirical evidence on student achievement growth due to the model was lacking.

Gentry & Owen (1999) interviewed teachers from the cluster-grouped school and also found positive perceptions of the program, especially when additional between-class grouping was used to meet the needs of a wider group of students, in particular content areas. Teachers reported that, although some students were not placed in the high-achieving cluster group, they needed educational services in a particular content area such as math. This flexible grouping in addition to clustering afforded teachers, of clusters of students achieving at low levels, the time to work at a much more appropriate pace to enable students to master content. This study was the first to provide empirical evidence of the usefulness of cluster grouping in meeting the needs of both gifted students and general population students.

Delcourt, Lloyd, Cornell, & Goldberg (1994) conducted a two-year investigation of grouping in 14 different school districts. Their objective was to investigate any differences in achievement, self-concept, and motivation based on grouping type and ethnicity. Delcourt et al. studied students who were identified as gifted in four different programming options: special schools, separate classes, pullout classes, and within-class grouping. These students were compared with high-achieving students and average-achieving students in school districts with no gifted education programming options. The authors found that cluster-grouped students scored lowest when compared to students in all other programming arrangements, which led them to conclude that cluster grouping must be done in an intentional manner with specific training and instructional differentiation for meeting the needs of gifted students. Without this attention to pedagogy and materials, cluster grouping can collapse into no model at all. Delcourt & Evans' (1994) follow-up study offered a cluster grouping program as an example of an exemplary gifted education program. The factors that set this example apart were strong leadership, well-articulated identification procedures, and close attention to the educational needs of gifted students. These aspects are part of the TSCG model and are critical to its success.

Delcourt et al.'s (1994) and Delcourt and Evans' (1994) findings that grouping itself is not enough to meet the needs of high-achieving students mirror the findings of Kulik (1992) and Lou et al. (1996) regarding ability grouping. Strong administrative support and changes in pedagogy are necessary in order to meet the needs of students who are grouped by ability. Homogenous grouping not only benefits those who are high achieving, but also benefits students of all ability levels when properly implemented. Kulik suggested that flexible grouping be a hallmark of any grouping program. Gentry & Owen's (1999) study provided further empirical support for this interaction between grouping, administrative support, and changes in pedagogy.

Research has been done to evaluate the effects of cluster grouping on academic achievement of elementary school students (Bear, 1998; Brulles, 2005). Bear investigated achievement for students who were placed in classrooms with a cluster of gifted students and a teacher who was trained to teach in a differentiated manner. However, the authors did not study the achievement of gifted students. Rather, they followed non-gifted students over a three-year period. Pre-test data were used to stratify students into high-achieving, above-average achieving, and low-achieving groups. Students in the control group were those who were never placed in a classroom with a cluster of gifted students.

They were also stratified according to initial test results. There were 60 students in each of the treatment and comparison groups for a total of 120 students in the sample from the entire school district. Overall, only students in the highest quartile were found to show significant increase in achievement over the three-year period. In addition, the main effect of clustering was not significant for either mathematics or reading; however, initial achievement was a significant factor in the variance of scores.

No discussion of identification practices was offered in Bear's (1998) study, so it is difficult to assess the level of performance required to place a student into the high-achieving cluster. Additionally, it is difficult to attribute differences to cluster grouping as teachers were neither all teaching in the same manner nor using the same sorts of high quality materials as suggested in the TSCG model. Essentially, the manner in which cluster grouping was implemented in this school district was not in the manner stipulated by the TSCG model.

Brulles (2005) also incorporated longitudinal data from schools in one school district. She investigated the difference in student achievement in a district using cluster grouping in comparison to the state statistics for student achievement and identification of students for gifted education programming. Students, identified as gifted, in cluster-grouped classrooms, were compared to students of all other levels of ability within those same classrooms. During the course of the study student identification for gifted education programming increased from 2.67% to 6% of the school district population. Over- and under-identification of students from culturally-diverse families still existed within the program with Caucasian and Asian students over-represented, and African American and Hispanic students under-represented.

Mathematics achievement growth over the course of one year was measured and analyzed for both gifted and non-gifted students in cluster-grouped classrooms. Overall, students in the cluster grouping model, regardless of ability level, showed an increase in mathematics achievement between the first and third quarters. Non-gifted females made greater progress in math than their gifted counterparts. Gifted and non-gifted males increased their mathematics achievement significantly when taught in a cluster-grouped classroom. ANCOVA performed for non-gifted, male students indicated significance in the cluster group placement. ANCOVA was not conducted for gifted males due to small sample size. Caucasian students of all ability levels showed significant increase in achievement over time. However, gifted, Hispanic students did not make significant gains in achievement in the cluster-grouped classroom, while their non-gifted counterparts did. Gifted students who were English Language Learners (ELL) did not show significant gains in mathematics achievement, however, non-gifted ELL students did show significant gains.

Brulles (2005) noted that, although frequency of identified high-achieving students increased using cluster grouping, not all subgroups increased their achievement once placed in cluster-grouped classrooms. She noted that this could be due to lack of teacher training and understanding of how to meet the needs of high-achieving learners in the classroom. This study was done over three grade levels, but during less than one year of schooling. This could also account for the inability to show growth for all students. The amount of time ELL students had been immersed in the American school system was not explained and may account for differences in math achievement.

Research on the total school cluster-grouping model. The TSCG model is a specific and formalized version of cluster grouping. The model specifies how students should be identified and how all students should be assessed and placed in groups appropriate to their level of achievement (Gentry & Owen, 1999; Gentry & Mann, 2008). Identification is done by teacher nomination with the use of test scores as a means of inclusion rather than a means of exclusion. Thus, a child would not be excluded from the high-achieving cluster based on low-test scores if all other evidence suggests that he/she should be in that cluster. Pedagogy that in the past has been reserved for gifted education must be used in all classrooms in order to expose learners of all achievement levels to high quality, challenging material. When students are offered curriculum at their level of achievement that requires them to think at high levels and justify their answers then the achievement of all students increases (Gentry & Owen; Gentry & Mann).

Gentry and Owen (1999) provided the first published empirical evidence of the efficacy of the TSCG model. This study was a causal-comparative, longitudinal study in which two schools were compared – one that implemented cluster grouping and one that did not. Quantitative and qualitative evidence was sought to explain the achievement growth and to understand teacher practices within the model and its usefulness in the elementary school setting. Reading scores for students in the treatment school were lower than in the comparison school at the beginning of the study. However, over the course of three years of cluster grouping scores equaled or rose above those of the comparison school. Mathematics scores for students from the treatment school were also higher than for the comparison students at the end of the study period. However, the change in scores was not significantly different. This may be because scores from the students in the treatment school were high to begin with. Students in this study were not only clustered into regular classrooms, but flexible grouping between classes, within grade level, was also used in order to further meet the needs of all students in specific content areas. Teacher preparation and strong administrative support provided a positive environment for both teachers and students to excel. Overall, teacher perceptions were positive with teachers enjoying the opportunity to be more deliberate in how they taught students and the time to meet the needs of a smaller achievement-range of students.

A major difference between Gentry and Owen's (1999) and Brulles' (2005) study is that in Brulles' study only high-achieving cluster teachers received training in differentiation and meeting the needs of high-achieving students. Additionally, teachers of students other than high-achieving students were not mandated to change their pedagogical practices. The results of this study support Delcourt and Evans' (1994) contention that grouping without changes in pedagogy results in no program at all.

Investigation of Identification of Diverse Students

The proposed study will investigate if the TSCG model increases representation and achievement of economically, culturally, and linguistically diverse students in the high achieving cluster. Although Brulles (2005) attempted to answer this question, her results were not conclusive due to the short time-period of her study. This study will use data from a three-year period allowing more time to show student achievement growth. In addition, a longer study time enables all students and teachers to become proficient with the model.

Under-representation of students in gifted education programs. Historically, providing an equitable education for students regardless of ethnicity and race has provided a major challenge to American educators. In the past, groups that have been regarded as minority groups have been just that. However, NCES statistics (2005) indicated that as of the 2000 census, these so-called minority groups comprise 42% of public, pre-K – 12 students enrolled in the United States. Gifted education programs should mirror this distribution. However, this has not been the case in past years, and continues to challenge schools today (Yoon & Gentry, 2009). Donovan and Cross (2002) analyzed OCR data and noted that in 1998 6.2 percent of all schoolchildren were placed in gifted and talented programs. This seems to be a typical percentage for identification of gifted and talented students, the disaggregated data tell another story. Cross and Donovan computed the odds ratios (the risk index of one racial group by the risk index of another) for each group. All odds ratios were computed against the risk ratio (a risk ratio is calculated by dividing the number of students in a given racial or ethnic group served in a given category by the total enrollment for that racial or ethnic group in the school population) for white students. When the 6.2 percent of identified students was disaggregated Donovan and Cross found that Asian/Pacific Islanders were one-third more likely to be placed in gifted programs than White students. American Indian/Native Alaskan students were next likely to be represented. Hispanic and Black students were least likely to be represented (Donovan & Cross).

The percentage of students identified as gifted has increased steadily since 1976, when only one percent of schoolchildren were identified as gifted and talented. Rates for all ethnic groups have increased during this time with American Indian/Native Americans increasing the most with their odds ratio increasing from 0.40 to 0.65. Asian/Pacific Islanders increased the least, with their odds ratio decreasing from 2.15 to 1.34. Rates of identification for Black and Hispanic students have not risen to rates proportional to their population distribution as documented by Ford (1998) and Plata, Masters, and Trusty (1999) using the same data from the OCR. The odds ratio for Hispanics has dropped from 0.97 to 0.60. This is a large decrease given the increase in the population of this particular ethnic group (Donovan & Cross, 2002).

Yoon and Gentry (2009) analyzed representation data from three different federal sources: the National Education Longitudinal Study of 1988 (NELS:88), the School and Staff Survey (SASS), and the Office of Civil Rights (OCR) data collections. Their findings confirm the over-representation of Asian/Pacific Islanders in gifted programs and the under-representation of Black, Hispanic, and Native American students.

Yoon and Gentry (2009) conducted an analysis of the Office for Civil Rights (OCR) data collection for years 2002, 2004, and 2006. As of 2006, Asian and Pacific Islanders were over-represented in gifted education programs in 41 states. Caucasian students were also over-represented, to a lesser degree, in 26 states. The authors also cautioned that using national trend data is not the ideal means of assessing representation and that these data need to be disaggregated by ethnic subgroups in order to understand the data. Generally, they also found that Alaska Native, American Indian (27 states), African American (42 states), and Hispanic students (43 states) are underrepresented. Although these data are informative, certain limitations inherently exist in a national dataset. Data for some states were unusable due to statistical uncertainties. In addition, identification of giftedness differs from state to state and therefore aggregated data such as these do not give a precise picture. However, these data do provide valuable information on national and state trends in identification of students for gifted education programs.

Smith, LaRose, and Clasen (1991) conducted an experimental study to assess under-representation in gifted programs and provide gifted education services to students who qualified and were selected for the treatment program. Today's ethical standards for human subjects research would prevent such a study; however, the results are pertinent to this discussion. The researchers selected the top 9% of each major ethnic group: White, Black, Hispanic, and Other and designated them as gifted. This resulted in a pool of 91 students who were selected from a class in Milwaukee, Wisconsin. Researchers randomly assigned these students to either a gifted education treatment or a control (general education program) treatment. From the 91 participants, 24 were chosen for the gifted program and comprised the treatment group, while 67 were placed in the regular classroom and comprised the control group. Twelve years after placement, researchers followed up with students to assess their achievement. Of the original 91 participants, 78 were still in the Milwaukee area. Of those in the treatment group (n = 24) none had dropped out of the program, while 45% of those placed in the control group had. Thirty-three percent of the treatment group participants graduated from high school, but did not attend college. Sixty three percent graduated from both high school and college, while those in the control group had not faired as well. Only 18% of the control group students graduated from high school, but did not attend college with 21% graduating from both high school and college. It must be noted that although this is a small sample, the results are undeniable. It is clear that the gifted and talented education program for the minority students in this study played an important role in their educational success.

Identification of low-income students. Low-income students of all ethnic and racial groups remain underrepresented in gifted education programs (Donovan & Cross, 2002; Ford, Grantham, & Whiting, 2008; Naglieri & Ford, 2003). Students from low-income families often do not have access to the experiences and resources of their peers from not-low-income families. Access to such resources and services predicts academic success (Valencia & Suzuki, 2001). Stambaugh (2007) summarized the findings of the *National Leadership Conference on Low-Income Promising Learners* and highlighted practices that would promote identification of students from low-income families for gifted education programming. Stambaugh suggested that students should be identified early, as low-income students tend to fall behind academically as time goes on if they are not served appropriately. Therefore, early identification is essential to "catching" these students before this occurs. This population of students should also be assessed for identification often, as low-income students may need time in the educational system in order to catch up before being able to demonstrate their abilities. Nomination scales were suggested as important tools in order for teachers to assess non-traditional markers of ability as well as affording them the opportunity to compare students of like circumstances and life experiences to one another. This is different from traditional testing where students are compared to all their age peers across either the region or country. A multiple informant approach using nomination forms and other measures is also important in this context.

Professional development in recognizing potential in low-income students is also imperative, as teachers tend to come from the majority culture. Wyner, Bridgeland, & Diiulio (n.d.) wrote the research report, *AchievemenTrap: How America is failing* *millions of high-achieving students from low-income families*. They found that teachers cannot expect students from low-income families to score in the same range as their peers from not-low-income homes because their home contexts do not provide the richness of resources. This reduces background knowledge necessary to perform well on achievement tests. The researchers suggest that a child from a low-income family scoring at the 75th percentile on a standardized achievement measure is equivalent in ability-level to a child from a not-low-income family who scores at the 90th percentile. Students from low-income families often do not have access to the same resources and information as their peers from not-low-income families. This may result in a lack of background knowledge and contribute to the lower scores on achievement tests. Therefore, lower scores are not always an indicator of lower ability, but could indicate a lack of background knowledge needed to answer test questions.

Robinson, Shore, and Enerson (2007) investigated best practices in identifying low-income students for gifted programs and for retaining them in programs. Their suggestions centered on identification that includes intensive observations, performance assessments, and allowing students to try out the program. Second, parents of students from low-income families must be involved in their child's program in order to retain these students in programs comprised typically of middle-class, Caucasian students. They noted that parents need to trust the program and the educational system in order to help their children understand the importance of the program. This is best done by actively involving parents within the classroom and at home. In addition, multiple-year involvement with students encourages retention in the program. Finally, connecting parents with the school is important as it facilitates communication and buy-in for students. It also encourages learning to continue in the home, which benefits both the student and the family.

Summary. The majority of researchers who have studied ability grouping suggest that homogenous grouping by ability is beneficial for all students regardless of ability level. It is particularly useful when students are clustered in an intentional manner, and teachers are trained to differentiate for all students as in the TSCG model. Caution must be taken to ensure that grouping is done intentionally and flexibly in order to allow students to move between groups as needed.

Students of diverse ethnic, linguistic, and socio-economic backgrounds must be adequately represented in gifted education programs. This requires new practices in identifying, retaining, and teaching students to meet their challenges while attending to their need for rigorous curricula. Historically, this has been a problem, but it has improved and must continue to be addressed. It remains to be seen if the TSCG model can alleviate these issues in a diverse population of learners.

CHAPTER THREE: METHODS

Introduction

This chapter describes the methods and procedures used in this study of the Total School Cluster Grouping (TSCG) model. This includes a description of the population from which the sample is drawn; the research sample, which includes a description of the treatment school and comparison school; instrumentation, which includes a description of the measures used as well as the interview protocol; and an overview of the treatment, which includes a description of how the treatment school implemented TSCG. Data collection and analysis procedures are described for both quantitative and qualitative data.

The purpose of this study was threefold, to investigate: (1) the effects of the TSCG model on student academic achievement in a diverse, urban school district over a three-year period, using a multi-level model for change; (2) the effects of the TSCG model on identification of students in all achievement categories, by socio-economic status and ethnicity; and (3) practices and perceptions of teachers in the school implementing the TSCG model.

The selection of a school district with a diverse student population afforded the opportunity to investigate the usefulness of the TSCG model in addressing educational issues experienced by students from low-income and/or culturally diverse families. This included: (1) under-identification of students from low-income and/or culturally diverse families in gifted education programs; (2) under-representation of these same populations in gifted education programs; (3) achievement gaps between underrepresented and represented groups of students as well as between students from low-income and/or culturally diverse families and students from not-low-income families. This provided an opportunity to extend the findings of Gentry and Owen's (1999) study, which was conducted in a rural setting.

For the purpose of this research, low SES was defined as students who qualify for free or reduced school lunch under the Richard B. Russell National School Lunch Act (2004). Not-low SES was defined as students who did not apply or did not qualify for free or reduced school lunch. Treatment school was defined as a school in which the TSCG model was implemented. Comparison school was defined as a school in the same school district that did not receive the treatment, but that was similar in size, demographics, achievement levels, and community characteristics.

Research Questions

The following research questions will guide this study:

1. What effects does TSCG have on student achievement as measured by the NWEA, specifically:
    a. What differences exist between treatment and comparison student growth curves among four groups based on income and representation status: (1)

low SES/underrepresented; (2) low SES/represented; (3) not-low SES/underrepresented; (4) not-low SES/represented[2].

b. Do treatment school students in a TSCG program out-perform matched comparison school students after three years in the program after controlling for SES and ethnic representation?

c. How do the learning trajectories differ between treatment and comparison students?

d. For treatment school students only, based on initial identification categories, how do student achievement growth trajectories differ over three program years?

e. For treatment school students only, using four groups of students based on income and representation status, how do student achievement growth trajectories differ over three program years?

2. What effects does TSCG have on identification categories?

    a. What changes occur in frequency of students in each of the five achievement groups?

        i. by income?

        ii. by ethnicity?

3. What factors exist within the classrooms and the school using TSCG that may influence student achievement?

---

[2] To account for the correlation between ethnicity and SES, we collapse ethnic groups into underrepresented (African American, Hispanic, Native American) and represented groups (White, Asian, Other), using categories of (1) low SES/underrepresented; (2) low SES/represented; (3) not-low SES/underrepresented; (4) not-low SES/represented.

Research Design

This study used a mixed-method design. The quantitative part of the study used a longitudinal model with quasi-experimental, nonequivalent group design. The qualitative part of the study used a grounded theory, semi-structured interview design.

The nonequivalent groups design is a frequently-used design in social sciences research (Trochim, 2006). The structure is much like a pretest-posttest, randomized design with one major difference – it lacks the randomization feature. In nonequivalent groups design, intact groups closely resemble treatment and control groups (Trochim, 2006). However, since random assignment is absent, groups are not always equal or identical in nature. Group matching is as close as possible based on the characteristics of the groups (Trochim). Lack of randomization may reduce the ability to attribute student achievement differences to the treatment effect as teachers from the comparison school may use some of the same practices as teachers in the treatment school.

In this study, one purposively selected treatment school was matched with a comparison school that was best approximated by size, demographics, achievement levels, and community characteristics as defined by the Indiana Department of Education (IDOE) and school district administrators. Random assignment was not used in this study, as the intent was to investigate the usefulness of the TSCG model in the context of the school district, in the school that chose to implement the model.

Qualitative data were used to extend and explain the quantitative findings. Teachers from the treatment school were interviewed to obtain their perceptions of the model and its efficacy in meeting the needs of students. This inquiry focused not only on those teachers who taught the high-achieving students, but also included information from teachers who taught a broad range of students at all grade levels under investigation. These qualitative data were gathered using a semi-structured interview protocol to provide a clearer picture of the phenomenon under investigation (Patton, 2002; Trochim, 2006). The focus of the qualitative inquiry was to understand the processes occurring in the school and classrooms where the TSCG model was implemented.

Population and Sample

Population. The urban school district under investigation was situated in a Midwestern, metropolitan area. School district enrollment ranged from 14,517 students for the 2004-2005 school year to 15,384 in the 2008-2009 school year. Ethnic distribution in the district was 48% White, 31% Black, 14% Hispanic, 5% Multi-racial, and 1% Asian (IDOE, 2009). The school district was comprised of three high schools, 2 middle schools, 11 elementary schools, and 1 special education school. School district demographics are summarized in Table 2.

Table 2

School District Demographics

	School Year			
	05-06	06-07	07-08	08-09
Student Enrollment	14,628	14,713	14,917	14,517
Special Education	18.2%	17.9%	16.4%	15.0%
Gifted Education	6.0%	8.0%	7.0%	N/A
Free/Reduced Lunch	69.0%	69.0%	76.0%	76%
Limited English Proficiency	10.0%	10.2%	11.5%	11.5%
Intra-District Mobility	2.9%	2.6%	2.9%	N/A
Inter-District Mobility	16.1%	15.0%	14.8%	N/A

Source: Indiana Department of Education, 2009.

Student Sample. Students from two elementary schools in the school district participated in this study. One school served as the treatment group and one as the comparison group. The treatment school was selected due to the diversity of the student population and the district personnel's desire to learn about the effects of the TSCG model. Specifically, students who will graduate in years 2014, 2015, 2016, 2017, and 2018 comprised the sample. Table 3 depicts demographic data for the treatment and comparison schools.

Table 3

Overall School Demographics for Treatment and Comparison Groups 2008-2009

Factor	Treatment School	Comparison School
Total Students[1]	825	644
Black	26.0%	24.0%
White	49.0%	39.0%
Asian	3.0%	2.0%
Hispanic	13.0%	27.0%
Multi-racial	10.0%	8.0%
Free/Reduced Lunch	50.0%	65.0%
Attendance	97.7%	96.0%
Stability Index	77.6%	81.5%
Cost Per Pupil (Teacher Salary)	$2,223	$2,530

Source: IDOE, 2009.

Note. [1] Data are not available for all students due to intra- and inter-district mobility.

Teacher Sample. Teachers from the treatment school were interviewed to assess their perceptions of the TSCG model and to learn about practices used by teachers in cluster-grouped classrooms. Students were identified for cluster grouping beginning in second grade. Therefore, the interview sample included all second through sixth grade, high-achieving cluster teachers (*n* = 5) and one teacher per grade level (Grades 2 – 6) from the treatment school (*n* = 5). This provided a sample of 10 teacher interviewees.

Teachers (Grades 2 – 6) who did not teach the high-achieving cluster at the treatment school were randomly selected and asked to participate in the interview process. Random selection was used because teachers who volunteer may influence the data positively or negatively depending on the reasons why they volunteer to be interviewed (Creighton, 2005). Teachers not included in the randomly-drawn sample were excluded from this investigation.

A list of teachers was made from the school website and information from the school administrator. The list was numbered consecutively for each grade level and then numbers chosen from a random number table for each list of non-high achieving cluster, grade-level teachers. These names were submitted to the school administrator to request permission to contact teachers individually. After permission was granted I contacted each randomly-selected teacher as well as the high-achieving cluster teachers to obtain consent to interview him/her.

Treatment: Total School Cluster Grouping Model (TSCG)

The treatment used in this study, namely the implementation of the TSCG model is discussed in two phases. The first section addresses how students were identified for each of the cluster groups in the model. The second section describes how teachers were selected to teach particular clusters as well as the training cluster teachers received to implement the model.

Identification of student achievement level. Training was provided in spring 2005 to all cluster-grouping teachers and the administrator on the identification process. In addition, one of the co-authors of the model facilitated the identification process the first year of the study (Gentry & Mann, 2008). Identification of achievement levels for placement in cluster groups was done annually using teacher nominations and NWEA (2005) test scores. The first step involved teachers identifying the achievement level of their students for placement in classes for the following year. This identification was done for all students who were placed in one of the five categories stipulated in the TSCG model:

1. *High Achieving* students are great at both math and reading
2. *Above Average Achieving* students are good at math and reading or are great at either math or reading.
3. *Average Achieving* students achieve on grade level; they neither struggle, nor do they excel.
4. *Low Average Achieving* students struggle slightly with reading and math, or they struggle with either reading or math.
5. *Low Achieving* students find school difficult; they struggle in all academic areas and are at risk of failure (Gentry & Mann, 2008, p. 26)

Teachers based their judgments on their knowledge of students' performance and their experiences with them during the year. Teachers identified students with special learning needs (those with an individual education plan) based on their academic potential, rather than on their deficits.

Each year classroom teachers completed placement cards with information about each student. This included information about students' strength and challenges. It also included information about students' work habits, students with whom they worked well; and other information that facilitated placement decisions. These cards were used in the placement process and provided information to the teacher for the following year. Grade level teams then met and placed students into classrooms according to the cluster definitions given in the model.

Teachers are able to evaluate their students' performance on tests and assessments. However, they are not always able to identify those high-ability students who do not perform in the classroom. This group is often characterized by careless work or no productivity at all (McCoach & Siegle, 2000). They often complain about boredom and lack interest in school. These students may have poor study habits and give up easily (McCoach & Siegle). They may be inattentive and have behavioral issues. Gifted students who are not challenged in the classroom often amuse themselves to the distraction of other students and the teacher (Plucker & McIntire, 1996). Some of their behaviors can be disruptive in the classroom setting. Their performance is average, which is far below the level of which they are capable (Gentry, Gable, & Springer, 2000; Kanevsky & Keighley, 2003).

Teachers and administrators used NWEA test scores to identify students with high-ability who under-achieved. If a child tested well (e.g., at or above a local norm of 90-95th percentile in mathematics and reading [Gentry & Owen, 1999]) then, despite performance in the classroom, the child was included in the high-achieving cluster.

Although this identification procedure may seem subjective, identifying students using a multi-dimensional approach provides a much broader picture than simply using a test score. A test score provides a picture of a child at a particular point in time; whereas, teachers base their nominations on an entire year of observations by the teacher (Gentry & Mann, 2008). In other words, teacher nominations may well identify those who do not score well on tests or even in school, but who have high potential to perform if given the resources and opportunities that a rigorous curriculum provides (Wyner, Bridgeland, & Diiulio, n.d.)

Teacher assignment designation and training. All teachers and the administrator in the treatment school attended a two-day training session in which the TSCG model was explained and during which they engaged in implementation planning. All teachers in the treatment school were provided with a general overview of gifted education practices and talent development. This training was based upon the three-ring conception of giftedness (Renzulli, 1978) and differentiation practices typically used in gifted education. Additional professional development was also provided regarding the practical implementation of differentiation practices, identification of students, and on topics requested by the cluster teachers such as compacting curriculum and tiering lessons. All treatment-school teachers received opportunities to attend local and state conferences and institutes throughout the study period, and time was scheduled during each semester for collaboration among teachers of high-achieving cluster students.

The school district administrators selected teachers to teach the high-achieving cluster classes. Teachers were given the opportunity to volunteer to teach these classes and from those volunteers, administrators selected those whose credentials and experiences matched the position. Gentry and Mann (2008) suggested that teachers identified to teach the high-achieving cluster be placed for a minimum of three years in order to allow for a year of transition and learning and then two years to solidify pedagogy and practices in the classroom.

Data Sources

Student achievement as measured by the northwest education assessment. Student achievement was assessed three times each year (fall, winter, and spring), beginning in spring 2007, using the Northwest Education Assessment Measures of Academic Progress (MAP) (NWEA: NWEA, 2005) in the areas of reading, language usage, and mathematics. The NWEA is a widely used, well-known test of academic achievement. Tests are state standards-aligned, computerized, adaptive assessments that provide information about student achievement progress.

Correlations between the reading and language-usage test scores were computed to determine if both these scores needed to be modeled. The dataset for this study did not contain item-level information for the three NWEA sub-tests. Therefore, validity and reliability evidences for this particular sample could not be established, however these evidences were available based on the general sample in the manual (NWEA, 2004).

Reliability. Reliability is a measure of the consistency of a test score. Several forms of reliability can be estimated including marginal reliability (indicating internal consistency) and test-retest reliability (indicating stability of test scores over time), these are calculated for the NWEA and reported in Tables 4 and 5 (NWEA, 2004). Marginal reliability can be calculated for tests, such as the NWEA (2005), that were created using Item Response Theory (IRT). The marginal reliability is in essence the average of the error across different points in the test (NWEA). *Cronbach's Alpha* is the commonly used index of internal consistency. However, this coefficient is sample-dependent and may vary across samples. Therefore, reliability is determined in the NWEA using *marginal reliability coefficients* (*r*).

Table 4

Marginal Reliability (r) for NWEA MAP Test for 1999

Content Area		Grade Level 3		4		5		6	
		Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
Reading	*r*	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94
	N	39,590	48,566	39,960	52,602	40,671	54,254	35,508	52,696
Math	*r*	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.95
	N	37,022	47,635	37,237	52,580	37,933	53,753	33,131	52,581
Language Usage	*r*	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	N	20,769	19,676	21,593	23,167	21,980	25,304	20,035	23,389

Note. Adapted from "NWEA Reliability and Validity Estimates: Achievement Level Tests and Measures of Academic Progress," by NWEA, 2004, unpublished report, p.5.

Table 5

Test-Retest Reliability for NWEA MAP Test for 1999

Content Area	Term		Grade Level 3	4	5	6
Reading	Fall to Spring	*r*	0.85	0.88	0.89	0.89
		N	27,460	30,091	34,525	30,079
Math	Fall to Spring	*r*	0.79	0.86	0.89	0.91
		N	26,522	30,100	34,073	29,730
Language Usage	Fall to Spring	*r*	0.85	0.89	0.89	0.90
		N	14,173	17,285	19,037	16,825

Note. Adapted from "NWEA Reliability and Validity Estimates: Achievement Level Tests and Measures of Academic Progress," by NWEA, 2004, unpublished report, p.5-6.

Validity. Validity is a measure of whether the test actually measures what it intends. Most of the validity evidence for the NWEA is concurrent validity expressed as a *Pearson correlation coefficient* (*r*). Students took both the NWEA and another standards-based test. Strong positive correlation provides evidence of the concurrent validity of a test. The NWEA yields data that are concurrently valid with the Arizona, Colorado, Illinois, Indiana, Minnesota, Nevada, Texas, Washington, and Wyoming state assessments as well as the *Iowa Test of Basic Skills,* and the *Stanford Achievement Test 9th edition (SAT-9).* Correlation coefficients of the NWEA with state assessments for the reading test range from .76 to .86; for mathematics they range from .72 to .84; and for language usage they range from .60 to .79, but not all states had a concurrent language usage test.

Indiana statewide testing for educational progress – plus (ISTEP+). The concurrent validity with the Indiana Statewide Testing for Educational Progress - Plus (ISTEP+: IDOE, 2008) is the second lowest of the reported concurrent validities with small normative populations (reading: r = .77, language usage: r = .79, mathematics: r = .72). The NWEA assessment used in the state of Indiana has been designed based on specific state standards (Adkins, 2007). This process led to the establishment of the concurrent validity with ISTEP+ data and also resulted in the development of state percentile rankings and proficiency cut scores for reading, language usage, and mathematics. The Indiana Department of Education offered little information regarding evidence of the reliability of the ISTEP+ data. The program manual for ISTEP+ 2008-2009 (IDOE, 2008) offered evidence of internal reliability and standard error of measurement for the 2003 test data. Table 6 summarizes these data.

Table 6

Reliability Estimates and Margins of Error for ISTEP+

Grade	English/Language Arts			Mathematics		
	N	Reliability	SEM	N	Reliability	SEM
3	6007	0.90	15	6025	0.90	15
4	N/A	N/A	N/A	N/A	N/A	N/A
5	N/A	N/A	N/A	N/A	N/A	N/A
6	7159	0.92	12	7171	0.92	13

Note. Adapted from “2008-2009 ISTEP+ Program Manual,” by IDOE, 2008, unpublished report, p.139-140.

These data indicated what appears to be high internal reliability estimates (Anastasi, 1988). However, these reliability estimates are for the 2003 test data, and there was no mention in the manual of whether the current test is the same as the 2003 test. The Language Arts prompts do change each year, which may change the reliability estimates and margins of error. Additionally, there were no data for the fourth or fifth grade tests. Therefore, in the absence of complete reliability evidence for the ISTEP+ it was not used as a measure of student achievement in this study.

Semi-structured interview protocol. This study replicated Gentry and Owen’s (1999) investigation of the efficacy of the cluster-grouping model. In order to assess differences in the outcomes of the model in an urban setting and reduce extraneous variables it is important to retain substantive components of the original study. Teacher practices and perceptions of the TSCG model were investigated using the original, 24-

question interview protocol with minor changes. Question two in the original protocol (Appendix A) noted that the cluster-grouping program had resulted in more students identified as high achieving or above average and fewer identified as low or low average. Gentry and Owen's (1999) interviews were done post hoc so results of the program were known. Analysis of the identification trends and interviews in this study were done simultaneously so these data could not be reported to teachers during the interview. Therefore, this portion of the question was removed. Questions 21 through 23 were added to the protocol to investigate teacher perceptions of the ability levels of students from low-income and/or culturally diverse families. These were added based on previous work by researchers who have shown that ethnicity and socio-economic status may influence the way in which educators perceive students' ability (Wyner, et al., n.d.). The questions are as follows:

21. How do you perceive the ability of students from underrepresented groups?
22. What factors influenced your perceptions?
23. Have these perceptions changed over time since the implementation of cluster grouping?

The semi-structured interview format provided a framework for structuring the data-gathering process in order to ensure some uniformity of information across interviewees. It also provided a guide to the interviewer as a reminder of the information that was important to gather from participants (Patton, 2002). The semi-structured interview process allowed for deviations from the protocol in the event that an interviewee offered information pertinent to the investigation that was not originally included in the protocol.

Qualitative inquiry is an attempt to understand the experiences of participants in the study setting (Patton). The purpose of the interview process was to attempt to capture the lived experience of those being interviewed, to learn about their experiences using their own terminology and perceptions within those experiences (Patton). The openness of the semi-structured interview provided a framework in which participants could express their own understandings, experiences, and perceptions in their own words, while ensuring that important information was gathered. Data were gathered regarding the pedagogy and practices used in classrooms as a result of implementing the TSCG model. All 24 questions were posed to teachers.

Most teachers remained as high-achieving cluster teachers for three years as suggested by the model (Gentry & Mann, 2008). Qualitative data were collected in the third year of the study in order to allow for in-depth information to be collected from these teachers. The number of interviews was determined by the number of teachers who agreed to participate from a random sample and a list of high-achieving cluster teachers. Data saturation was defined as the point where no new information concerning the phenomenon was gained through the addition of more informants (Bowen, 2008). This point was determined post hoc, as it was impossible to determine prior to the study (Patton, 2002).

Information gathered focused on teacher practices that might influence student achievement as well as on elements of successful programming for gifted and talented students. This information was asked of all teachers regardless of which clusters of students they taught. Questions were posed regarding the types of grouping used within the classroom; whether more students were identified as high achieving since the inception of the model; what ways teachers recognized talent in their students; what strategies teachers were used to provide challenge for students; and what curricula modifications were made to meet the needs of students. In addition to this information, teachers were asked how their classrooms were set up to meet the challenges of grouping students such as whether they used centers in the classroom for enrichment or interest; the extent of seatwork time in an instructional day; and the atmosphere in the classroom.

Teachers provided information regarding administrator support and perceptions of the TSCG model. Teachers were also asked how much administrative support they received in implementing the model. In addition, questions were asked about the general atmosphere and environment of the school, communication regarding the implementation of the model between teachers, parents, and administrators. Perceptions were also elicited about the usefulness of the model in meeting the needs of students; the usefulness of the model in increasing the identification of students from culturally diverse families as high achieving; staff perceptions of the ability of students from culturally diverse families; and what factors influenced these perceptions.

Data Collection

Quantitative Data. This study used extant data from the treatment and comparison schools. All students in grades two through six in the two schools took part in the NWEA (2005) testing. These data were provided for analysis along with identification data for each student in the treatment school. NWEA data were collected three times per year beginning in spring 2007. Identification data were collected once per year. After obtaining these data from the participating school district, they were assembled into a data file for analysis.

Qualitative data. The curriculum director for the school district oversaw implementation of Total School Cluster Grouping. Recruitment of interviewees was done in conjunction with her. Teachers were chosen using the above-described procedures and then recruited using a letter and consent form, which I disseminated according to the curriculum director's wishes. Participants were asked to provide consent by signing a consent form. Before the interview commenced, I reviewed the consent form with each participant in order to ensure that each interviewee was aware of what they had consented to. Participants were informed that they may withdraw from the study at any time or refuse to answer any interview question(s) without penalty. Interviews lasted approximately 20 minutes and were conducted during the school day with classroom coverage from a roving substitute teacher. Information about each interviewee was kept confidential, and a pseudonym was used when reporting any responses from an interview.

Maintaining confidentiality. Names, grade taught, school, cluster class taught, date, and time were recorded during the interview to maintain detailed records. This information, along with audio files and transcripts of the interviews, was maintained in a password protected, electronic file, and/or locked file cabinet in the principal investigator's office. These data were reported using pseudonyms for participants' names to maintain confidentiality.

Audio files and transcripts will be maintained in the manner described for an indefinite period of time. Consent forms and identifiable information, and the code-key for pseudonyms used will also be maintained in the Principal Investigator's locked file cabinet for an indefinite period of time.

Quantitative Data Analysis

Preliminary analysis with descriptive data. Preliminary data analysis included inspecting the data for errors and duplications. The datasets were cleaned, standardized, and merged for analysis prior to computing descriptive statistics. Plots of the achievement versus time, including ordinary least square regression lines, were produced for a random subset of the sample. Descriptive statistics included sample demographics, identification frequencies for each year, and the number of students at each assessment time-point. Descriptive statistics including means and correlations for NWEA subtest data were computed in order to determine those that should be included in the individual growth curve model.

Missingness. The missingness of these data was investigated. Frequencies of students at each assessment time period were calculated. In addition, frequencies of students were calculated for repeated assessments sessions. Identification descriptive statistics were computed using only those participants for whom there were complete identification records (i.e., data for all four identification periods). This was necessary to facilitate identification comparisons across socio-economic and ethnic groups as well as to allow for comparison of those identified for each cluster in each year of the study.

Although the extent of missing data was included in the analysis and was necessary for the analysis of identification data, the maximum likelihood method was used in individual growth curve model estimation. This method allowed parameters to be estimated based only on the existing data for each individual participant. In particular, the process of fitting sums individual likelihoods computed for individual participants and does not require imputation or deletion of cases for handling missing data (Bollen & Curran, 2006).

Inferential analysis with multi-level growth curve model. Research question one, *what effects does TSCG have on student achievement as measured by the NWEA*? was addressed using a multi-level growth-curve model. In order to answer a complex question such as this, a multi-level technique was the most appealing to use. It afforded me the opportunity to determine the direct effects of variables at different levels. This method of analysis provided the opportunity to determine whether higher-order variables served as moderators of lower-level variables (Hox, 2002). In this study model-fitting was done to determine if change in student achievement over time (lower-level variable) was moderated by cluster identification, socio-economic status, and ethnic background (high-level variables).

Historically, individual change has been measured over time using a pre-test / post-test perspective (Willett, Singer, & Martin, 1998). This method did not allow for the estimation of a true trajectory of individual change that can only be reflected through multiple measures logically spaced over time. A clear quantitative definition of the trajectory of change allows for the possibility of summarizing evidence across trajectories. Therefore, multi-level growth curve models define the parameters of change for an individual (Raudenbush, 2001). Individual change over time is tracked using an *individual growth model* (Singer & Willett, 2003; Willett, Singer, & Martin). This model tracks the individual change within a person as a function of time and is mediated by predictors. Change is modeled at level 1 as a function of within-person change over time.

One key point to note in the level-1 model is that the researcher may control the interpretation of the intercept parameter estimate. Often this parameter is re-centered to a point that makes sense for the data. For instance, if analysis only begins at age 11, an intercept point of age zero has no practical significance in the analysis and actually draws the trajectory too far down. Centering the intercept at age 11 provides a more substantive interpretation of the intercept, allowing researchers to understand levels of the outcome variable at a point relevant to the study. Centering the intercept at age 11 changes the focal point of the interpretation. In this instance, that is the individual's initial status at the beginning of the study. Although significant difference between treatment and control groups is not evidenced at their initial status, it provides a point at which to start observing the trajectories. Equally possible is centering the intercept at the end of the study. This indicates where individuals from both groups (treatment and comparison) were at the end of the study. This may show a significant difference and often offers a clear way to show treatment effects. This study centered the intercept at the last measurement occasion as changes in student achievement after three years of treatment is the major focus of the study.

An important assumption at level 1 is that the functional form (linear or non-linear) of the trajectory is the same for all participants, with differences found in magnitude and starting-point of that trajectory. Defining the functional form beforehand is based upon the researchers' knowledge of the phenomenon under investigation, results of descriptive analysis, as well as, the theory behind the outcome variable (Singer & Willett; Willett, Singer, & Martin). Leaving analysis at this level only allows researchers to determine the significant difference among people with regard to their intercepts (starting point) and slopes (magnitude of change), which does not take into account other possible predictor variables that may moderate the change.

Multi-level growth modeling affords researchers the ability to look at additional levels of analysis. A level-2 model allows for estimation of inter-individual differences in change based on time-invariant descriptors of the individuals in the sample. The coefficients in the *intercept-as-outcome* model summarize the magnitude of the relationship between individual growth estimates and the characteristics under investigation (Singer & Willett, 2003; Willett, Singer, & Martin, 1998; Snijders & Bosker, 1999). Estimation of the parameters and corresponding significance tests can be done by fitting both the level-1 and level-2 models to the data simultaneously. Level-2 models must account for both the average change as well as the heterogeneity within the sample.

An additional advantage of applying the *multi-level growth modeling* approach to longitudinal analysis is the ability to deal with unbalanced data structures. For instance, repeated measures data with fixed measurement intervals where some data are missing for all or some of the participants can be easily accommodated (Singer & Willett, 2003; Snijders & Bosker, 1999). Even participants with only one data point are included in the model as their data are used in the estimation of the intercept and slope, although their contribution is minor by comparison to the other participants.

Two separate sets of *multi-level model* fittings were performed in this study to model predictors of student achievement for both the treatment and comparison schools as well as for the treatment school alone. This was done to compare the treatment and comparison groups as well as to explain differences between identification sub-groups in the treatment group. Both models are described in the following sections.

Variables in the multi-level models a through d. At level 1 the outcome variable was *student achievement* as measured by NWEA (2005), denoted in the model by y_{ti}. The within-person predictor at level 1 was *time*, which referred to the testing occasion. NWEA was administered three times per year beginning in spring 2007 for a total of seven testing occasions during the three-year course of this study.

At level 2 three predictors were added. These were between-person predictors and resided at the individual level: First, the *schooltype* predictor was a dummy-coded variable (1, 0) specifying whether the individual received the treatment or comparison condition. This variable was added to the model to predict whether the treatment influenced individual achievement in a significantly different way from the comparison condition.

Second, *SES_ETH* was a combined variable in order to account for the correlation between ethnicity and socio-economic (SES) status. Analysis of the demographic data from the school district (see Table 3) indicated that several ethnic groups should be collapsed in order to allow for sufficient sample size for modeling. Ethnicity has been collapsed into *underrepresented* (i.e., African American, Hispanic, and Native American) and *represented* (i.e., White, Asian, and Other). These groups were based on the literature regarding underrepresentation of certain ethnic groups in gifted education programs (e.g., Donovan & Cross, 2002; Yoon & Gentry, 2009). Socio-economic status (SES) was operationally defined using the federal lunch program eligibility data supplied by the school district.

To account for the possible multicollinearity between ethnicity and SES, these two variables were combined into 4 groups: Low-SES/underrepresented, (2) low SES/represented, (3) not-low SES/underrepresented, and (4) not-low SES/represented. Frequencies of students who fall into these four categories were examined to determine whether sufficient numbers of students existed for each category.

Third, a set of *interaction* effects between *Schooltype* and the four categories of *SES_ETH* variables were investigated in the level-2 model to assess the effects of treatment or comparison conditions on students from different levels of SES and different ethnicities.

Data analysis for multi-level models a through d. Data were arranged in person-periods nesting *time* (testing occasions) within the individual. Given the longitudinal nature of this study and the inter- and intra-school mobility depicted in Table 3, some participants did not have scores for each of the seven testing occasions. The model was fitted twice, once using the full dataset and once using only those participants with all seven waves of data in order to examine whether missingness had an effect on the estimations in the model.

Several preliminary models (Models A, B, and C) noted in Table 8 had to be fit prior to fitting the models (Models D and E) to answer the research questions posed in this study. This was done as multi-level modeling requires fitting the *fully-unconditional model* and the *means-as-outcomes model* in order to calculate the intraclass correlation and test the functional form of the trajectory (linear or curvi-linear) respectively prior to adding predictors that would answer specific research questions. These models are described in the following sections.

Step one involved fitting the *fully unconditional model* (Model A). This model did not have any predictors at any level. Instead of modeling change over time, this model separated the variation in outcome:

Level 1: $y_{ti} = \pi_{0i} + \varepsilon_{ti} \qquad \varepsilon_{ti} \sim N(0, \sigma_{\varepsilon}^{2})$

Level 2: $\pi_{0i} = \beta_{00} + r_{oi} \quad r_{0i} \sim N(0, \sigma_{0}^{2})$ Model A

This initial fitting of the *fully-unconditional model* was done in order to assess whether the variation in outcomes lay within the individual or between individuals (Singer & Willett, 2003). The variance estimates calculated when fitting this model allowed for calculation of the Intraclass Correlation (ICC) $\rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{\varepsilon}^2}$, which describes the percentage of the total outcome variation under investigation that occurs between people (Singer & Willett). In addition, specification of this model provided a baseline for assessing further model fit when adding predictors to the model.

The second step in modeling was to add *time* (Model B) at level 1 to show the within-person time-effect association on achievement. Altering the level-1 model by adding the predictor of time alters what the level-1 residuals mean (Singer & Willett, 2003). The model contained a second residual variation at level 2 that depicted variation in the rate of change between individuals. There were still no predictors added to the level-2 model. Therefore, this model only stipulated that the individual growth parameter was the sum of an intercept and a level-2 residual.

Level 1:

$$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \varepsilon_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i} \qquad \text{Model B}$$

In addition the level-2 model had additional residual variance that depicted between-person variation in rate of change. This indicated between-person variation in initial status and rate of change. This estimation allowed me to determine whether between-person differences in change were due to between-person differences in true initial status or true rate of change.

The two models fit thus far assumed a linear form. A model was then tested for non-linear change as the data seemed to support this notion as well as theoretical and research evidence (Raudenbush & Bryk, 2002; Singer & Willett, 2003). This study involved children in elementary school, a time of great growth and development that may not be uniform in change over time. Therefore, this model was fit with a quadratic term for *time* (*time*[2]) to assess if change in student achievement was non-linear during the course of the study. Model C fit was:

Level 1:

$$y_{ti} = \pi_{0i} + \pi_{1i}(time)^2_{ti} + \varepsilon_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i} \qquad \text{Model C}$$

Once the functional form was set, parameters were added to the level-2 model in order to answer research questions posed in this study. This multi-level model analysis was used to answer Research Question (RQ) one: *What effects does TSCG have on student achievement as measured by the NWEA?* In order to answer parts a through c of the question different predictors were fit to the model in order to fit the most parsimonious model as well as adding parameters to the model that allowed me to answer the questions posed. Discussion of this model fitting follows:

This study investigated the difference in student achievement between a treatment and a comparison school. Model D was used to model data from both the treatment and comparison conditions. *Schooltype* was fit to the model in order to make the distinction between these two conditions.

Model D[3] was proposed to answer RQ 1 a - c is therefore:

Model D, Level 1:

$$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \varepsilon_{ti}$$

---

[3] Model D as specified here is the most complex model possible. Depending on the results of the foregoing analyses, some predictors may be removed.

Model D, Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01}(Schooltype)_i + \beta_{02}(SES\_ETH1)_i + \beta_{03}(SES\_ETH2)_i +$$
$$\beta_{04}(SES\_ETH3)_i + \beta_{05}(Schooltype)(SES\_ETH1)_i + \beta_{06}(Schooltype)(SES\_ETH2)_i$$
$$+\beta_{07}(Schooltype)(SES\_ETH3)_{7i} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(Schooltype)_i + \beta_{12}(SES\_ETH1)_i + \beta_{13}(SES\_ETH2)_i +$$
$$\beta_{14}(SES\_ETH3)_i + \beta_{15}(Schooltype)(SES\_ETH1)_i + \beta_{16}(Schooltype)(SES\_ETH2)_i$$
$$+\beta_{17}(Schooltype)(SES\_ETH3)_{7i} + r_{1i}$$

Table 7

Model Fitting Process for Multi-Level Models of Change

Model	Level-1 Model	Level-2 Model
A	$y_{ti} = \pi_{0i} + \varepsilon_{ti}$	$\pi_{0i} = \beta_{00} + r_{oi}$
B	$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \varepsilon_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$
C	$y_{ti} = \pi_{0i} + \pi_{1i}(time)^2_{ti} + \varepsilon_{ti}$	$\pi_{0i} = \beta_{00} + r_{0i}$ $\pi_{1i} = \beta_{10} + r_{1i}$
D	$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \varepsilon_{ti}$	$\pi_{0i} = \beta_{00} + \beta_{01}(Schooltype)_i + \beta_{02}(SES\_ETH1)_i + \beta_{03}(SES\_ETH2)_i +$ $\beta_{04}(SES\_ETH3)_i + \beta_{05}(Schooltype)(SES\_ETH1)_i + \beta_{06}(Schooltype)(SES\_ETH2)_i$ $+\beta_{07}(Schooltype)(SES\_ETH3)_i + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(Schooltype)_i + \beta_{12}(SES\_ETH1)_i + \beta_{13}(SES\_ETH2)_i +$ $\beta_{14}(SES\_ETH3)_i + \beta_{15}(Schooltype)(SES\_ETH1)_i + \beta_{16}(Schooltype)(SES\_ETH2)_i$ $+\beta_{17}(Schooltype)(SES\_ETH3)_i + r_{1i}$
E	$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{ti} + \varepsilon_{ti}$	$\pi_{0i} = \beta_{00} + \beta_{01}(Identification)_{1i} + \beta_{02}(SES\_ETH1)_i + \beta_{03}(SES\_ETH2)_i +$ $\beta_{04}(SES\_ETH3)_i + \beta_{05}(Identification)(SES\_ETH1)_i +$ $\beta_{06}(Identification)(SES\_ETH2)_i + \beta_{07}(Identification)(SES\_ETH3)_i + r_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(Identification)_i + \beta_{12}(SES\_ETH1)_i + \beta_{13}(SES\_ETH2)_i +$ $\beta_{14}(SES\_ETH3)_i + \beta_{15}(Identification)(SES\_ETH1)_i +$ $\beta_{16}(Identification)(SES\_ETH2)_i + \beta_{17}(Identification)(SES\_ETH3)_i + r_{1i}$

Note. Model D used for modeling data from full sample. Model E used for modeling data from treatment school data only

Research question 1a, *What differences exist between treatment and comparison student trajectories among four groups based on income and representation status: (1) low SES/underrepresented; (2) low SES/represented; (3) not-low SES/underrepresented; (4) not-low SES/represented*? assessed any differences that existed in student growth, between the treatment and comparison schools, among the four *SES_ETH* groups. Parameter estimates and interaction effects for each *SES_ETH* group in both the treatment and comparison conditions were evaluated in order to answer this question. The amount of variation explained by *SES_ETH* was calculated for this model.

Research question 1b used the same model to investigate the second part of question one: *Do treatment school students in a TSCG program out-perform matched comparison school students after three years in the program after controlling for SES and ethnic representation?* Since *SES_ETH* was controlled for in the RQ 1a analysis, this question was answered by calculating the amount of variation explained by the main effect of *schooltype* after controlling for *SES_ETH.*

Research question 1c*, How do the learning trajectories differ between treatment and comparison students?* involved interpretation of the β_{11} fixed effect of Model D and its associated significance test. The β_{11} parameter is a level-2 slope that represented the effect of the predictor, namely *Schooltype*, on individual achievement growth rates. Estimation of this parameter addressed the question of the average trajectory of true change that was associated with *schooltype.*

Variables in multi-level model e. Model E was used for modeling data from the treatment school only.

The model specified at level 1 remained the same as Model D. This was due to the need to track individual *i*'s achievement over *time*.

Model E, Level 1:

$$y_{ti} = \pi_{0i} + \pi_{1i}(time)_{it} + \varepsilon_{it}$$

The predictors in the level-2 model that may have influenced the change in achievement over time for students in the treatment school were different in order to answer the research questions posed and are discussed below:

Model E, Level 2:

$$\begin{aligned}\pi_{0i} &= \beta_{00} + \beta_{01}(Identification)_{1i} + \beta_{02}(SES\_ETH1)_i + \beta_{03}(SES\_ETH2)_i + \beta_{04}(SES\_ETH3)_i \\ &+\beta_{05}(Identification)(SES\_ETH1)_i + \beta_{06}(Identification)(SES\_ETH2)_i \\ &+\beta_{07}(Identification)(SES\_ETH3)_i + r_{0i}\end{aligned}$$

$$\begin{aligned}\pi_{1i} &= \beta_{10} + \beta_{11}(Identification)_i + \beta_{12}(SES\_ETH1)_i + \beta_{13}(SES\_ETH2)_i + \beta_{14}(SES\_ETH3)_i \\ &+\beta_{15}(Identification)(SES\_ETH1)_i + \beta_{16}(Identification)(SES\_ETH2)_i \\ &+\beta_{17}(Identification)(SES\_ETH3)_i + r_{1i}\end{aligned}$$

The predictor, *schooltype,* was removed and *identification* was added into the level-2 model both as a predictor on its own as well as in an interaction term with *SES_ETH*. *Schooltype* was no longer needed as only the data from the treatment school was modeled to answer RQ 1 d and e. *Identification* was coded into the five ordinal categories as defined by the TSCG model and coded as such. This *identification* parameter denoted the individual's *identification* category at the beginning of the treatment period, with acknowledgment that this category may have changed during the course of the treatment. However, adding *identification* as a time-varying predictor in the model was not possible due to the difficulty of coding both the treatment *identification* categories as well as the comparison category of no *identification* at all. Therefore, in order to prevent errors in estimation, *identification* was entered as a time-invariant predictor of initial *identification* at the beginning of the treatment and modeled to assess the influence of initial group membership as a predictor of change in *achievement*. For example: how does *achievement,* measured by NWEA scores, change over time for individuals originally categorized in the *low-achievement* group? This question was posed for all 5 *identification* categories.

Data analysis for multi-level model e. Research question 1 d, *for treatment students only, based on initial identification categories, how do student achievement growth trajectories differ over three program years?* fit model E using *identification* as a predictor. The *SES_ETH* variable continued to be modeled. The β_{11} parameter estimates were evaluated for an answer to this question. β_{11} described the degree of difference in achievement between children in each of the five *identification* categories. Parameter estimates provided information about whether this difference was significant or not. Modeling of this level-2 model only provided a partial answer to the question therefore further descriptive information including figures and graphs added information to the results and interpretation of these data. Graphs of fitted trajectories of prototypical individuals provided graphical representation about differences or similarities between intercepts and rates of change of student achievement based on individual initial cluster group identification at the beginning of this study. Descriptive statistics describing the tenability of the model assumptions and mean student achievement level by initial *identification* category were also reported.

Research question 1e*,* specifically: *for treatment school students only, using four groups of students based on income and representation status, how do student achievement growth trajectories differ over three program years?* used the interaction term of *identification* and *SES_ETH* to investigate if there was a significant moderating effect of *SES_ETH* on initial *identification* as well as growth trajectories during the course of the three treatment years.

Identification data analysis. Research question two: *What effects does TSCG have on identification categories? What changes occur in frequency of students in each of the five achievement groups? by income? by ethnicity?* was answered in several ways. *Chi-square test for goodness of fit* was used to determine whether a significant difference existed in proportions of students identified for the high-achieving cluster for different socio-economic and ethnic groups (Kleinbaum, Kupper, Muller, & Nizam, 1998). The null hypothesis was that there was no difference in proportions between students identified as High Achieving and the school population; therefore a statistically significant difference indicated that the number of high-achieving students was not proportional to the school population. *Chi-square* analysis required the use of frequency data. Expected cell counts could not be less than five as this could lead to a false positive result indicating that there was a difference where one did not exist (Kleinbaum, Kupper, Muller, & Nizam). *Chi-Square* analysis in this study utilized identification data stratified by ethnic group to assess representation of culturally diverse students in the different cluster groups proportional to the school ethnic distribution.

These data were used to assess whether the model increased proportional representation of students in the high-achieving cluster compared to the school district population for each year of the study. *Chi-square* statistics and standardized residuals were calculated for each year of the study in order to address this research question. *Chi-Square* statistics were calculated for collapsed identification categories (below average, average, and above average), by race for the treatment school. Standardized residuals were then calculated to explain which ethnic and socio-economic groups in each school contributed to the overall significant *chi-square* statistics that may have indicated over- or under-representation.

Frequency data were also btracked for those students who had data for all testing occasions. These data were analyzed to investigate any changes in the number of students identified for different clusters over time. These data were also analyzed for any changes in ethnic and socio-economic proportions by cluster. It was only possible to analyze frequency data in this way for students who were present for the entire treatment period. More in-depth analysis using the entire sample was done using the multi-level modeling method.

Qualitative data analysis. Research question three: *What factors exist within the classrooms and school using TSCG that may influence student achievement?* was answered using qualitative, interview data. Interview transcripts from the treatment school were used to answer research question three.

A grounded theory approach was used to elucidate the themes that are present in the interview data (Glaser & Strauss, 1967; Strauss & Corbin, 1998). The grounded theory approach involves the process of deriving a theory from a body of qualitative data through repeated readings and coding of the data. Grounded theory involves comparison of different sets of data or cases (interviews in this case) for the presence of similar themes. The deduction of these themes then becomes generalized to form theories regarding the phenomenon under investigation (Glaser & Strauss).

I transcribed and hand-analyzed interviews through multiple readings of the transcripts. During this process of open-coding (Strauss & Corbin, 1998) important themes and categories were recorded in the research codebook. Multiple readings of the transcripts provided the opportunity to gain a more holistic understanding of the data. This coding process began as an inductive process in which themes emerged from the data based on content alone. A second stage in the analysis included a deductive analysis of the data to find illustrations related to the research question.

To address credibility issues one third of the transcripts were coded along with another coder (S. Peters, personal communication, September 4, 2010) in order to train the coder in the processes being used in the analysis. After training, the coder and I each coded another third of the transcripts individually. Once individual analysis was completed we conferred as to our emergent themes and categories in order to assess the level of agreement in our analyses. Discrepancies in themes were resolved through discussion and consensus.

CHAPTER FOUR: RESULTS

This chapter includes the results of the study. Results of the multi-level model for change are presented and are followed by the results of the effects of the TSCG on identification categories for students from different income and ethnic groups. Qualitative results are presented about the teaching practices and TSCG model perceptions of both high-achieving cluster teachers and teachers of other clusters.

Research Question One

Missingness. In order to address any of the quantitative research questions missingness in the dataset was investigated due to the high transiency rate of the student sample in this study. This was necessary to ensure that the correct data analysis techniques were used. Several types of missingness were investigated.

Missingness is a fundamental part of research done with real data (Little, 1995; McKnight, McKnight, Sidani, & Figueredo, 2007). However, the reasons for missingness are many and varied. Missingness may hinder the ability of the researcher to explain and understand the phenomenon under investigation. Therefore, in general, missingness can threaten the validity of study results (McKnight et al., 2007).

There are different reasons why data are missing: (1) Study participants drop out prior to the beginning of treatment or during the study. (2) Missingness is also caused by the interaction of participants with the study design. This may occur when a participant is placed in a treatment that results in improvement of symptoms for example, depression. If a participant feels less depressed he may be less likely to continue in the study causing missingness. These reasons may affect the amount and patterns of missing data and dictate the appropriate data analysis techniques (Little, 1995; McKnight et al., 2007). The stage of the study when missing data occur can also affect results and interpretation. Data can be lost during recruitment, treatment, and/or follow-up. Data loss during recruitment may be associated with criteria for choosing participants, drop out before treatment commences, or participant change-of-mind after signing the consent form. Data loss during treatment may be due to skipped questions or testing occasions, absence during testing windows, or participant drop out. (3) Follow-up drop out is also prevalent in longitudinal studies as participants may move and become difficult to locate (McKnight et al.). (4) Study design incorporates the use of cohorts as participants advance through the treatment. This will result in attrition when participants reach a particular level. This study used a cohort design with four graduation years as students moved through grades two through six. Therefore, when students completed grade six they exited the study causing missingness.

The effects of missingness. The most important issue regarding missingness is the extent to which the missingness influences study results (Hedeker & Gibbons, 1997; Little, 1995; McKnight, et al., 2007). The types of data that are missing might influence outcomes of an intervention or treatment. For example, benefits may be overestimated due to a lack of negative outcome data and vice versa (McKnight et al.). Another aspect of missingness is the process that might cause the data to be missing. If a post-test is perceived to be difficult then those who think they may not pass the test may choose not to take it. Therefore, any inferences made about the effectiveness of the treatment would be invalid based on the lack of poor mathematics results (Hedeker & Gibbons; Little; McKnight et al.). These mechanisms of missingness were not present in this study. Students who were present at school were all tested during each testing window regardless of previous test results. This removed performance-based dropout as an issue to be resolved.

Multiple measures of the same construct, in this study, achievement, can reduce the bias caused by missingness (McKnight et al., 2007; Singer & Willett, 2003). If one or two measurement occasions have missing data, the others can still offer information about the construct under investigation. However, missingness can still pose a problem for multiple measures. For example if ethnic information is missing and the study requires an assessment of achievement by ethnicity no other data source will fill in those gaps (Singer & Willett; McKnight et al.). This was only a minor issue in this study as most of the ethnicity data were present in the data.

Types of missingness. Missingness Completely At Random (MCAR) exists when missing values are distributed randomly across all observations. The missing data mechanism is not related to the value of any variable, either observed or missing. In a true MCAR situation, complete case analysis will yield the same results as full dataset analysis. This is not a realistic assumption in most real-life datasets (McKnight et al., 2007). Missingness At Random (MAR) occurs when the missingness is not dependent on the missing values, but may be dependent on the value of other observed variables. MAR missingness is not randomly distributed across all observations, but is randomly distributed within one or more subsamples in the dataset. Covariate Dependent missingness (CD) occurs when missing values are related not to the values of the response variables, but only to the covariates. For example missing achievement data may not be related to a low or high previous score, but may be related to the socio-economic status of the participant. Missing Not At Random (MNAR) is missingness that is not randomly distributed. This type of missingness is sometimes referred to as non-ignorable missingness. This type of missingness is related to the missing values in the dataset. This can occur when participants do not want to reveal personal information that may prejudice them either negatively or positively. For example students who previously performed badly on a test are less likely to retest than those who previously performed well. Whether the data are missing or not is related directly to the value. Complete case analysis in the presence of this type of missingness is highly biased.

Ignorable missingness investigation. Diagnostic procedures were run to diagnose the type, extent, and pattern of the missing data. These diagnostics informed me about inferential limitations and how best to handle the missing data. Little (1988) proposed a test to reliably detect whether data are MCAR. If data are found not be MCAR then the missingness must be considered either MAR or MNAR. Although MAR and MNAR are clearly distinguished by definition it is difficult to distinguish between MAR and MNAR within the data. One conclusion that can be drawn from the distinction between MCAR and either MAR or MNAR is that data MCAR can be ignored in the analysis, while only certain types of data MAR can be ignored. If MCAR is ruled out then the researcher must try to distinguish between MAR and MNAR using logic and a good understanding of the study design as no empirical test is available to make the distinction.

Data MAR can be ignored in four situations proposed by Schafer (1997): (1) double sampling where a subsample of those participants who are missing data are re-sampled to obtain further information on the variable under investigation for data are missing; (2) survey sampling in which groups of non-respondents are subject to intensive follow-up to determine the reasons for non-response; and (3) in unbalanced randomization trials MAR can be assumed when the unbalanced nature of the data is by chance. .Missing data in the double sampling are assumed to MAR. Little (1995) further noted that while MAR is dependent only on the observed data and has been termed random dropout there is also a class of ignorable missingness referred to as covariate-dependent dropout (CDD). The main difference between MAR and CDD is that while MAR is dependent on observed values previous and future, the CDD assumption is true when missing data are dependent both on between-subjects and within-subject covariates that are treated as fixed effects in the model (Hedeker & Gibbons, 1997; Little, 1995). Fairclough (2002) also noted that the CDD assumption is true when participants with missing data are a random sample within each treatment group. For example, culture may influence the follow-up response in an international study of breast cancer. Missingness has no bearing on the outcome, but on the cultural morés of the participants. In this study, missingness is not dependent on previous scores as students were not given the choice to test. If they were present at school on test day, they are tested. Missingness seemed to be dependent on the socio-economic status of students. Different patterns of missingness were evident in the four SES_ETH groups present in this study. This type of missingness can be viewed as the Covariate Dependent Dropout special case of MAR and treated as such in the model.

Missingness Test Procedures

To determine whether missingness was completely at random, Little's MCAR Test (Little, 1988) was used. This test yields a non-significant result if missingness is completely at random (MCAR). A significant test result indicates that missingness is either MAR or MNAR. Results of Little's MCAR test were significant indicating that further testing was needed to determine whether missingness in this dataset was MAR or MNAR. I conducted Little's MCAR Test again with scores broken into SES_ETH groups. Results of these analyses only yielded two out of eight estimates significant. Further, I tested the montonicity of the missingness. I did this as there were at least two mechanisms of missingness in this study – cohort dropout and transiency. Monotonic missingness prevents the multiple imputation of data that could help yield unbiased results.

Although overall results of Little's MCAR Test were significant for both reading and mathematics scores, when scores were broken into subsets by SES_ETH categories only two out of the eight tests were significant at the $p < 0.05$ level. These data satisfied Fairclough's (2002) requirements for MAR. The additional dependence on SES_ETH also fulfilled the requirements for CDD (Hedeker & Gibbons, 1997; Little, 1995). These results are depicted in Table 8.

Table 8

Little's Test of MCAR for Overall Reading and Math Data and by SES_ETH Category.

Category	Reading	Math
SES_ETH1	152.24*	126.48
SES_ETH2	38.97	37.92
SES_ETH3	157.56*	135.03
SES_ETH4	32.62	46.51

Note. * significant at $p < 0.05$ level.

These results indicated that missingness was most likely associated with the social and cultural characteristics of the students. This was further substantiated by the district's curriculum director (personal communication, November 11, 2008) who noted that the area in which the school district is located was an area of high transiency where families were often from minority groups and were migrant workers who moved frequently. This accounted for some of the intermittent missingness patterns found in the data as students moved away and returned. However, the majority of missingness patterns were monotone in nature, where once a student dropped out of the school they did not return within the timeframe of the study. Socio-economic status was a covariate built into the model and therefore was accounted for in the model specifications. Therefore, random effects modeling as initially proposed were appropriate for this study.

Missingness by design. McCoach (personal communication, March 25, 2010), Little (1995), and Hedecker and Gibbons (1997) noted that missingness that was built into the study design was ignorable. This study employed a cohort design following students in graduation years 2014 through 2017 from grade two through six. This design automatically resulted in attrition as students moved out of sixth grade and therefore out of the study.

Missingness descriptive statistics. Eight hundred two usable lines of data were used in these analyses. That is lines of data with at least one subtest score out of a possible seven. Missingness data were compiled for each of the waves of the outcome variables. Tables 9 and 10 show the frequency and percent of missing data for each of the waves, for both mathematics and reading subtests. For the total sample, missingness of the mathematics subtest scores ranged from a low of 14.1 percent for the second mathematics subtest to a high of 22.9 percent for the third mathematics subtest. Missingness of reading subtest scores ranged from a low of 14.0 percent for the second reading subtest to a high of 20.1 percent for the first reading subtest.

Table 9

Amount of Missingness for Mathematics Subtest for Total Sample.

Variable	Missing	
	Count	Percent
Mathematics 1	159	19.8
Mathematics 2	113	14.1
Mathematics 3	184	22.9
Mathematics 4	115	14.3
Mathematics 5	140	17.5
Mathematics 6	143	17.8
Mathematics 7	123	15.3

Table 10

Amount of Missingness for Reading Subtest for Total Sample.

Variable	Missing	
	Count	Percent
Reading 1	161	20.1
Reading 2	112	14.0
Reading 3	184	22.9
Reading 4	116	14.5
Reading 5	138	17.2
Reading 6	142	17.7
Reading 7	124	15.5

These tables indicate that even with the missingness there were substantial observed data for use in fitting the multi-level model. Figures 1 and 2 indicate the patterns of missingness in the seven waves of data for both the mathematics and reading subtests.

These charts are ordered so that variables and patterns reveal monotonicity in the data. Variables are ordered from left to right in increasing number of missing values. Therefore, the third wave of mathematics testing has the most missing data points while the second wave has the least. In addition patterns are sorted from right to left with non-missing values first and then missing values. In order for the missingness to be monotonic all missing cells and all non-missing cells in the chart will be contiguous. There should be no islands of non-missing data (white lines) in the bottom, right corner. In addition there should be no islands of missingness (dark lines) in the upper left corner. These data show that the mathematics subtest data were not completely monotonic in nature, although there were very few islands in the data. This indicated that imputation of the data might be possible to yield unbiased results (SPSS, n.d.).

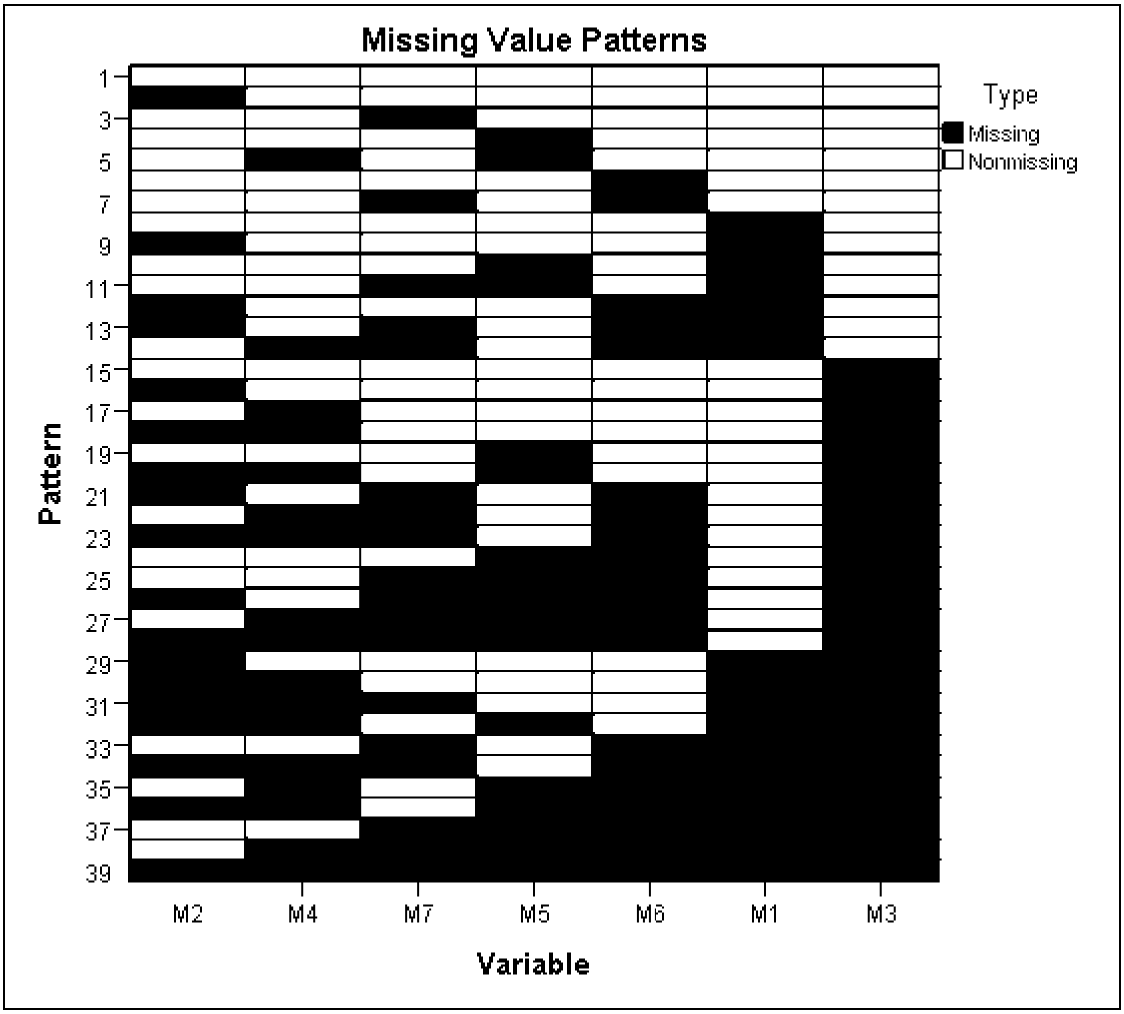

Figure 1. Patterns of missingness for Mathematics Subtest by Prevalence of Pattern.

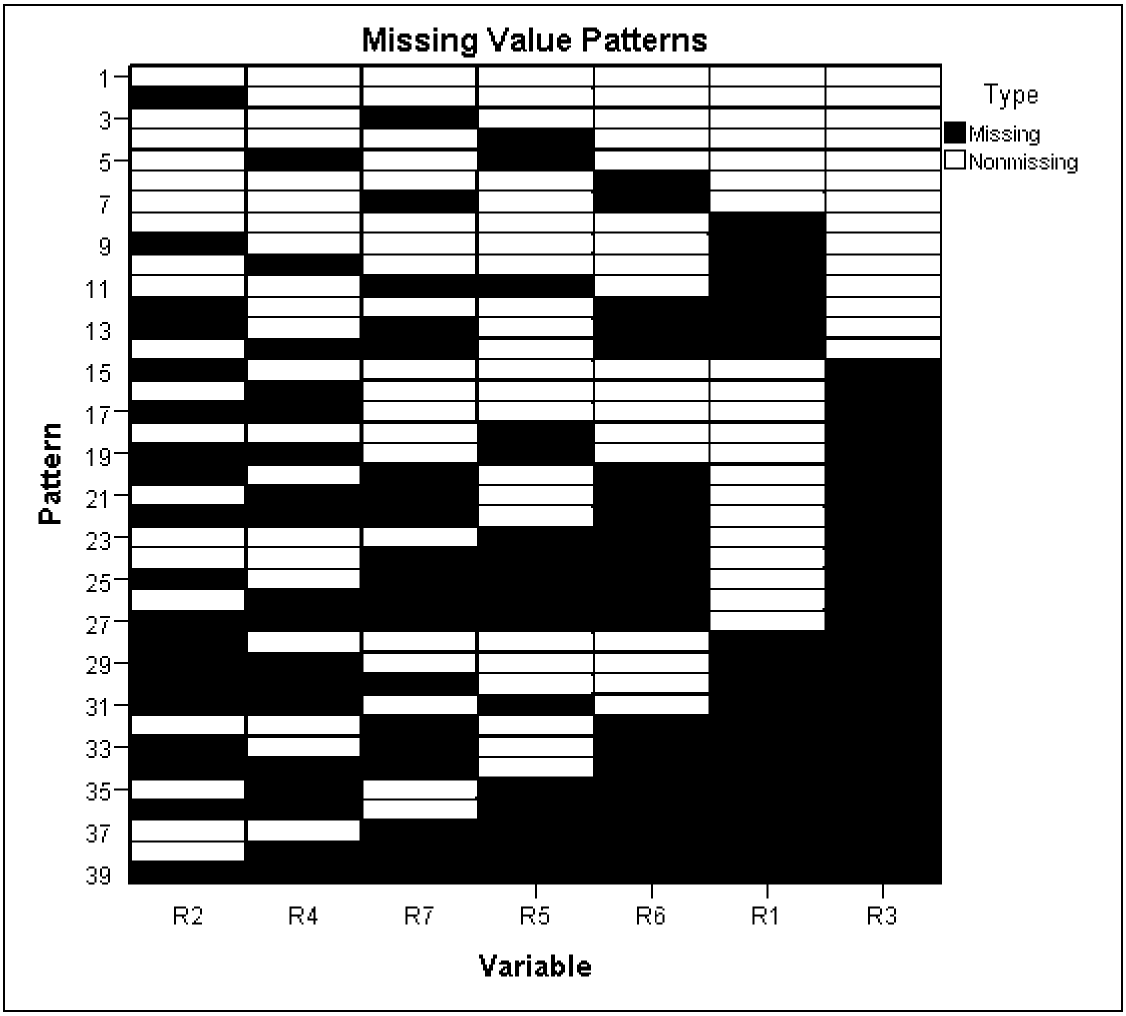

Figure 2. Patterns of missingness for Reading Subtest by Prevalence of Pattern.

In addition to the above figures the following two figures (Figures 3 and 4) are companion bar charts that indicate the percentages for each of the top ten patterns of missingness. The chart of mathematics missingness patterns shows that more than 70 percent of the data is present in missingness pattern number 1. Referring back to Figure 1, pattern number one denotes no missing data points over the course of the study. The rest of the charted patterns make up less than ten percent of the missing data patterns respectively indicating little missing data in this data set. Figure 3 for the reading subtest shows a similar trend. Over 70 percent of the data falls into missingness pattern number one, no missing data points.

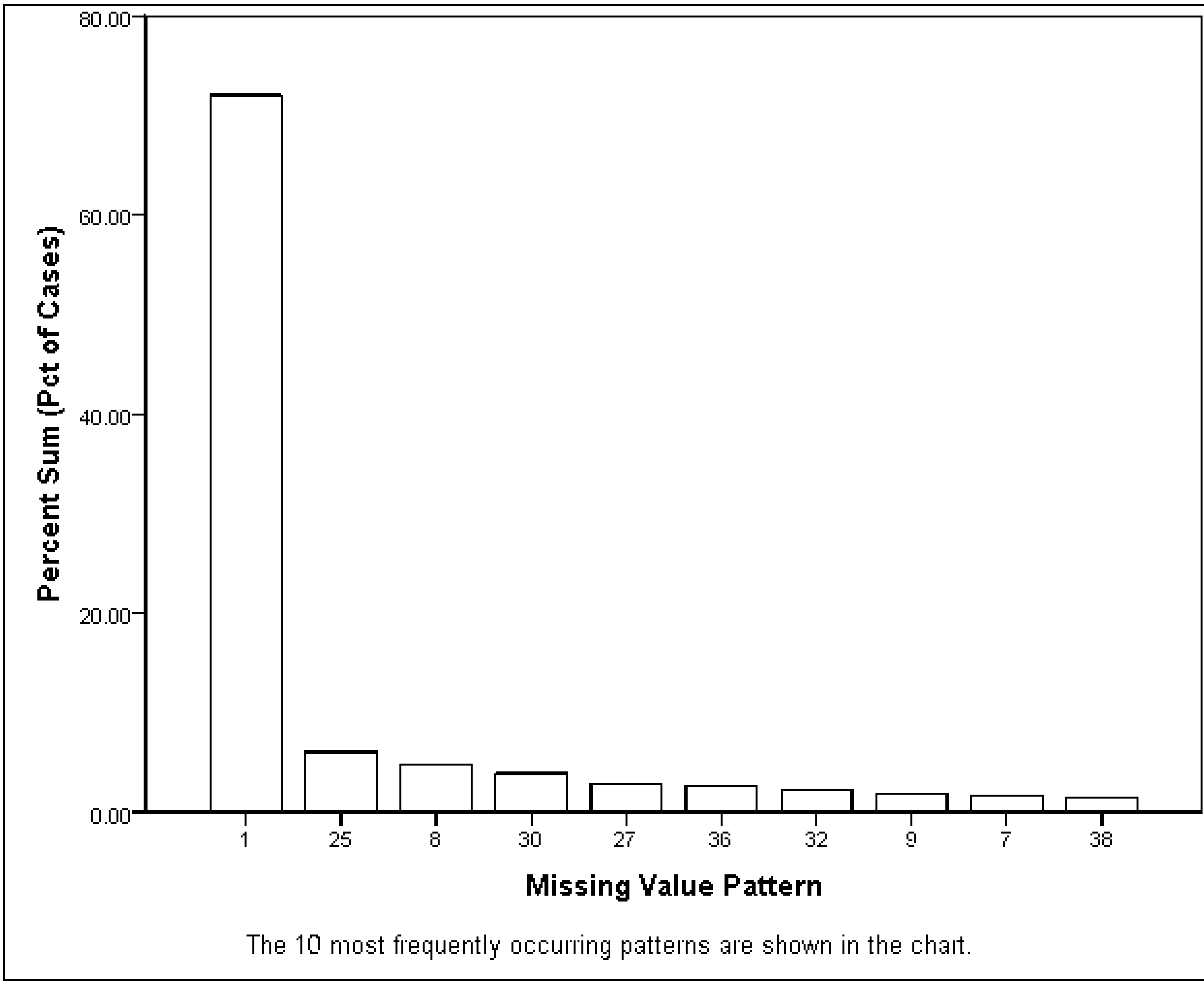

Figure 3. Most Frequently Occurring Missingness Patterns for Mathematics Subtest.

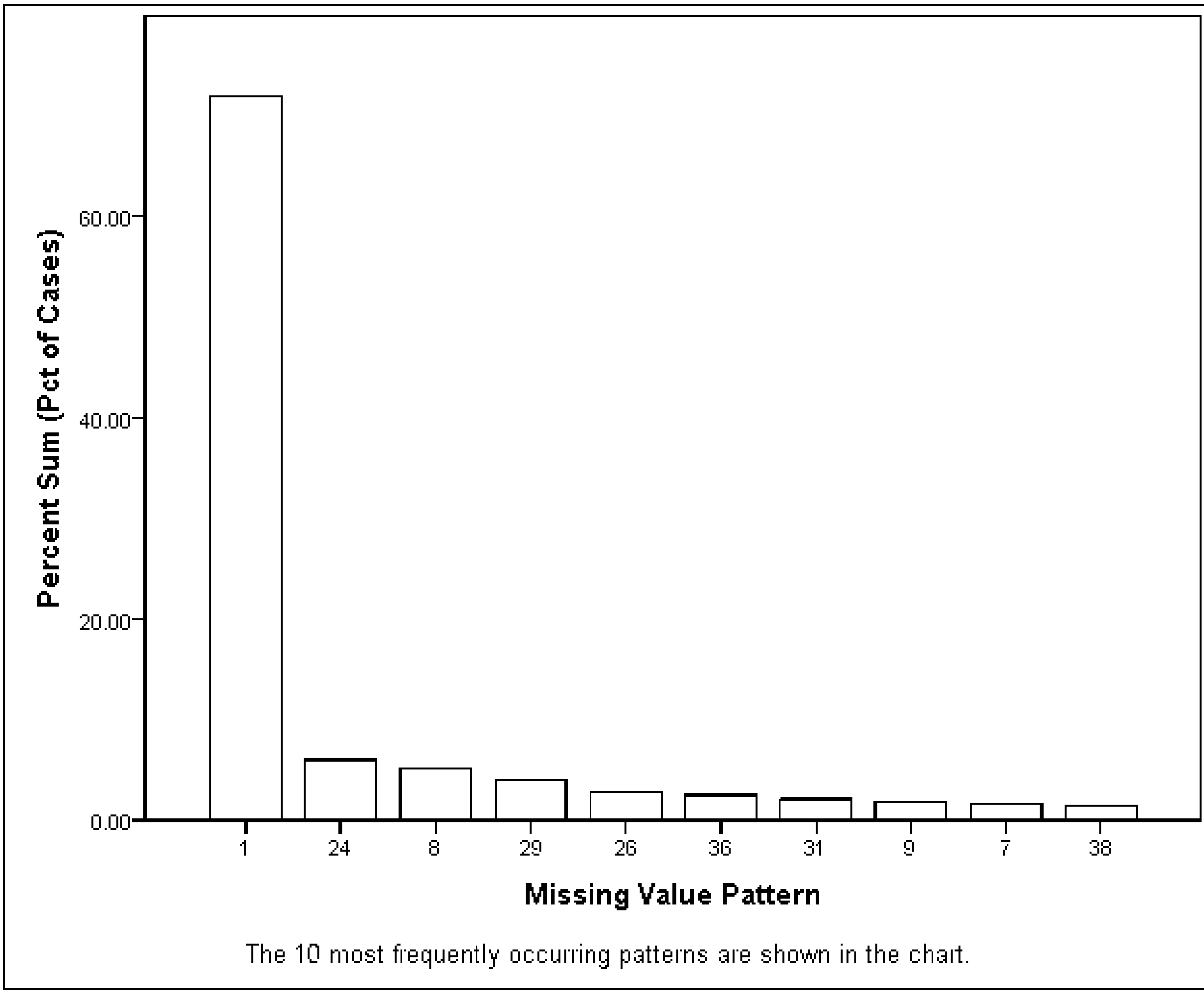

Figure 4. Most Frequently Occurring Missingness Patterns for Reading Subtest.

Further analysis was conducted to determine how the missing patterns of various variables influenced the outcome variables. Comparisons were computed for categorical variables and continuous variables. Frequencies and percentages were generated. This enabled me to determine any patterns of covariate missing values. Tables 11 and 12 display results for the total sample, by SES_ETH variables one through four.

Table 11

Percentage Missing Mathematics Subtest Scores by SES_ETH

Variable	Percentage Missing			
	SES_ETH1	SES_ETH2	SES_ETH3	SES_ETH4
Math 1	27.6	20.0	16.1	10.8
Math 2	20.4	15.6	12.9	4.6
Math 3	32.1	18.9	21.5	13.1
Math 4	19.9	13.3	10.8	6.5
Math 5	24.5	6.7	17.7	9.5
Math 6	28.6	6.7	18.3	10.1
Math 7	25.0	7.8	18.3	9.8

Table 12

Percentage Missing Reading Subtest Scores by SES_ETH

Variable	Percentage Missing			
	SES_ETH1	SES_ETH2	SES_ETH3	SES_ETH4
Reading 1	27.6	20.0	16.7	11.1
Reading 2	20.4	15.6	12.4	4.6
Reading 3	32.7	18.9	21.0	13.1
Reading 4	20.4	13.3	10.8	6.5
Reading 5	24.0	6.7	17.2	9.5
Reading 6	28.6	6.7	17.7	10.1
Reading 7	25.0	6.7	18.3	10.1

Missingness was highest for the Low-SES, underrepresented (SES_ETH1) group for all testing occasions on both subtests. The second highest rate of missingness was for the not-Low SES, underrepresented group (SES_ETH3), except for the first testing occasion. The lowest incidence of missingness was for the not Low-SES, represented group (SES_ETH4). Percent of missingness was fairly stable for three of the SES_ETH groups throughout the study with the exception of testing occasion three, which showed a marked increase in missingness across all SES_ETH groups. The data for SES_ETH2 group indicated that missingness dropped substantially at testing period five and continued at the same level for the duration of the study. This may be due to changes in boundaries which moved some middle-income neighborhoods out of the treatment school catchment area and brought more students from Caucasian, lower-income families into the school zone (Principal Treatment School, May 15, 2009).

Conclusion. Data must be assumed as MAR or MNAR given the MCAR test results. Based on the analyses of the patterns of missingness it is plausible that the missingness can be categorized as covariate-dependent dropout and therefore is ignorable when the covariate is included in the model. Laird (1988) noted that when maximum likelihood is used for random-effects longitudinal modeling, valid inferences can be made in the presence of ignorable missing data. However, given the deviation of missingness pattern for graduation year 2014 it was appropriate to model the data in several ways to evaluate whether this had a significant effect on the parameter estimates obtained in modeling these data.

Comparison analyses. These data were modeled assuming the data were missing at random (MAR). Next, data were modeled using a Markov Chain Monte Carlo multiple imputation method with ten iterations. Lastly, the data were modeled using the available case method to evaluate if parameter estimates for initial Models A through C remained reasonably close. Imputed datasets were combined using SAS procedure MIANALYZE and parameter estimates generated for the longitudinal models. Imputed parameter estimates were the most divergent of the three methods and standard errors were the largest for this analysis indicating that imputing was not the most reliable method.

Model D and E parameter estimates were almost unchanged between modeling with missing data and with complete cases and are discussed in the context of answering the research questions. The lack of variance in parameter estimation between these data sets supports the randomness of the missingness patterns (McKnight et al., 2007). Details of the Model A through C analyses are depicted in Table 13.

Table. 13

Comparison of Achievement Parameter Estimates Across Analysis Methods.

	Model					
	Reading A	Math A	Reading B	Math B	Reading C	Math C
With Missing Data	200.86	205.5	208.35	215.76	205.85	212.11
Complete Case	200.85	205.5	208.34	215.77	205.85	212.1
Multiple Imputation	201.128	205.612	204.636	210.46	201.307	205.861

Descriptive analyses. ***Correlations***. Reading and mathematics were negatively correlated with SES_ETH1, SES_ETH2, and SES_ETH3, but positively correlated with SES_ETH4. Thus students from families of lower socio-economic status achieved at lower levels than those who were from not-low income, represented families. The association between reading and mathematics achievement and socioeconomic factors was consistent with previous findings (Wyner, Bridgeland, & Diiulio (n.d.). School attended and reading and mathematics scores were positively correlated indicating that scores were somewhat dependent on where students were, i.e. whether they were in the treatment or comparison school. These are depicted in Table 14.

Table 14

Correlations Between NWEA Achievement Scores and Socioeconomic Status and School

Reading	SES_ETH1	SES_ETH4	School
Wave 1	-0.176**	0.202**	0.063*
Wave 2	-0.152**	0.168**	0.068*
Wave 3	-0.196**	0.189**	0.031
Wave 4	-0.201**	0.200**	0.056*
Wave 5	-0.181**	0.191**	0.019
Wave 6	-0.155**	0.208**	0.103*
Wave 7	-0.076**	0.254**	0.180**
Math	SES_ETH1	SES_ETH4	School
Wave 1	-0.149**	0.198**	0.054*
Wave 2	-0.135**	0.162**	0.056*
Wave 3	-0.17**	0.181**	0.047
Wave 4	-0.173**	0.195**	0.038
Wave 5	-0.172**	0.211**	0.028

Wave 6	-0.172**	0.198**	0.091*
Wave 7	-.052**	.0283**	.211**

Note. *$p < .05$ ** $p < .0001$.*

Descriptive analysis of individual change over time. A random selection of participants was selected to conduct exploratory analyses to describe how individuals in this study changed over time. Empirical growth plots were generated for 100 individuals. Plots were broken up into reading and mathematics plots and plotted over the seven testing intervals. Although summer losses can be seen in the dips in the plots, overall growth is positive. Another trend of note is that most students' achievement dipped at the last testing occasion. This was consistent across reading and mathematics. Figures 5 and 6 show a representative subsample of the 100 plots.

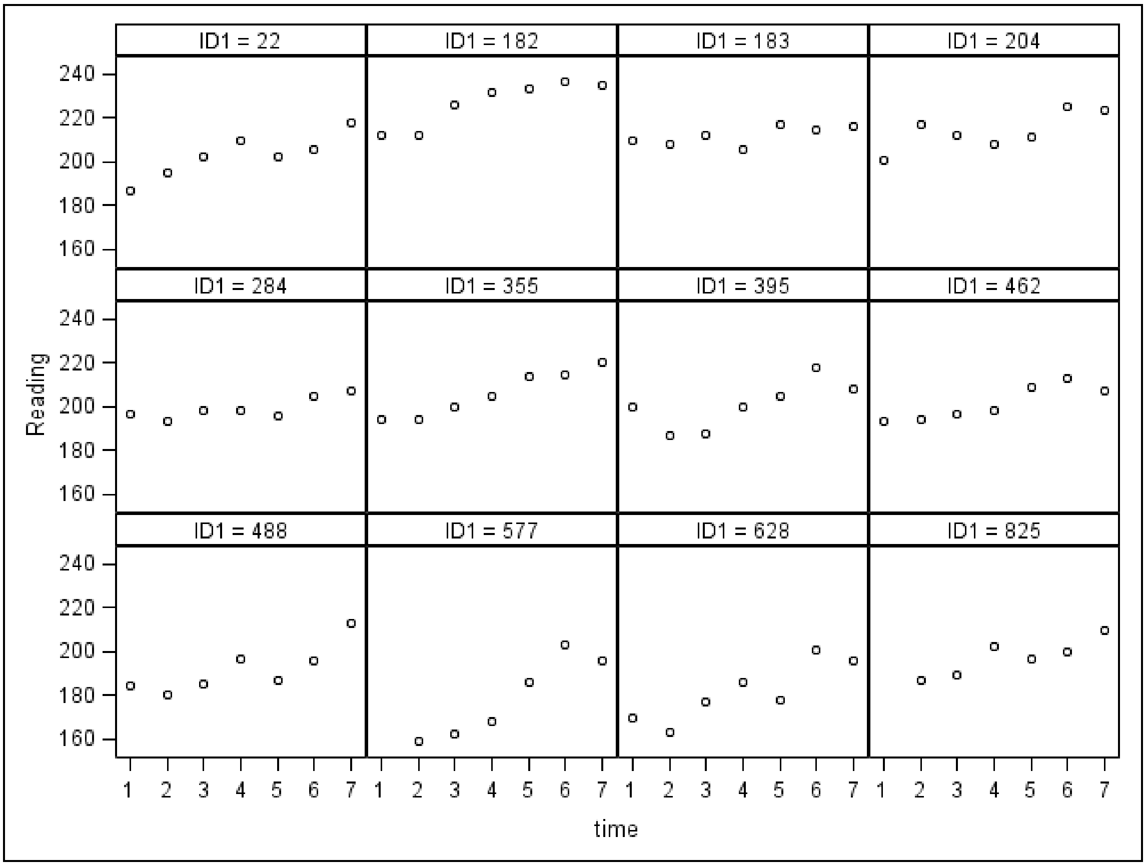

Figure 5. Empirical Growth Plots for Reading Scores for Twelve Study Participants.

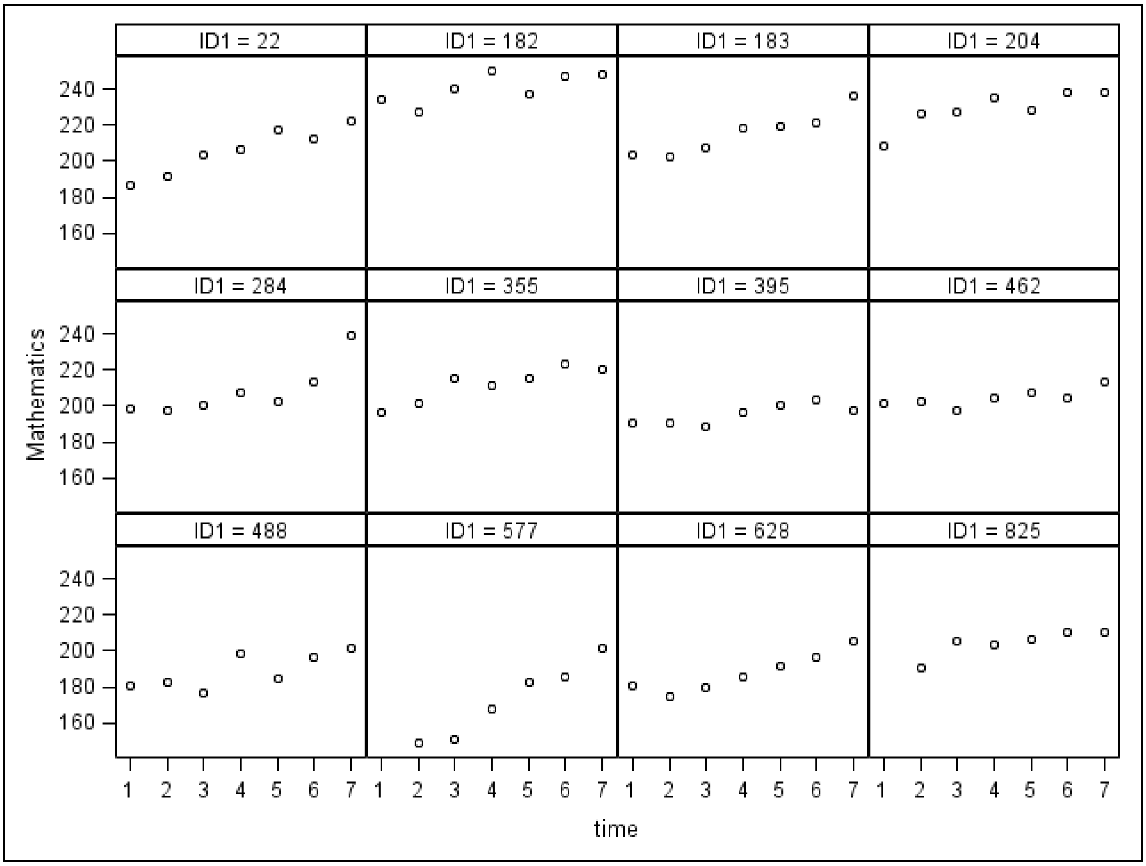

Figure 6. Empirical Growth Plots for Mathematics Scores for Twelve Study Participants.

In addition to the empirical growth plots, non-parametrically smoothed trajectories were also generated. The benefit of these plots was that no functional form was assumed for growth and could therefore be informative in the choosing of the functional form for model specifications. The same subsample was used for this analysis and their trajectories are summarized in figures 7 and 8. Examination of these trajectories as a group indicated that all students made some gains, some more than others. Participant 577 started low at the beginning of the study and made gains consistently throughout the study while some students began at a higher level and made small gains, such as participant 182. Several students made greater gains towards the end of the study after making smaller gains at the beginning, such as participant 395. No students showed negative gains throughout the course of the study.

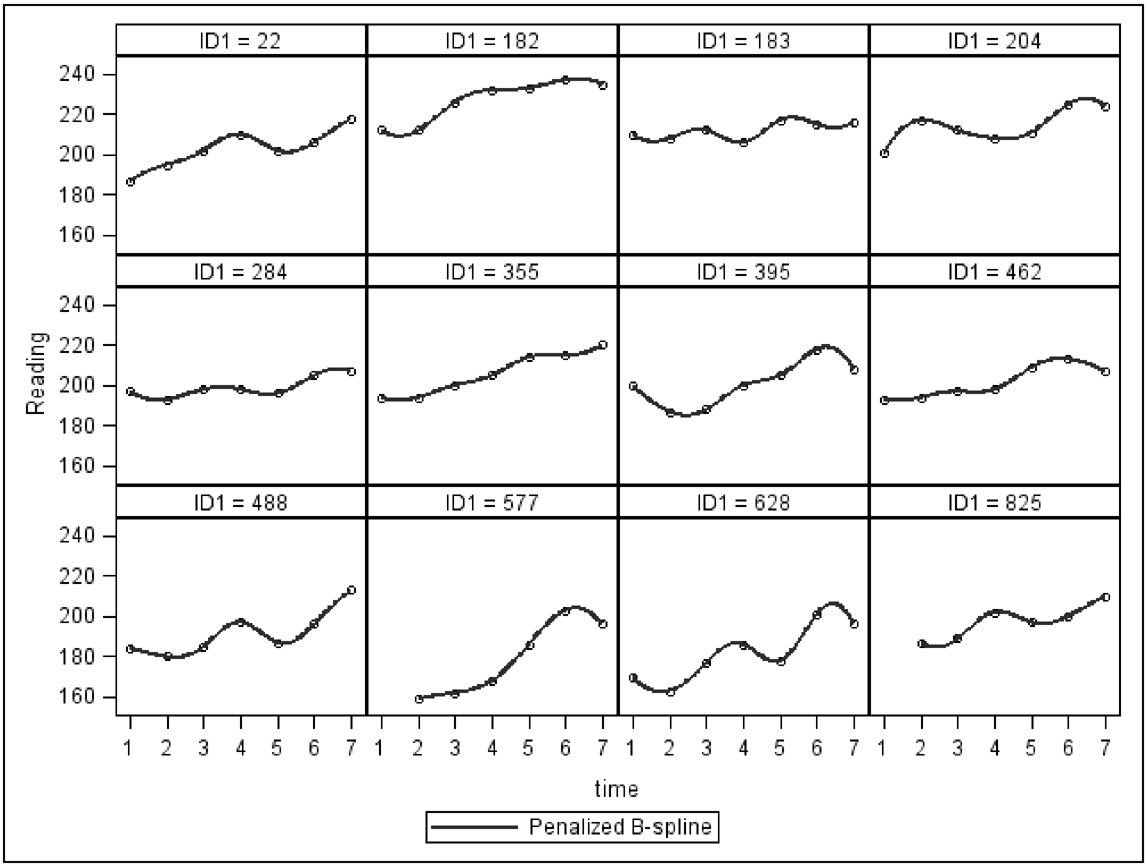

Figure 7. Non-Parametrically Smoothed Trajectories for Reading.

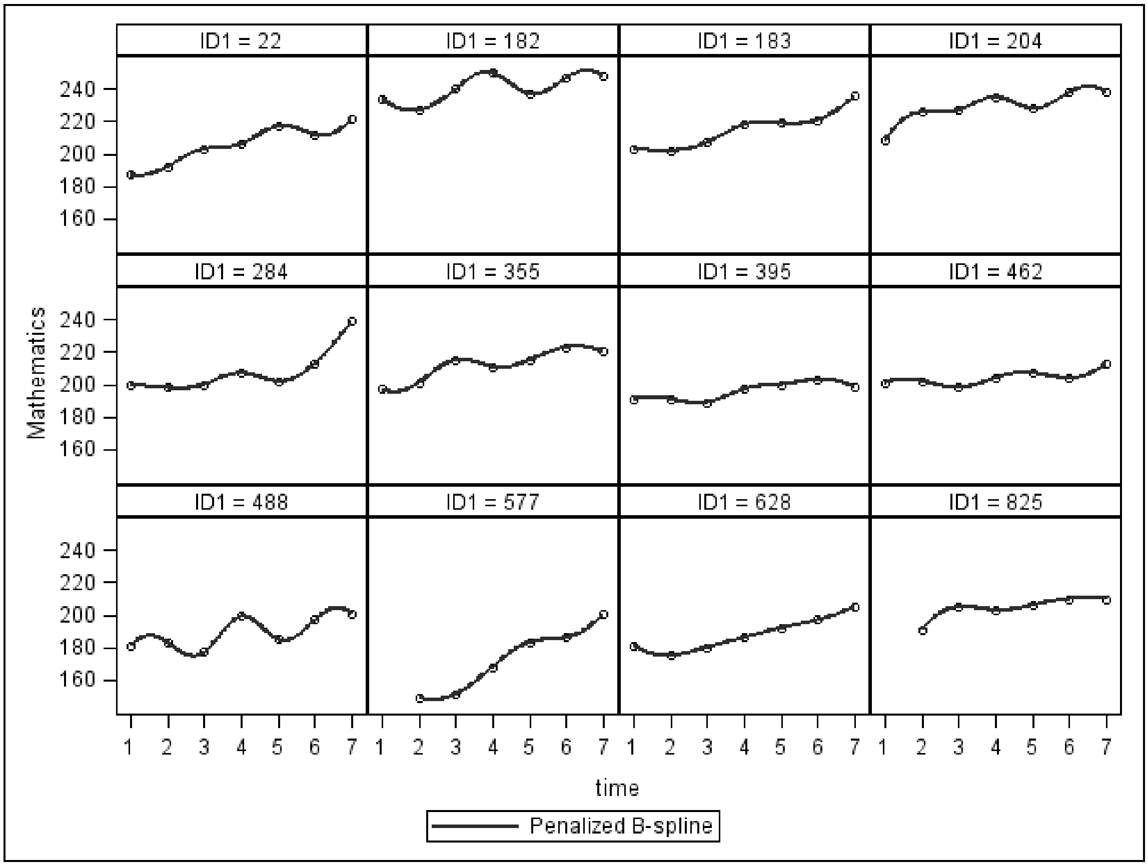

Figure 8. Non-Parametrically Smoothed Trajectories for Mathematics.

The next analysis was to fit a smoothed trajectory to each individual's data using Ordinary Least Squares (OLS) regression. Although there was some fluctuation in the individual growth of students through the course of the study a linear trajectory was adequate for exploratory purposes. Figures 9 and 10 depict the OLS summaries for the subsample. Although there was some deviation from the trajectory on some plots, overall it seemed that a linear trajectory was appropriate for these data. This was further investigated in the model fitting process.

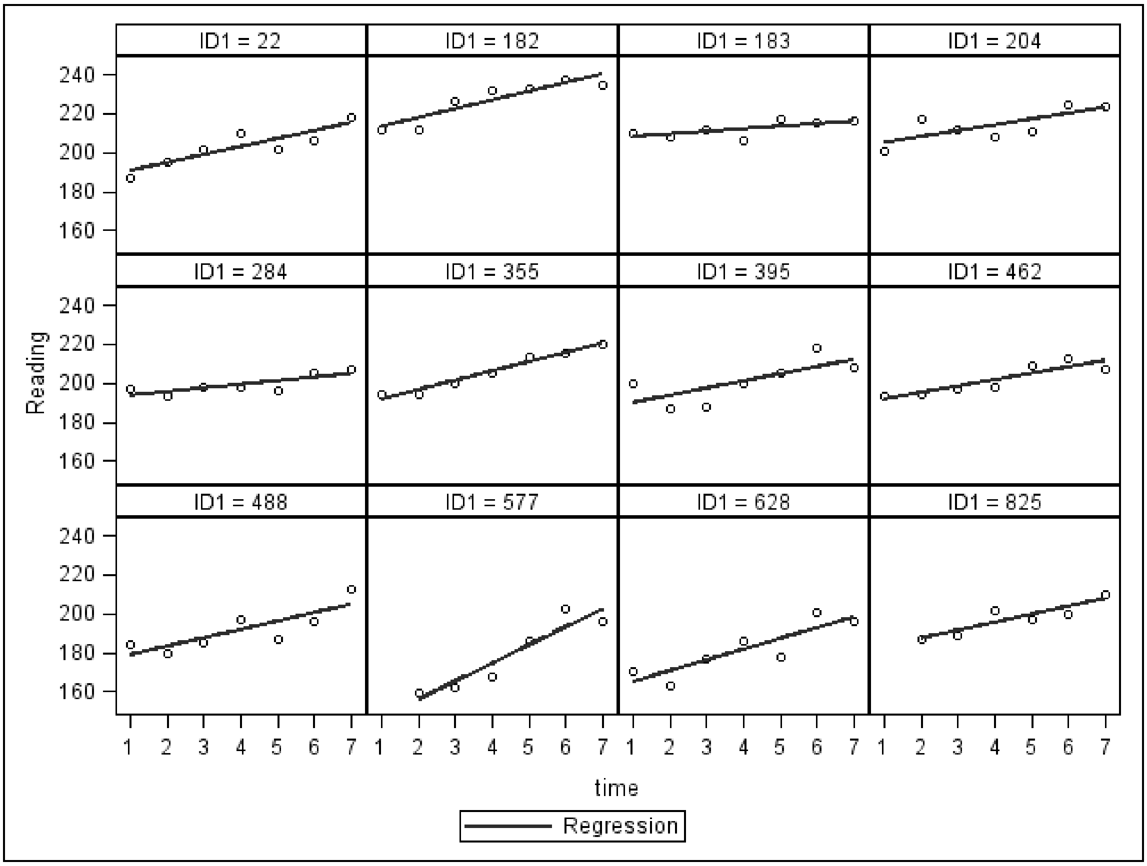

Figure 9. Ordinary Least Squares Summaries of Reading Change Over Time.

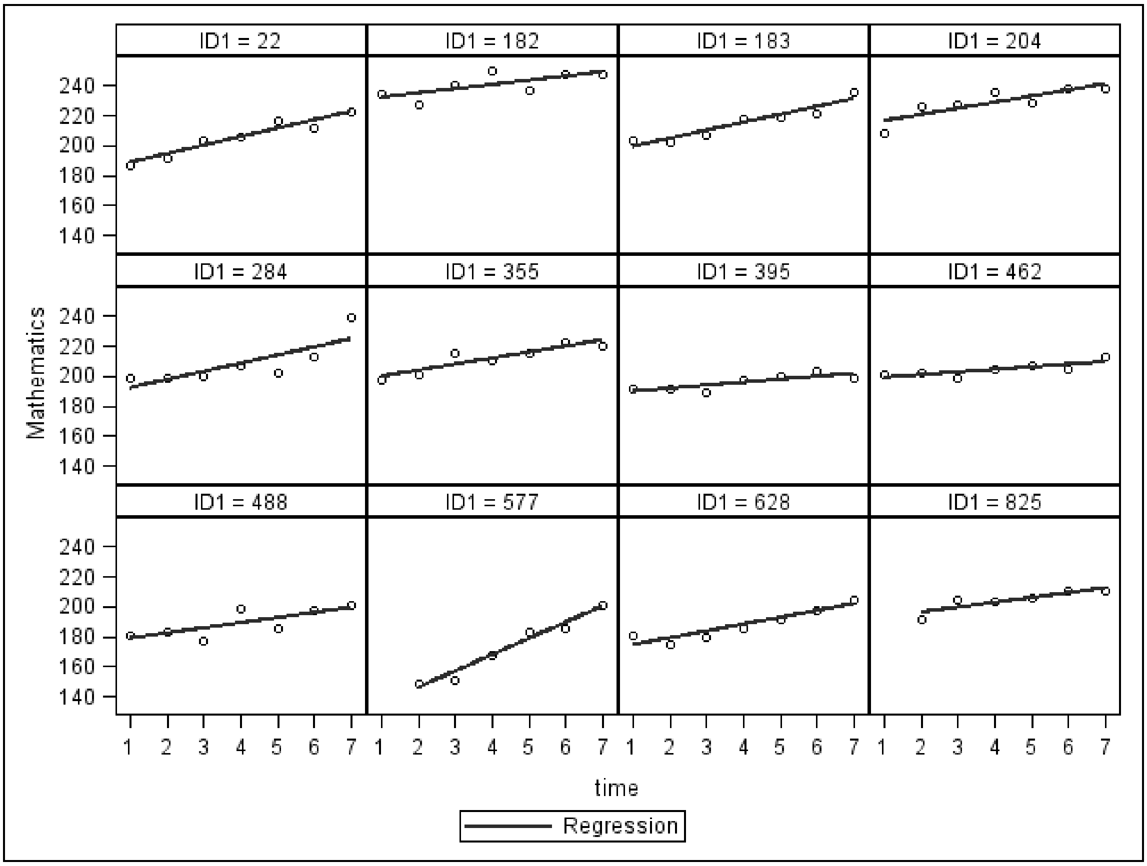

Figure 10. Ordinary Least Squares Summaries of Mathematics Change Over Time.

Descriptive analysis continued with a summary of the non-parametric and parametric trajectories in order to look at the subsample as a whole. The dark line in each panel indicates the average change trajectory for the subsample. Examination of these plots indicated that although change was gradual, it was still positive for both reading and mathematics. However, the range of change was greater for mathematics than for reading. Inspection of the non-parametric growth plots showed that a linear change model was still appropriate. These analyses are depicted in Figures 11 and 12.

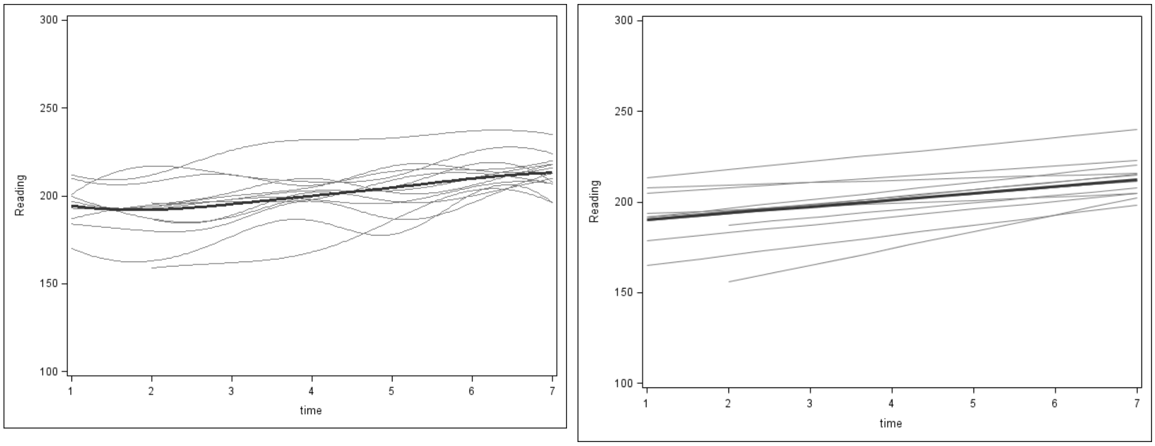

Figure 11. Collection of Smooth Nonparametric and OLS Reading Trajectories Across Participants.

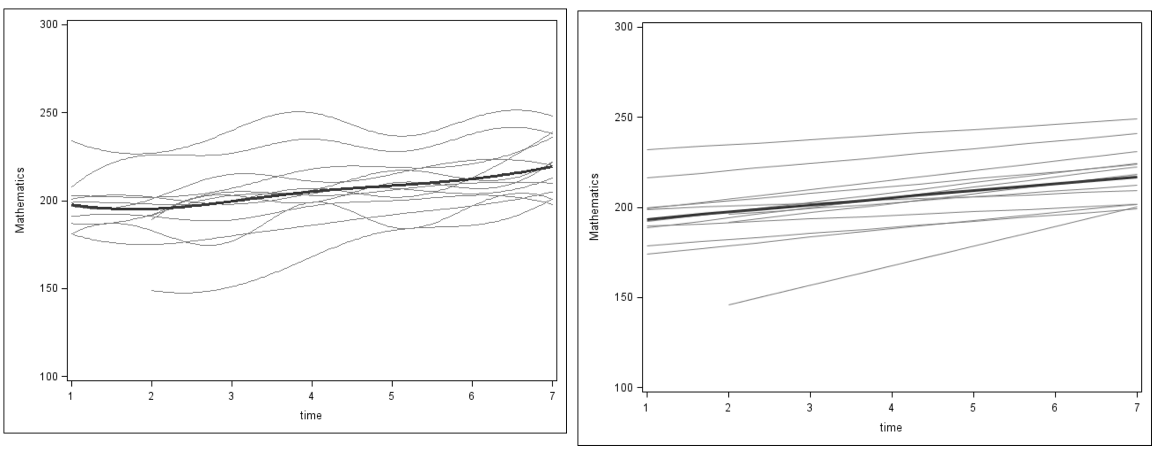

Figure 12. Collection of Smooth Nonparametric and OLS Mathematics Trajectories Across Participants.

Finally, the summary growth plots by ethnicity (represented and underrepresented) and socio-economic status (Low and not-low) were generated to evaluate whether trajectories were affected by these factors. The reading by ethnicity plots (Figure 13, bottom panels) indicated that in general, both groups began the study at the same point, but students from represented groups ended the study at a slightly higher point than their underrepresented peers. The reading by SES plots (Figure 13, top panels) indicated that students from not-low income families began and ended the study at a higher point than their low-income peers. These same trends were evident for the mathematics growth trajectories as well (Figure 14). Further analysis of these results was done in the model fitting process.

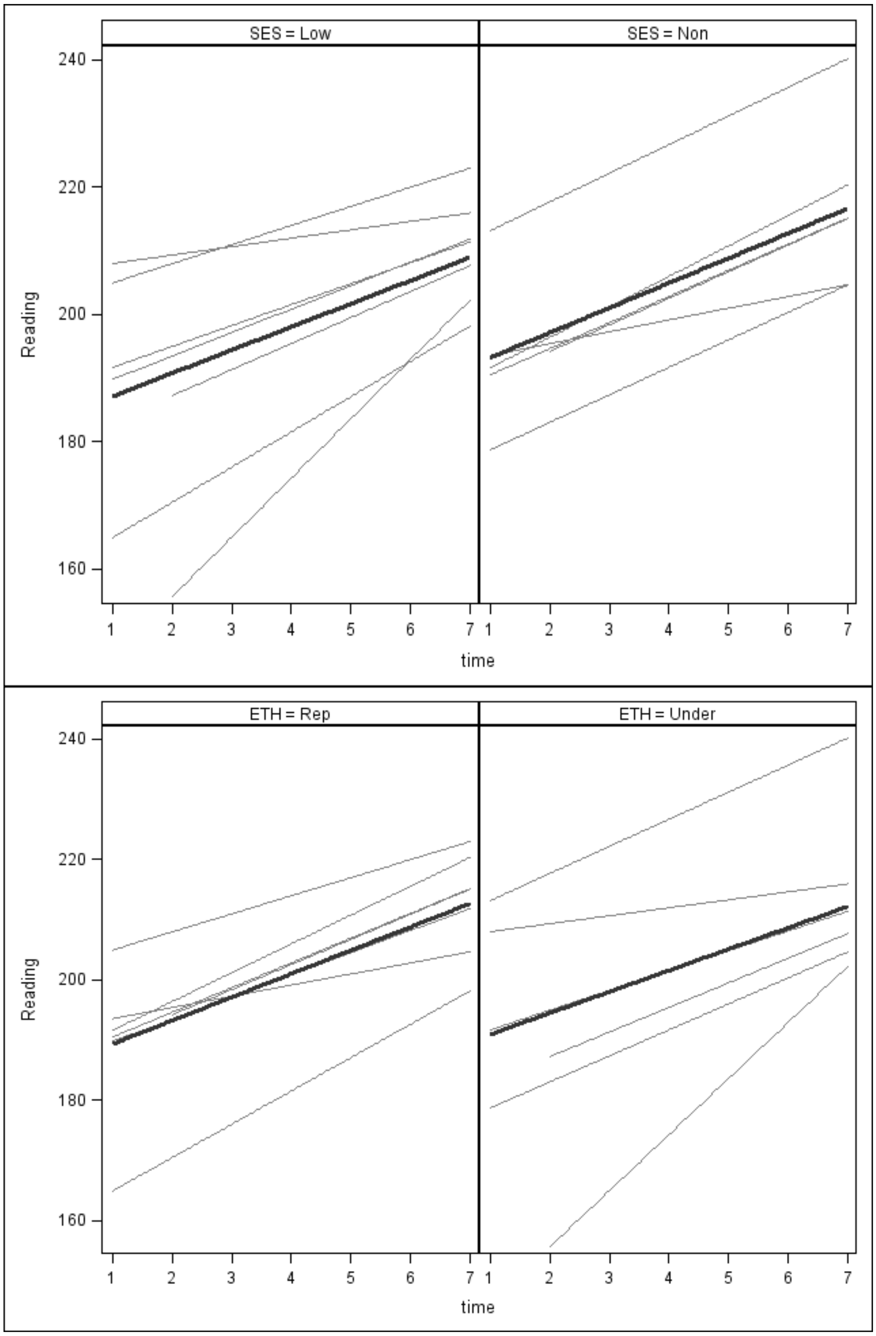

Figure 13. Identification of Potential Predictors of Reading Change by Examining OLS Fitted Trajectories for Ethnicity and Socio-Economic Status.

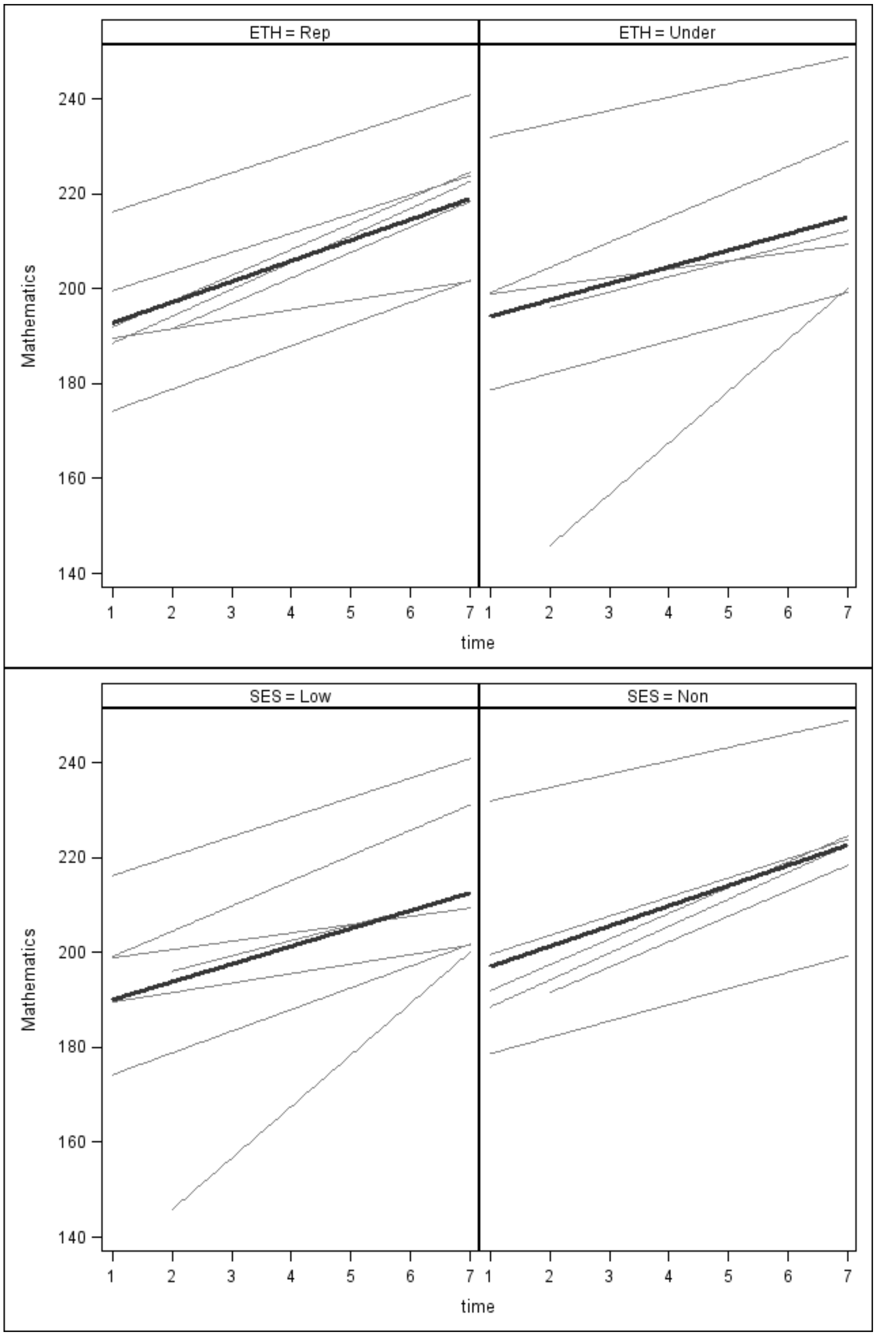

Figure 14. Identification of Potential Predictors of Mathematics Change by Examining OLS Fitted Trajectories for Ethnicity and Socio-Economic Status.

The descriptive analyses discussed above provided summary information about student reading and mathematics achievement growth over the course of the study.

Furthermore, potential predictors of this change were summarized to evaluate their potential influence on growth rates. However, no inferences were made about the form of the trajectories or the influence of the predictors. These analyses are discussed in the proceeding sections.

Model A**.** Model A specification produced parameter estimates for the intercept for both reading and mathematics achievement.

Level 1: $y_{ti} = \pi_{0i} + \varepsilon_{ti}$ Level 2: $\pi_{0i} = \beta_{00} + r_{oi}$

The intraclass correlation (ICC) for reading indicated that 71.9 percent of the variation in reading scores was between-subject variability. For mathematics the ICC indicated a between-subjects variation of 71.45 percent. This showed that there were potential level-two covariates that could explain this variation. These covariates were modeled in subsequent models.

Model B. Model specification in Model B introduced time into the model. Time was centered at the end of the study since potential growth in achievement through the course of the study was the primary focus. Model B, for the total sample indicated that the average reading score for students at the end of the study was 208 points with a gain of 2 points per testing occasion. For mathematics, the average score at the end of the study was 216 with a gain of 4 points per testing occasion. These results were all significant at the $p < .0001$ level indicating that the null hypothesis was rejected that either of these parameters was zero in the population. Examination of the covariance parameter estimates indicated that there was variation in both the intercept and slope that could potentially be explained by person-level covariates. Therefore, further modeling was required.

Model C. Model C was posited to determine if a curvilinear growth was plausible in this context. Therefore, a quadratic term for time was entered into the model. An examination of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) indicated that model fit was better with simple time than with the quadratic term for time. Descriptive results also indicated that growth was linear. Therefore, any further modeling was done using simple time centered at the end of the study (Ctime).

Effects on TSCG on achievement. To answer research question 1a *What effects does TSCG have on student achievement as measured by the NWEA, specifically: (a) What differences exist between treatment and comparison student growth curves among four groups based on income and representation status: (1) Low-SES/underrepresented (SES_ETH1), (2) low-SES/represented (SES_ETH2), (3) not Low-SES/underrepresented (SES_ETH3), and (4) not Low-SES/represented (SES_ETH4)*, model D parameter estimates were examined as well as descriptive statistics. Parameter estimates indicated that, on average, comparison students ended the study with higher achievement scores in both reading and mathematics. However, the comparison slope was marginally less steep than that of the treatment group. Since the intercept was centered at the end of the study mean scores were generated for reading and mathematics achievement for the first testing occasion to evaluate whether the comparison group began the study at a higher achievement level as well. The mean scores for the initial reading and mathematics achievement scores for the treatment school were 198.64 and 202.51 respectively. Mean scores for the comparison school, in reading and mathematics achievement were 200.91 and 205.05 respectively. Therefore, the comparison group also began the study with higher average achievement scores in both content areas. The average reading achievement score for the treatment group at the end of the study was 203.46 and the mathematics score was 209.84. The end-of-study reading and mathematics achievement scores for the comparison school were 206.14 and 213.40 respectively. The treatment effect was also significant.

To evaluate achievement by SES_ETH group, parameter estimates for the intercepts and slopes were generated. Growth plots were also generated for a visual inspection of reading and mathematics achievement. The analyses were further broken down into sections by dataset used. The first analysis was done with the full dataset, including data lines that contained missing data. The second analysis was done using only those data lines that had complete data for all testing occasions.

Full dataset results. Parameter estimate for the time is the main effect in the model. Time is centered at the end of the study. This parameter estimate is significant for both reading and mathematics (R=3.9993, t=5.38, $p<.001$; M=4.3827, $t=6.70$, $p<.001$). This indicates the typical change in score for treatment students per testing occasion is positively affected by the treatment. The main effect for Schooltype indicates the difference in average achievement score at the end of the study between the treatment and comparison students. For reading the parameter estimate is 22.7329 ($p<.001$) indicating that students in the comparison school ended the study with reading achievement scores 22.73 points higher than their treatment peers. Comparison mathematics scores at the end of the study were 28.1647 (p<.001) points higher than those for treatment students. However, the reading and mathematics slope parameter estimates for the comparison school were -.2105 (p=.0913) and -.2776 ($p=.0096$) indicating that, although significant at the $p<.05$ level, these slopes were lower than those for the treatment school. Therefore, it can be said that treatment students increased their achievement scores at a faster rate than their comparison peers. Trajectories are depicted in Figures 15 and 16.

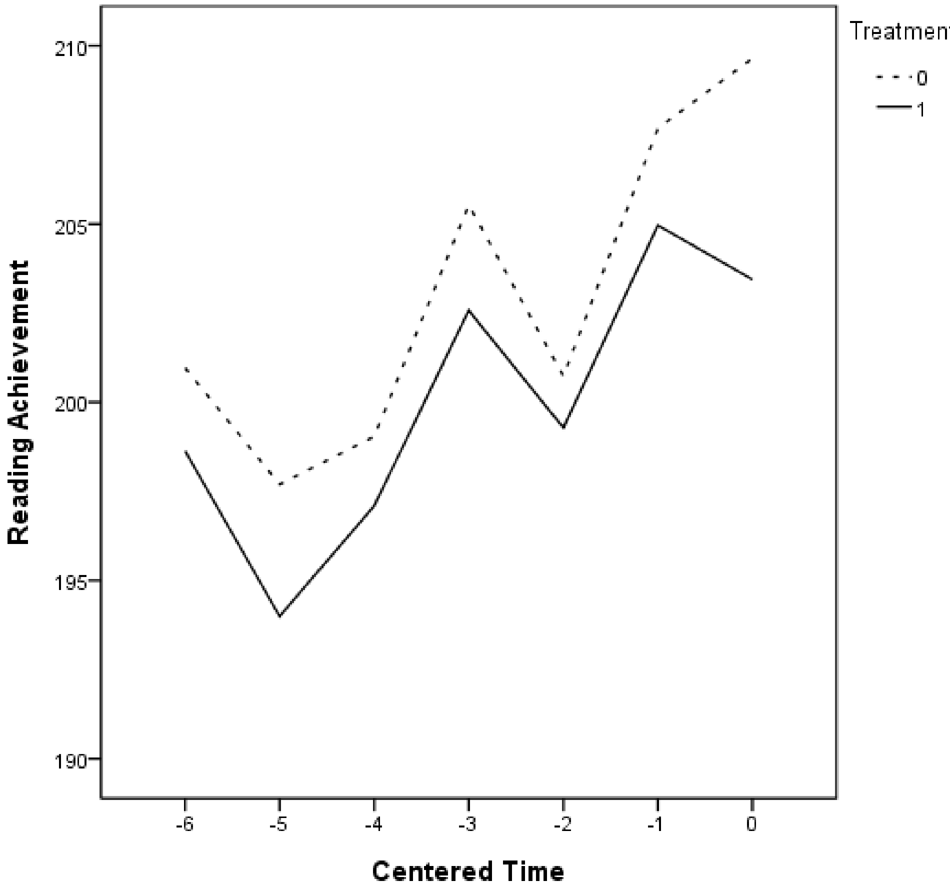

Figure 15. Comparison of Reading Achievement Growth Throughout the Study. Treatment = 1 is the treatment school.

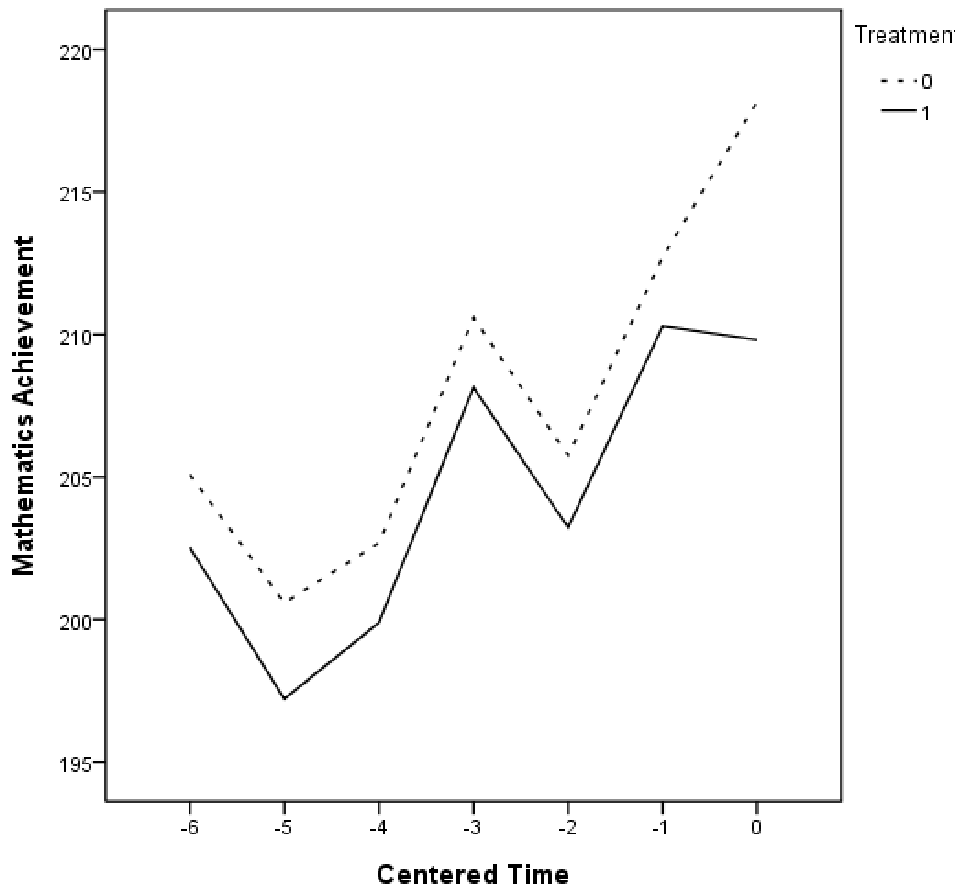

Figure 16. Comparison of Mathematics Achievement Growth Throughout the Study. Treatment = 1 is the treatment school.

Parameter estimates for each SES_ETH group provided evidence that students from low-income, underrepresented (SES_ETH1) families showed lowest achievement levels (R=206.47; t=13.38, p<.0001; M=213.00; t=14.33, p<.0001). The effect of representation and ethnicity was significant for the level of achievement at the end of the study. The slope for students from the low socio-economic/underrepresented group (SES_ETH1) for reading was -.9563 (p=.17) and mathematics was -.6553 (p=.26).

The slope for students from the low socio-economic/represented group (SES_ETH2) for reading was -1.0039 (p=.15) and for mathematics was -.4010 (p=.44).

The slope for students from the not-low socio-economic/underrepresented group (SES_ETH3) for reading was -1.0369 (*p*=.15) and for mathematics was -.6314 (*p*=.28). These estimates indicate that rate of growth in both reading and mathematics achievement was not influenced by a students' representation and ethnicity.

Since, the model used to answer this question is complex, I reduced the model by removing Schooltype from the model to see if the slopes would be significant in a reduced model. This did not occur. At no point in the model fitting process did the slopes for representation and ethnicity become significant. Therefore, it can be said that for each of these four groups, based on representation and ethnicity, rate of growth in achievement in reading and mathematics is not affected. However, the level of achievement at the end of the study was dependent on a child's socio-economic status and ethnicity.

A second interaction between SES_ETH and schooltype (treatment versus comparison) was modeled as well. This interaction was significant for all SES_ETH groups for both reading and mathematics achievement. The negative parameter estimates indicate that the rate of growth in achievement for all groups based on representation and ethnicity were lower for comparison students than their treatment peers. These data are depicted in Table 15 and 16.

Table 15

Model D for Reading Parameter Estimates (Full Dataset)

Predictor		Estimate	Standard Error	T Value
Intercept	Reading	185.5200	1.2284	151.02*
Ctime		3.9993	0.7432	5.38*
Schooltype		22.7329	3.0888	7.36*
SES_ETH1		20.9584	1.5686	13.36
SES_ETH2		22.7849	1.9064	11.95*
SES_ETH3		24.1776	1.5682	15.42*
Ctime*Schooltype		-0.2105	0.1246	-1.69
Ctime*SES_ETH1		-0.9563	0.7553	-1.27
Ctime*SES_ETH2		-1.0039	0.7652	-1.31
Ctime*SES_ETH3		-1.0369	0.7549	-1.37
SES_ETH1*Schooltype		-22.9594	3.4519	-6.65*
SES_ETH2*Schooltype		-20.0581	3.7315	-5.38*
SES_ETH3*Schooltype		-24.6435	3.4971	-7.05*

Note: * indicates significant findings at p<.05.

Table 16

Model D Mathematics Parameter Estimates (Full Dataset)

Predictor		Estimate	Standard Error	t Value
Intercept	Mathematics	186.46	1.4451	129.03*
Ctime		4.3827	0.6543	6.70*
Schooltype		28.1647	3.6508	7.71*
SES_ETH1		26.6106	1.8536	14.36*
SES_ETH2		28.8007	2.2628	12.73*
SES_ETH3		31.0076	1.8548	16.72*
Ctime*Schooltype		-0.2776	0.1070	-2.60
Ctime*SES_ETH1		-0.6553	0.6646	-0.99
Ctime*SES_ETH2		-0.4010	0.6727	-0.60
Ctime*SES_ETH3		-0.6314	0.6641	-0.95
SES_ETH1*Schooltype		-27.0694	4.0752	-6.64*
SES_ETH2*Schooltype		-23.7578	4.4231	-5.37*
SES_ETH3*Schooltype		-31.2816	4.1411	-7.55*

Note: * indicates significant findings at $p<.05$.

Downward slopes in the trajectories occurred at the fall testing occasions indicating that summer loss was an issue for students regardless of school placement. A spline model might have alleviated this drop somewhat and reduced the standard errors, but given that the errors were already small with the use of this model, and the fact that there is little difference in trajectory, a spline model did not dramatically change results. Therefore, a linear model was retained.

Complete case results. In comparison to the results using the full dataset, results using only the complete cases were similar. No slope parameter estimates for SES_ETH were statistically significant and the main effects of treatment and time remained significant. In other words, socio-economic status and ethnicity had the same effect on achievement. Students from lower-income families and underrepresented groups did not perform as well on reading and mathematics achievement tests as their counterparts from not-low income, represented families.

Table 17

Model D Reading Parameter Estimates for Complete Case Dataset

Predictor		Estimate	Standard Error	t Value
Intercept	Reading	185.52	1.2284	151.03*
Ctime		3.9992	0.7428	5.38*
Schooltype		22.7340	3.0888	7.36*
SES_ETH1		20.9344	1.5689	13.34*
SES_ETH2		22.7801	1.9064	11.95*
SES_ETH3		24.1779	1.5684	15.42*
Ctime*Schooltype		-0.2030	0.1246	-1.63
Ctime*SES_ETH1		-0.9567	0.7550	-1.27
Ctime*SES_ETH2		-0.9884	0.7648	-1.29

Ctime*SES_ETH3	-1.0395	0.7546	-1.38
SES_ETH1*Schooltype	-22.9451	3.4520	-6.65*
SES_ETH2*Schooltype	-19.9103	3.7323	-5.33*
SES_ETH3*Schooltype	-24.6261	3.4972	-7.04*

Note: * indicates significant findings at p<.05..

Table 18

Model D Mathematics Parameter Estimates for Complete Case Dataset

Predictor		Estimate	Standard Error	t Value
Intercept	Mathematics	186.46	1.4455	128.99*
Ctime		4.3823	0.6548	6.69*
Schooltype		28.1644	3.6517	7.71*
SES_ETH1		26.6101	1.8541	14.35*
SES_ETH2		28.8009	2.2634	12.72*
SES_ETH3		31.0091	1.8554	16.71*
Ctime*Schooltype		-0.2791	0.1071	-2.60
Ctime*SES_ETH1		-0.6521	0.6651	-0.98
Ctime*SES_ETH2		-0.3999	0.6733	-0.59
Ctime*SES_ETH3		-0.6295	0.6646	-0.95
SES_ETH1*Schooltype		-27.0721	4.0762	-6.64*

SES_ETH2*Schooltype	-23.7590	4.4242	-5.37*
SES_ETH3*Schooltype	-31.2836	4.1421	-7.55*

Note: * indicates significant findings at p<.05..

Analysis of the parameter estimates for each SES_ETH group provided interesting information. Students in SES_ETH2 and SES_ETH4 groups performed better than those in SES_ETH1 and SES_ETH3 groups. SES_ETH2 and SES_ETH4 are the groups that included students from represented groups. Therefore, it seems that socio-economics was less of a predictor of lower achievement in this sample than representation status. Analysis of correlations between SES_ETH and reading and mathematics scores also provided evidence for this trend. SES_ETH1 (R=-.128, p<.0001; M=-.110, p<.001) SES_ETH2 and SES_ETH3 (R=-.040, p=.001; M=-.031, p=.01) were negatively correlated with reading and mathematics achievement and SES_ETH4 (R=.186, p<.0001; M=.174, p<.001) was positively correlated. SES_ETH2 correlations were not significant.

Students in the treatment school outperformed their comparison peers in achievement growth by a marginal difference. When socio-economic status and ethnicity was controlled for, students from represented groups performed better in both reading and mathematics than their underrepresented peers regardless of socioeconomic status. Overall, students in the treatment school from represented, not low-socioeconomic status outperformed all other groups in both reading and mathematics.

Achievement of treatment students. To answer question 1b, *Do treatment school students in a TSCG program out-perform matched comparison school students* *after three years in the program after controlling for SES and ethnic representation?* Model D parameter estimates were also consulted. Covariance estimates changed from the model with only time in it (Model B) and the model including the covariates (Model D). When it came to estimating the achievement status of students at the end of the study, inclusion of the covariates reduced the size of the variance component for the intercept from 190.81 to 135.55. The difference in intraclass correlation indicates that 28.96 percent of the difference in level of reading achievement at the end of the study was explained by Schooltype and SES_ETH. However, again the covariance estimate for slope changed only slightly and accounted for only 1.6 percent of the explainable variance in rate of change. Therefore, it can be said that although treatment students ended the study at a lower achievement level in reading they made greater gains than their comparison counterparts by a small margin.

Analysis of mathematics scores after controlling for Schooltype and SES and ethnic representation indicated a slightly different result. Both covariance estimates for intercept and slope increased when these covariates were added to the model. Covariance estimates for achievement levels at the end of the study changed from 190.81 to 210.08 and for rate of change from 1.2602 to 1.2654. Mathematics achievement did not seem to change due to placement in treatment or comparison after SES_ETH was accounted for.

Students in the treatment school did make greater gains than their comparison peers in reading by a small margin in reading achievement. However, treatment students did not show greater gains in mathematics after controlling for Schooltype, socioeconomic status and ethnicity.

Differences in achievement growth by condition. To answer research question 1c, *How do the learning trajectories differ between treatment and comparison students?* parameter estimates and plots from Model D were examined. These data are depicted in Tables 17 and 18. Overall achievement in both reading and mathematics increased during the course of the study. For the full dataset, the rate of change for students in the treatment school for reading was 3.98 (t=5.36, p<.0001) points per testing occasion and for mathematics was 4.36 (t=6.66, p<.0001). The growth rate for students from the comparison school for reading was 3.77 (t=1.69, p=.0913) and for mathematics was 4.08 (t=2.60, p=.0096). Results for the complete case dataset showed slopes for reading for the treatment school as 4.00 (t=5.38, p<.0001) and 4.38 (t=6.69, p<.0001) for mathematics. Rate of change for reading for the comparison school was 3.80 (t=1.63, p=.1036) and for mathematics was 4.10 (t=2.60, p=.0093). This indicated an increased growth for students in the treatment school compared to their comparison counterparts. Comparison students showed significant growth in mathematics achievement during the course of the study, but showed negligible growth in reading achievement.

Differences in achievement growth based on initial identification. To answer research question 1d, *For treatment school students only, based on initial identification categories, how do student achievement growth trajectories differ over three program years?* Model E parameter estimates were examined. These data are depicted in Tables 20 and 21. When the full model was run, parameter estimates for time (R=1.7133, t=5.01, p<.0001; M=3.3348, t=10.33, p<.0001), and Ident1 (Intercept based on initial cluster identification) for the special-education (R=205.98, t=2.79, p=.006; M=210.64, t=3.39, p=.0008), low-achieving (R=205.77, t=4.20, p<.0001; M=211.67, t=4.72, p<.0001), low-average (R=207.80, *t*=3.59, *p*=.0004; M=215.25, *t*=3.84, *p*<.0001), average (R=213.38, *t*=2.96, *p*=.0033; M=223.75, *t*=3.00, *p*=.0029), and high-achieving clusters (R=221.76, *t*=88.61, *p*<.0001; M=234.73, *t*=71.74, *p*<.0001) were significant. Reading achievement slopes for special education (R=2.44, *t*=4.43, *p*<.0001), low-achieving (R=1.26, *t*=2.97, *p*=.0032, low-average (R=0.7859, *t*=1.98, *p*=.048), and high-achieving (R=1.71, *t*=5.01. *p*<.0001) clusters were also significant as shown by the interaction effect of centered time and initial identification (Ctime*Ident1). All other estimates were non-significant. I reduced the model to just including intercept and slope parameters for SES_ETH and Ident1. This did not result in other parameter estimates becoming significant so the full model was fit again. Refitting the full model did not change covariance estimates or fit statistics a great deal, therefore, the full model is useful in that the data provide information about non-significance of the treatment for certain groups.

Students originally identified for the special-education cluster ended the study with an average reading achievement score of 205.98 (*p*=.005) and mathematics achievement score of 210.64 (*p*=.0008). Students originally identified for the low-achieving cluster ended the study with an average reading achievement score of 205.77 (*p*<.0001) and mathematics achievement score of 211.67 (*p*<.0001). Students originally identified for the low-average cluster ended the study with an average reading score of 207.80 (*p*=.0004) and mathematics score of 215.25 (*p*=.0001) . Students originally identified for the average cluster ended the study with an average reading score of 213.38 (*p*=.003) and a mathematics score of 223.75 (*p*=.0029). Those individuals originally identified for the above-average cluster ended the study with an average reading score of

221.58 (p=.6593) and mathematics score of 233.08 (p=.9509), a negligible difference. Students originally identified as High Achieving ended the study with reading and mathematics achievement scores only slightly higher than those students identified originally as Above Average. Their scores were 221.76 (p<.0001) for reading and 234.73 (p<.0001) for mathematics. These results are summarized in Table 19.

Table 19

Achievement Subtest Scores at End of Study by Original Cluster Identification

Cluster	Reading Achievement Score	Standard Error	Mathematics Achievement Score	Standard Error
High Achieving	221.76	2.50	234.73	3.27
Above average	221.58	2.87	233.08	3.74
Average	213.38	2.83	223.75	3.66
Low Average	207.80	3.88	215.25	5.07
Low Achieving	205.77	3.81	211.67	4.88
Special Education	205.98	2.87	210.64	7.11

Since the intercepts and slopes were significant for special-education, low-achieving, low-average, and high-achieving clusters, at the p<.05 level, it can be said that initial identification for these students had an effect on the rate of growth in reading achievement and level of reading achievement at the end of the study. Intercepts for special-education, low-achieving, low-average, average, and high-achieving clusters for mathematics were significant, but the growth rate parameters were not. Therefore it can be said there is a relationship between initial cluster identification and level of mathematics achievement at the end of the study for these students.

Students originally identified as lower-achieving tended to make greater gains in reading achievement while mathematics gains were not as great. Students identified as higher-achieving did not make the same gains as their lower-achieving peers. However, examination of Figures 15 and 16 indicate that treatment students' achievement dipped at the final testing occasion for both reading and mathematics. This may have been due to changes made in the identification process or due the changes in the school boundaries that occurred for the treatment school during the study.

Treatment student achievement growth by income and representation. To answer research question 1e, *For treatment school students only, using four groups of students based on income and representation status how do student achievement growth trajectories differ over three program years?*

None of the intercept or slope parameter estimates for SES_ETH groups were significant at the $p < .05$ level, indicating that there was no relationship between income and representation and achievement in the treatment school. This is true for both reading and mathematics. The parameter estimates are depicted in Tables 20 and 21.

Table 20

Model E Reading Parameter Estimates

Predictor		Estimate	Standard Error	t Value
Intercept	Reading	221.76	2.5027	88.61*
Ctime		1.7133	0.3417	5.01*
Ident1	0	-15.7785	5.6458	-2.79*
Ident1	1	-15.9945	3.8079	-4.20*
Ident1	2	-13.9590	3.8842	-3.59*
Ident1	3	-8.3813	2.8297	-2.96*
Ident1	4	-0.1771	2.8709	-0.06
Ctime*Ident1	0	2.4421	0.5518	4.43*
Ctime*Ident1	1	1.2571	0.4231	2.97*
Ctime*Ident1	2	0.7859	0.3968	1.98*
Ctime*Ident1	3	0.3900	0.3677	1.06
Ctime*Ident1	4	0.4274	0.3777	0.26
SES_ETH1		2.1842	6.1322	0.36
SES_ETH2		8.7778	5.0582	1.74
SES_ETH3		8.2993	6.4785	1.28

Note. * indicates significant findings at $p<.05$..

Table 21

Model E Mathematics Parameter Estimates

Predictor		Estimate	Standard Error	t Value
Intercept	Mathematics	234.73	3.2719	71.74*
Ctime	Mathematics	3.3348	0.3228	10.33*
Ident1	0	-24.0924	7.1118	-3.39*
Ident1	1	-23.0631	4.8824	-4.72*
Ident1	2	-19.4771	5.0742	-3.84*
Ident1	3	-10.9843	3.3665	-3.00*
Ident1	4	-1.6492	3.7373	0.44
Ctime*Ident1	0	-0.4090	0.5283	-0.77
Ctime*Ident1	1	0.3874	0.4012	0.97
Ctime*Ident1	2	-0.1866	0.3754	-0.50
Ctime*Ident1	3	0.2701	0.3476	0.78
Ctime*Ident1	4	0.5190	0.3568	1.45
SES_ETH1		-7.1429	7.7755	-0.92
SES_ETH2		9.6681	6.3899	1.51
SES_ETH3		14.4282	7.8734	1.83

Note. * indicates significant findings at $p<.05$..

Research Question Two

Overall effects of total school cluster grouping model on identification categories. Descriptive data and pre/post Chi-square analyses for goodness of fit were used to address research question 2: *What effects does TSCG have on identification categories?* Changes in frequency of students identified in each of the six achievement groups for three identification periods were analyzed. This analysis was performed for the treatment school as a whole, by graduation year, and then by income and ethnicity. Three waves of identification data were present in the dataset. Identification occurred in the spring prior to the beginning of the study and at the end of each of the next two treatment years.

This study utilized graduation years as cohorts of students as they advanced from second through sixth grade. Therefore, not all students were present for all three identification periods. Students from graduation years 2015, 2016, and 2017 were present for all identification periods. Students from graduation year 2014 were present for two identification waves. This represents all graduation years for which data were gathered. Table 22 summarizes the treatment school sample and the frequency of students available for different identification periods for the total treatment school (grades two through six) by graduation year. Table 23 summarizes the treatment sample and the frequency of students available for different identification periods by income group, and by ethnic and cultural groups.

Table 22

Identification Frequencies by For Treatment School by Graduation Year

Group	3 ID Points	2 ID Points
Treatment	256	386
2017	75	84
2016	102	122
2015	79	89
2014	N/A	91

Table 23

Identification Frequencies for Treatment School by Socioeconomic Status and Ethnicity

Group	3 ID Points	2 ID Points
Treatment	256	386
Low-Income	140	215
Not-Low-Income	116	171
Underrepresented	84	137
Represented	172	249

Trends in identification categories were noted for students who were present in the study for different lengths of time. For students present for three identification periods the following trends were noted. From identification period 1 (Spring 2006) to identification period 2 (Spring 2007) changes in identification categories were noted with the number of students identified as Low Achieving, Above Average, and High Achieving increasing. Students identified as Special Education increased by one. The number of students identified as Low Average and Average decreased in number. From identification period 2 (Spring 2007) to identification period 3 (Spring 2008) changes in identification categories were also noted with the number of students identified as Special Education, Low Achieving, Average, and High Achieving increasing. The number of students identified as Above Average decreased, as did the number of students identified as Low Average, but only by one. These changes are depicted in Figure 17.

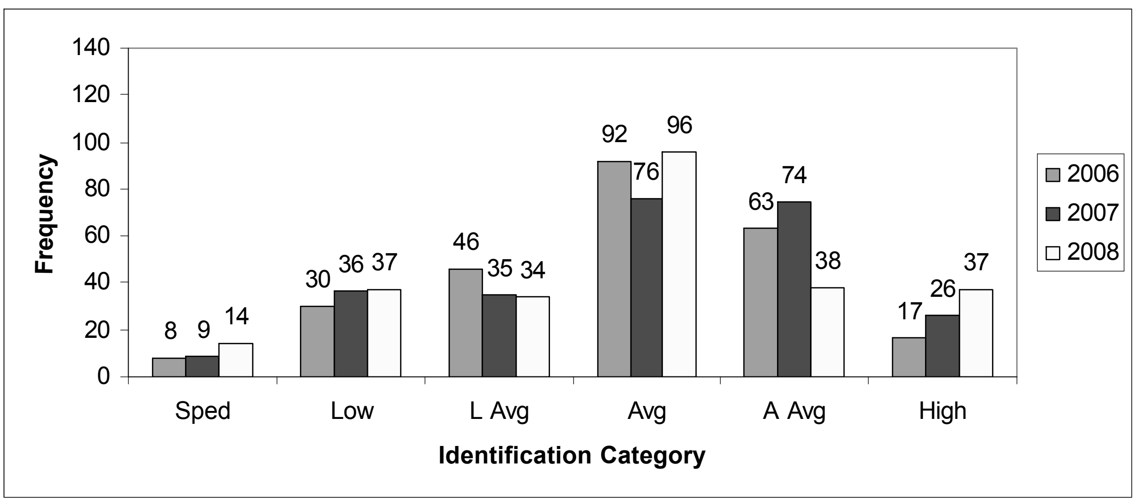

Figure 17. Treatment School Identification Categories for 3 Identification Periods.

For students present for two identification periods, similar trends were noted. From identification period 1 (Spring 2006) to identification period 2 (Spring 2007) the number of students identified as Special Education, Above Average, and High Achieving increased while the number of students identified as Low Achieving, Low Average, and Average decreased. These trends are depicted in Figure 18.

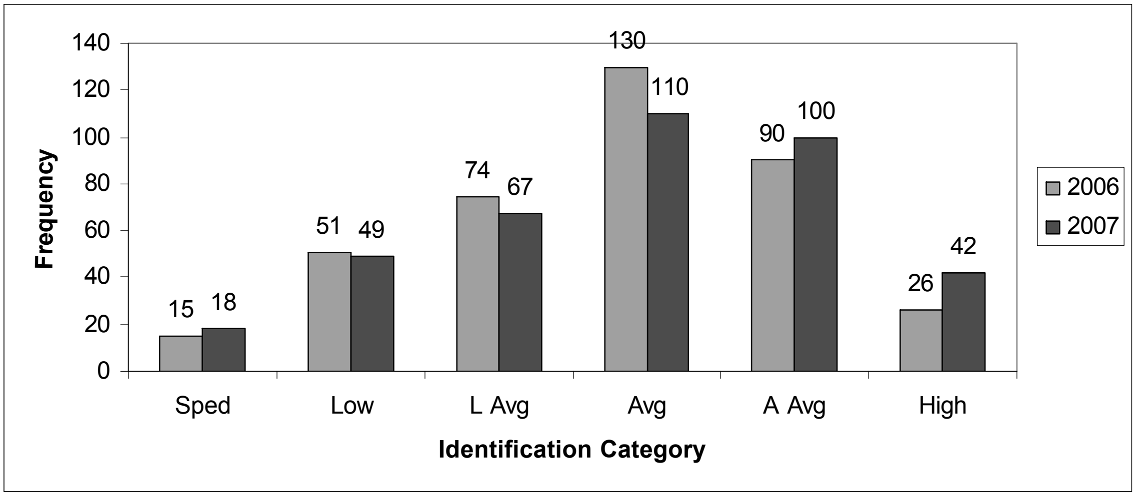

Figure 18. Treatment School Identification Categories for 2 Identification Periods.

Effects on identification categories by graduation year. ***Effects on identification categories for graduation year 2017.*** Students in graduation year 2017 began the identification process in second grade and were potentially present for all three identification periods. Trends for students present for three identification periods were as follows: From identification period 1 (Spring 2006) to identification period 2 (Spring 2007) changes in identification categories were somewhat different from those of the treatment school overall with the number of students identified as Special Education remaining the same, the number of students identified as Low Achieving, Low Average, Above Average, and High Achieving increasing, and the number of students identified as Average decreasing. From identification period 2 (Spring 2007) to identification period 3 (Spring 2008) more changes were noted with the number of students identified as Low Achieving, Low Average, and Average increasing, and the number of students identified as Special Education, Above Average, and High Achieving decreasing. Figures 19 and 20 depict these trends.

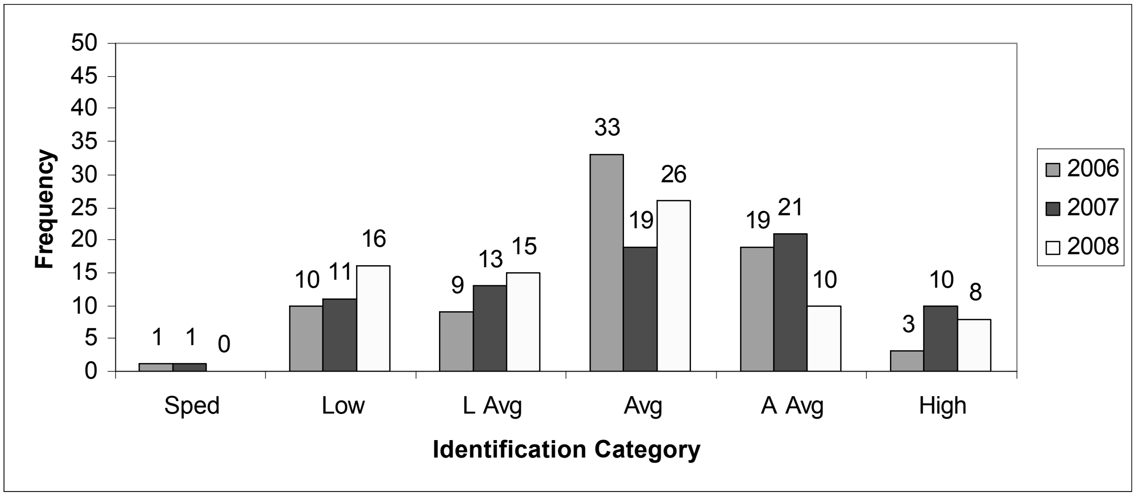

Figure 19. Identification categories for Students in Graduation Year 2017 With 3 Identification Points.

For students who were present for two identification periods, from identification period 1 (Spring 2006) to identification period 2 (Spring 2007) changes in identification categories were seen with the number of students identified as Low Achieving, Low Average, and High Achieving increasing and the number of students identified as Average decreasing. No changes were noted for the number of students identified as Special Education and Above Average.

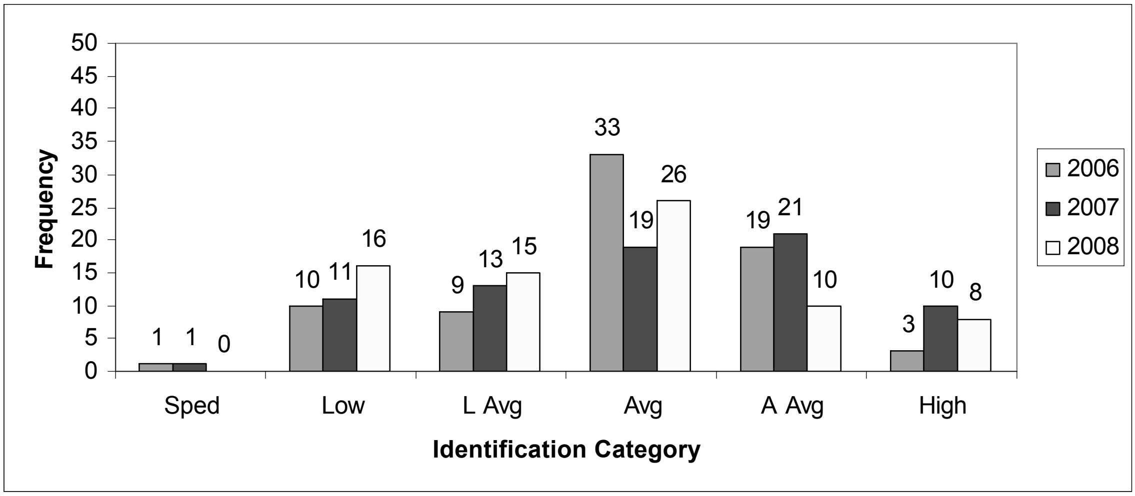

Figure 20. Identification categories for Students in Graduation Year 2017 With 2 Identification Points.

Effects on identification categories for graduation year 2016. Students in graduation year 2016 began the identification process in third grade. For those students present for three identification periods, from identification period 1 (Spring 2006) to identification period 2 (Spring 2007) changes in identification categories were seen for all categories, except High Achieving, which remained the same. Changes were noted with the number of students identified as Special Education, Low Achieving, and Above Average increasing and the number of students identified as Low Average, and Average decreasing. From identification period 2 (Spring 2007) to identification period 3 (Spring 2008) changes were noted with the number of students identified as Average and High Achieving increasing and the number of students identified as Low Achieving and Above Average decreasing. No changes were seen for students identified as Special Education and Low Average. These changes are depicted in Figure 21.

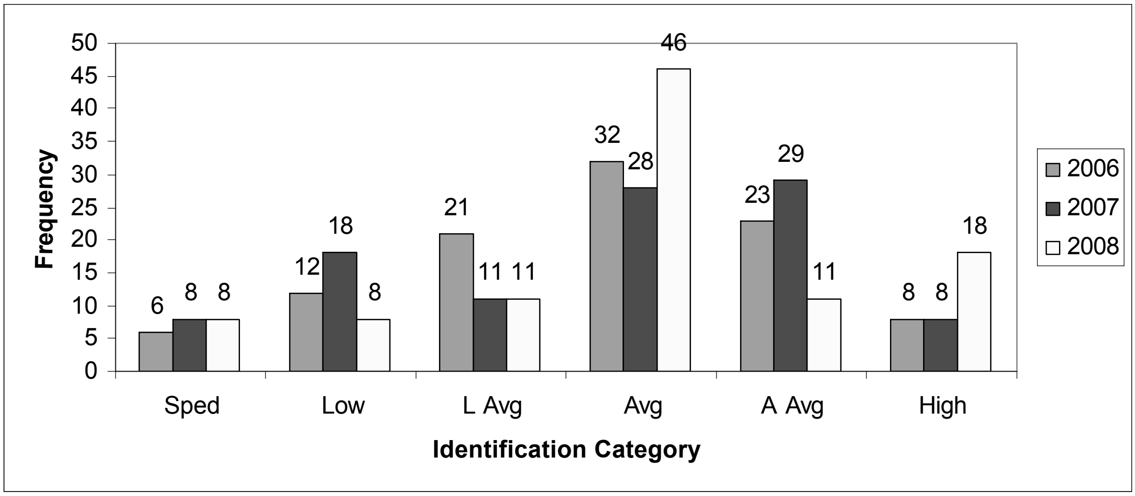

Figure 21. Identification Categories for Students in Graduation Year 2016 with 3 Identification Points.

For students present for two identification periods, changes were noted in all categories with the number of students identified as Special Education, Low Achieving, and Above Average increasing and the number of students identified as Low Average, Average, and High Achieving decreasing. These changes are depicted in Figure 22.

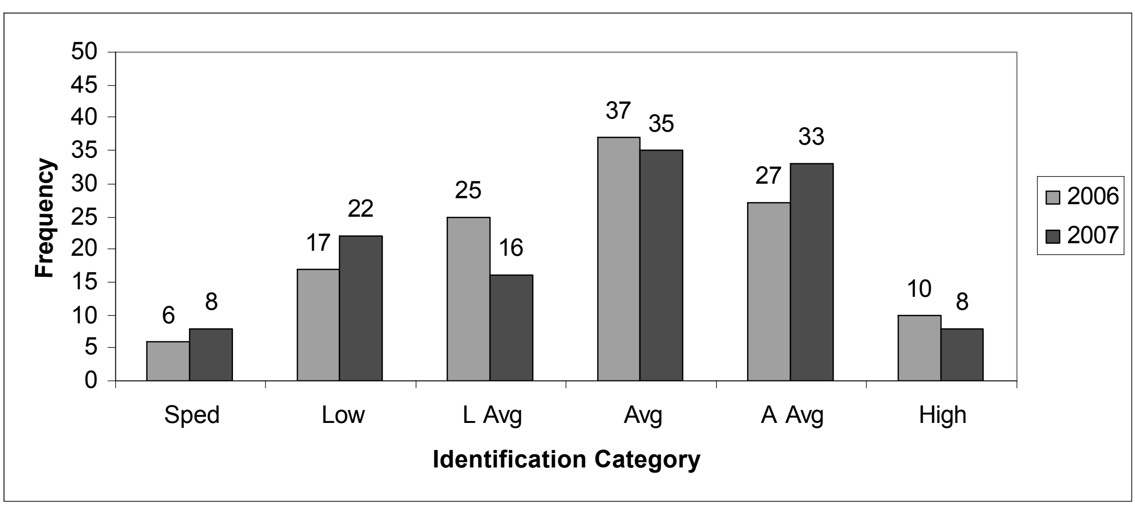

Figure 22. Identification Categories for Students in Graduation Year 2016 with 2 Identification Points.

Effects on identification categories for graduation year 2015. A maximum of three identification points existed for graduation year 2015. These students began the identification process in grade four. From identification period 1 (Spring 2006) to identification period 2 (Spring 2007), changes in identification categories were seen across all categories with the number of students identified as Average, Above Average, and High Achieving increasing and the number of students identified as Special Education, Low Achieving, and Low Average decreasing. From identification period 2 (Spring 2007) to identification period 3 (Spring 2008) changes were also seen in identification categories with the number of students identified as Special Education, Low Achieving, and High Achieving increasing and the number of students identified as Low Average, Average, and Above Average decreasing. These changes are depicted in Figure 23.

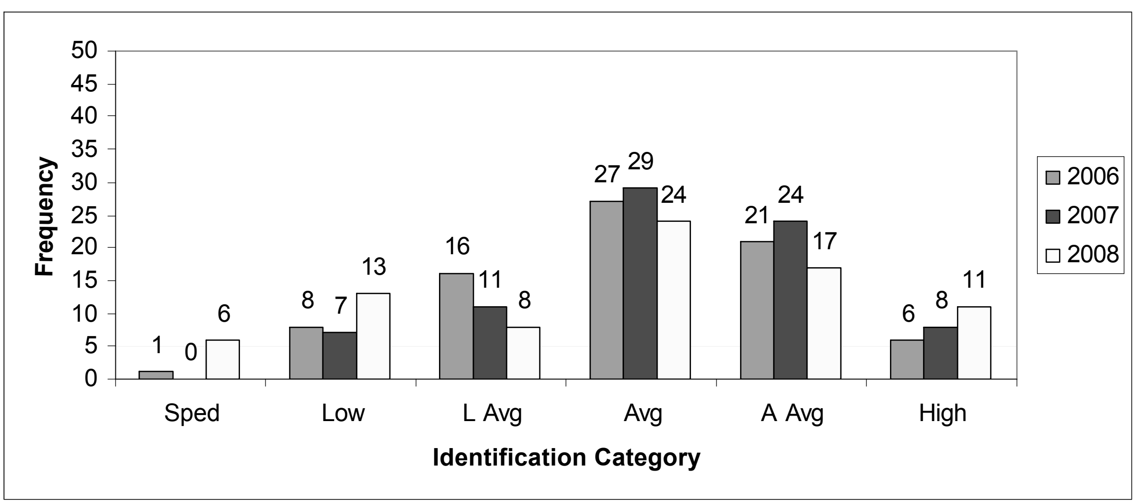

Figure 23. Identification Categories for Students in Graduation Year 2015 With 3 Identification Points.

For students present for two identification periods changes were seen in all categories except for students identified as Average. The number of students identified as Low Achieving, Above Average, and High Achieving increased while the number of students identified as Low Average and Special Education decreased.

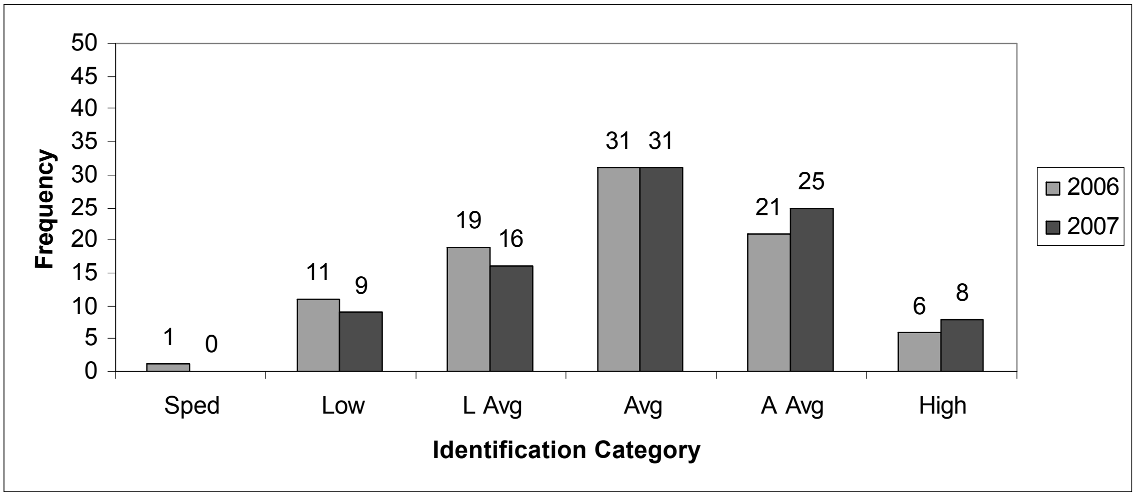

Figure 24. Identification Categories for Students in Graduation Year 2015 With 2 Identification Points.

Effects on identification categories for graduation year 2014. A maximum of two identification points were available for graduation year 2014. Students in this graduation year entered seventh grade in school year 3 and were therefore not available for identification. From identification period 1 (Spring 2006) to identification period 2 (Spring 2007) changes in identification were noted with the number of students identified as Special Education, Low Average, and High Achieving increasing and the number of students identified as Low Achieving and Average decreasing. No changes were seen for students identified as Above Average. These trends are depicted in Figure 25.

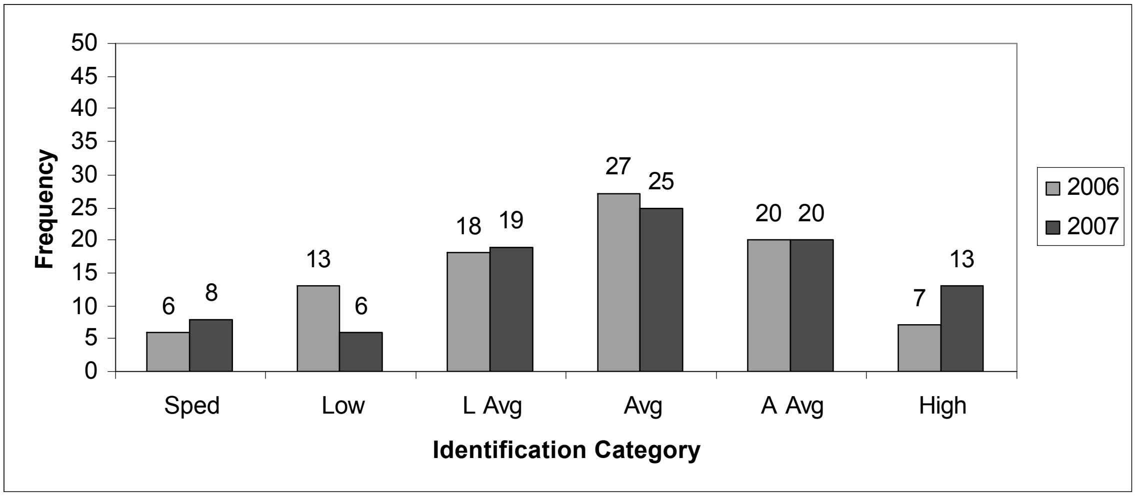

Figure 25. Identification Categories for Students in Graduation Year 2014 with 2 Identification Points.

Overall for the treatment school (graduation years 2017 through 2014), more students were identified as high achieving for the first three identification periods (the first two program years), while fewer were identified as low achieving. Students only present for two identification periods were less likely to be identified as High Achieving.

Effects on identification by income status. *What changes occur in frequency of students identified in each of the five achievement by income?*

Three identification points. Data from students present for three identification periods were analyzed for proportional representation in clusters by income status. Figure 26 indicates that non-proportionality was an issue in the special-education, low-achieving, and high-achieving clusters. Students from low-income families were overrepresented in the lower clusters and underrepresented in the upper clusters. The final year of the study showed some improvement for these students as the number identified as Above Average increased.

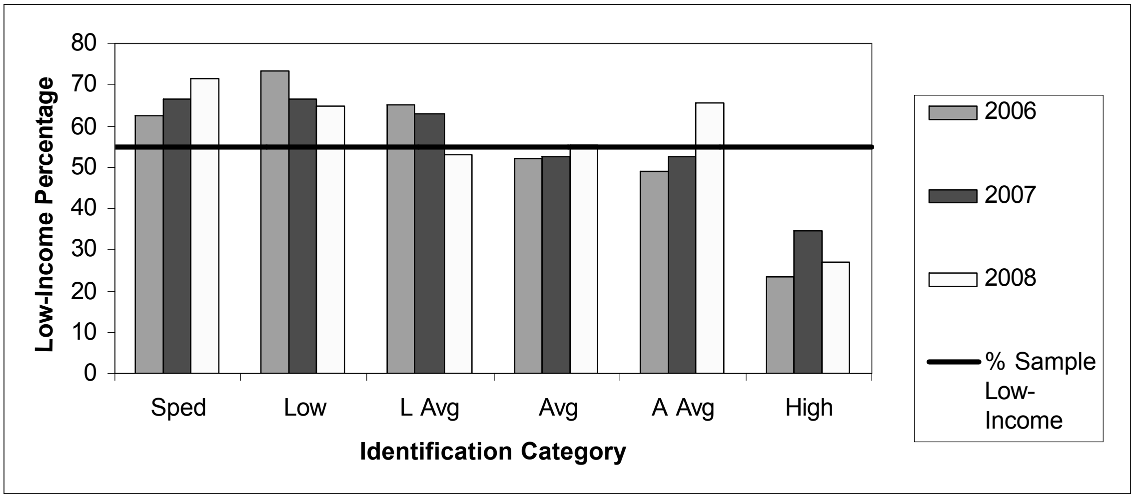

Figure 26. Low-Income Students Percentages by Categories Compared to Total Students with Three Identification Points.

Chi-Square analyses for identification period 1 (Spring 2006) and identification period 3 (Spring 2008) were conducted to see if non-proportionality issues were statistically significant. Chi-square for identification period 1 (Spring 2006) was 14.12 at the p=.15 significance level indicating a non-significant finding. Chi-square for identification period 3 (Spring 2008) was 16.50 at the p<.05 level indicating that non-proportionality increased over the three years this group was in the study. Standardized residuals were generated for these analyses to examine where the disparity occurred. At the end of this study students from low-income families were underrepresented in the high-achieving cluster while students from not-low-income families were overrepresented. This indicates that students from low-income families were less likely to be identified as high-achieving than their not-low-income peers even when present for three identification periods. This disparity at the end of the study may be due to the change in identification procedures that occurred at the end of year three. These results are depicted in Tables 24 and 25.

Table 24

Chi-Square Results for Students with Three Identification Points

	2006	2008
X^2	14.12	16.50
Df	5	5
P	0.15	<0.05

Table 25

Standardized Residuals for Students with Three Identification Points

	2006 Low	2006 Not Low	2008 Low	2008 Not Low
Sped	0.3	-0.3	0.8	-0.9
Low	1.4	-1.5	0.8	-0.9
L Avg	1.0	-1.1	-0.1	0.2
Avg	-0.3	0.4	0.1	-0.1
A Avg	-0.6	0.6	0.9	-1.0
High	-1.7	1.9	-2.3	2.5

Note. Important contributors to the overall effect are highlighted.

Two identification points. For students who were present for two identification periods, from identification period 1 (Spring 2006) to identification period 2 (Spring

2007) the number of students from low-income families identified as Above Average and High Achieving increased. Chi-Square analysis (Year 1 X^2=28.28, K=5, p<.0001 and Year 2 X^2=17.68, K=5, p<.05) revealed issues of non-proportionality across identification groups – both above average and below average.

Table 26

Chi-Square Results for Students with Two Identification Points

	2006	2007
X^2	28.28	17.68
Df	5	5
P<	0.0001	0.05

At the beginning of the study, students from low-income families were underrepresented in the high-achieving cluster. Students from not-low-income families were underrepresented in the low-achieving cluster and overrepresented in the high-achieving cluster. By the end of the study proportionality was no longer an issue in the high-achieving cluster, but not-low-income students continued to be underrepresented in the low-achieving and low-average clusters. It is not clear whether this is due to the over-identification of low-income students in the lower-achieving clusters or is not-low-income students do not experience the same learning challenges as their low-income peers. More research into this trend is needed to understand it more fully. These trends are depicted in Table 27 and Figure 27.

Table 27

Standardized Residuals for Students with Two Identification Points

	2006 Low	2006 Not Low	2007 Low	2007 Not Low
Sped	0.2	-0.3	0.3	-0.3
Low	1.8	-2.0	1.7	-2.0
L Avg	1.7	-1.9	1.7	-2.0
Avg	-0.3	0.3	-0.3	0.3
A Avg	-1.6	1.8	-1.0	1.2
High	-2.0	2.2	-1.5	1.7

Note. Important contributors to the overall effect are highlighted.

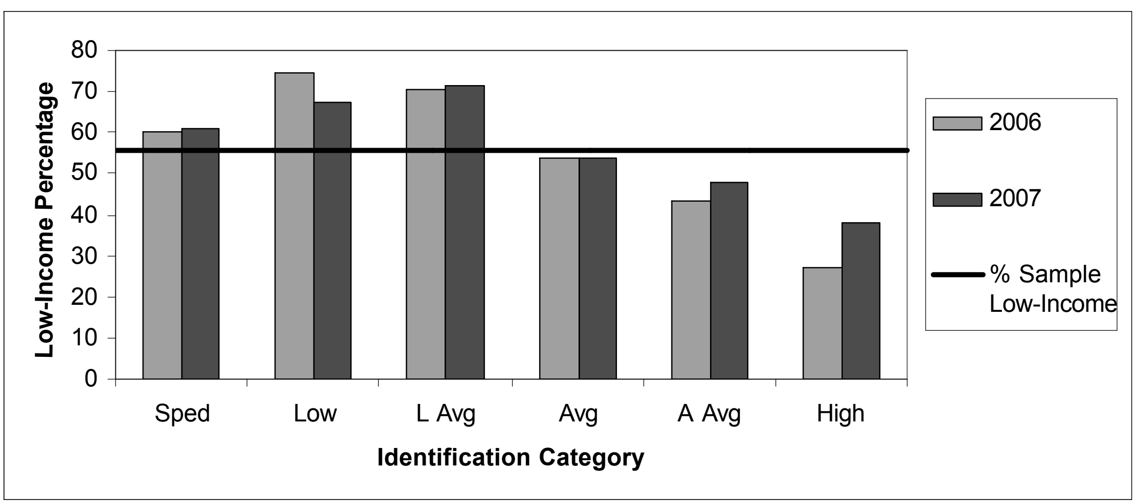

Figure 27. Low-Income Students Percentages by Categories Compared to Total Students with Two Identification Points.

Graduation year 2014. Ninety-one treatment students from graduation year 2014 were present for two identification points. Chi-square analyses were not conducted for graduation year 2014 due to insufficient numbers to satisfy cell size requirements for the test.

Graduation year 2015. Data for students in graduation year 2015 were used to analyze the proportionality of students from low-income families across cluster groups. Pre/post Chi-square analysis for students with two identification points was non-significant at the beginning of the study (Year 1 X^2=8.52, K=5, p=.13) and significant for identification period 2 (Spring 2007) (Year 2 X^2=9.50, K=5, p<.05). This indicates an increase in disparity in cluster placement that occurred over the two study years. Examination of the standardized residuals indicated no significant contributors to this disparity.

Table 28

Chi-Square Results for Graduation Year 2015 with Two Identification Points.

	2006	2007
X^2	8.52	9.50
Df	5	5
P	.13	.05

Table 29

Standardized Residuals for Graduation Year 2015 with Two Identification Points.

	2006 Low	2006 Not Low	2007 Low	2007 Not Low
Sped	0.4	-0.6	0.3	0.0
Low	0.2	-0.2	-0.1	-0.5
L Avg	1.1	-1.6	1.2	-1.8
Avg	-0.3	0.4	-0.1	0.1
A Avg	-0.4	0.5	-0.5	0.8
High	-1.0	1.5	-1.1	1.6

Data for students in graduation year 2015 present for three identification periods were used to examine the proportionality of students from low-income families across cluster groups. Pre/Post Chi-square analysis was non-significant for both identification periods (Year 1 X^2=7.30, K=5, p=.2; Year 2 X^2=4.79, K=5, p=.4). Table 30 shows the Chi-square analysis results.

Table 30

Chi-Square Results for Graduation Year 2015 with Three Identification Points.

	2006	2008
X^2	7.30	4.79
Df	5	5
p<	0.2	0.4

Graduation year 2016. Chi-square analysis for proportionality of representation for students in graduation year 2016 across clusters indicated some disparity in identification based on income status at the beginning of the study. Initial Chi-square results (Year 1 X^2=15.82, K=5, p=.05) and those for the following year are depicted in Table 31. Chi Square analysis for identification period two was non-significant (Year 2 X^2=7.69, K=5, p=.17). Standardized residuals were examined to determine where the disparity lay and can be seen in Table 32. Standardized residuals were generated to investigate areas of non-proportionality. These are shown in Table 32.

Table 31

Chi-Square Results for Graduation Year 2016 with Two Identification Points.

	2006	2007
X^2	15.82	7.69
Df	5	5
$p<$	.05	.17

Table 32

Standardized Residuals for Graduation Year 2016 with Two Identification Points.

	2006 Low	2006 Not Low	2007 Low	2007 Not Low
Sped	0.1	-0.1	0.6	-0.6
Low	2.0	-2.0	1.3	-1.3
L Avg	-0.3	0.3	0.5	-0.4
Avg	0.5	-0.5	-0.7	0.7
A Avg	-0.8	0.8	-0.5	0.5
High	-1.7	1.7	1.0	0.9

Issues of non-proportionality were seen in the low-achieving cluster at the beginning of the study. These were rectified by the second identification period. Pre/post Chi-square analysis could not be conducted for students in graduation year 2016 present for three identification periods as cell counts were too low.

Graduation year 2017. Eighty-four treatment school students in graduation year 2017 were present for two identification periods and seventy-five were present for three identification periods. Neither of these subsamples had sufficient numbers to conduct Chi-square analysis.

Disaggregation of data by income status provided insight into the identification trends in the treatment school across three identification periods. When data were disaggregated by graduation year Chi-square analysis was not possible for some groups. However, those groups for which Chi-square analysis was performed indicated the presence of the same trends. The longer students remained in the program the less likely that income status played a role in cluster identification.

Effects on identification by ethnicity. *What changes occur in frequency of students identified in each of the five achievement groups by ethnicity?* Data were collected for students in six ethnic groups: White, African American, Hispanic, Asian, American Indian, and Other. In order to simplify analysis and try to ensure cell sizes were sufficient for Chi-square analysis these categories were collapsed into two groups. White, Asian, and Other were combined to make a group called, represented. African American, Hispanic, and American Indian were combined to make a grouped called, underrepresented. These groupings were based on those ethnic groups that are traditionally represented or underrepresented in gifted programs according to the literature.

Three identification points. The percentage of students from underrepresented groups was calculated to examine the proportionality of representation by cluster as a comparison to the overall percentage of students from underrepresented groups in the treatment sample. These results are depicted in Figure 28.

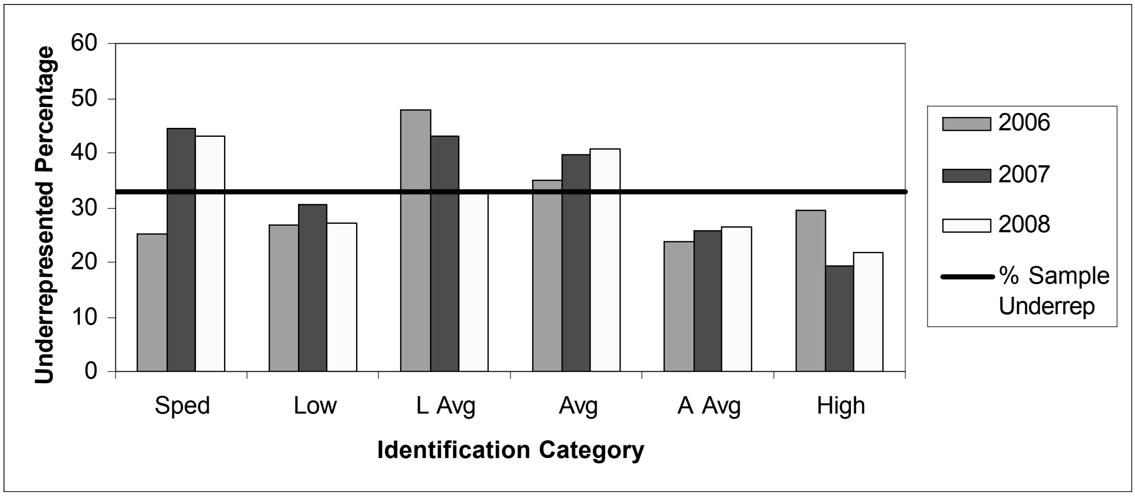

Figure 28. Sample of Underrepresented Students with 3 ID Points by Category Compared to Total Percentage of Sample

Based on examination of the descriptive data, treatment students from underrepresented groups who were present for three identification periods appeared to experience some disparity in cluster group identification. However, Chi-square results were non-significant. These students were overrepresented at the beginning of the study in low-average and average clusters while they were underrepresented in the low, above-average, and high-achieving clusters. This disparity was reduced by the third identification period. Chi-square analysis and examination of standardized residuals for these students' data substantiated this. Pre Chi-square analysis was non-significant for both years. The results are shown in Table 33.

Table 33

Chi-Square Results for Students with Three Identification Points.

	2006	2008
X^2	8.00	6.69
Df	5	5
p	.156	.244

Two identification points. For treatment school students who were present for two identification periods visual analysis indicated that students from ethnic groups traditionally underrepresented in gifted programs were overrepresented in the lower-achieving clusters and underrepresented in the higher-achieving clusters. Figure 29 depicts the percentage of students from underrepresented ethnic groups in each cluster compared with the overall percentage of underrepresented students in the sample.

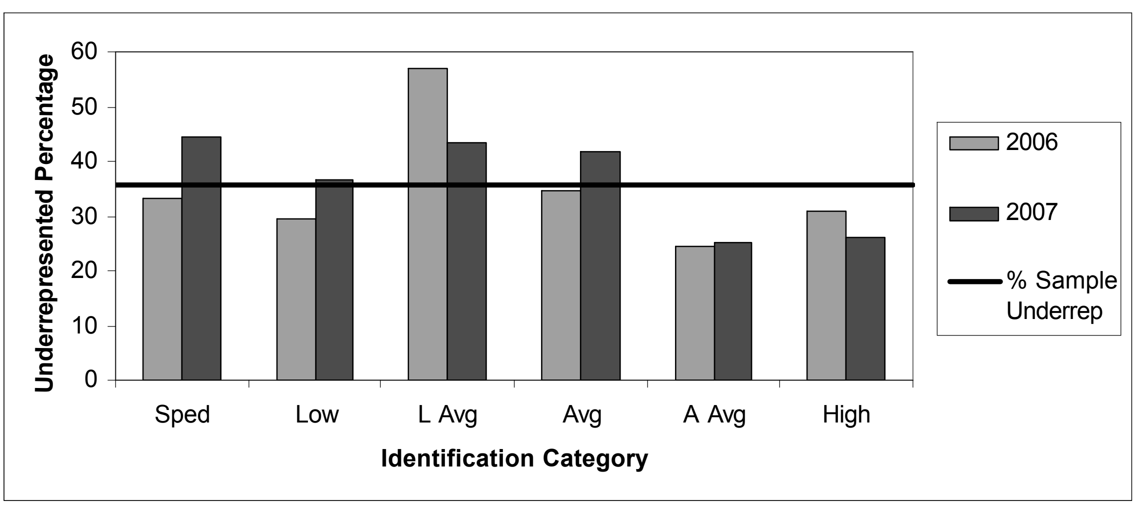

Figure 29. Sample of Underrepresented Students with 2 ID Points by Category Compared to Total Percentage of Sample

Chi-square analyses were performed to see if representation in each cluster was proportional by ethnicity. Both pre and post Chi-square analyses for this group were significant at the *alpha of* .05 level. This indicated disparity in the representation of students from underrepresented groups in the clusters. These results are depicted in Table 34. Standardized residuals were examined to determine where the disparity lay. Students from underrepresented groups were overrepresented in the low-average cluster. Students from represented groups were underrepresented in the low-average cluster, but were proportionally represented in other clusters.

Table 34

Chi-Square Results for Treatment Students with Two Identification Points.

	2006	2007
X^2	20.56	10.76
Df	5	5
$p<$	.001	.056

Table 35

Standardized Residuals for Treatment Students with Two Identification Points.

	2006 Rep	2006 Underrep	2007 Rep	2007 Underrep
Sped	0.1	-0.1	-0.5	0.6
Low	0.5	-0.7	-0.1	0.1
L Avg	-2.3	3.1	-0.8	1.1
Avg	0.1	-0.2	-0.8	1.1
A Avg	1.3	-1.8	1.3	-1.8
High	0.3	-0.4	0.8	-1.0

Note. Important contributors to the overall effect are highlighted.

Cell counts were not sufficient for Chi-square analysis by individual graduation year and ethnicity and therefore were not calculated.

Disaggregation of data by ethnicity provided insight into the identification trends for the treatment school over the three identification periods. Again, length of time in the program reduced the likelihood of disparity in placement by ethnicity. Disparity that was evident in the overrepresentation of students from underrepresented groups in the low-achieving cluster was no longer present at the third identification period. Marked disparity was evident for students who were only present for two identification periods. Both Chi-square analyses were significant, and standardized residuals provided evidence at the beginning of the study of overrepresentation of students from underrepresented groups in the low-achieving cluster as well as a corresponding underrepresentation of students from represented groups in this cluster.

Question Three

To address question three, *What factors exist within the classrooms and school using TSCG that may influence student achievement?* data from interviews with teachers ($N = 10$) were collected and analyzed. Interviews were conducted in May 2009. Interviews lasted approximately 30 minutes and were completed in one day. Twenty-two initial codes were used for inductive coding. These were merged into six major themes: grouping, identification, differentiation, meeting the needs of learners, teacher's influence, and school environment.

Context of the participants. Teachers in the treatment school teach at an elementary school with a low socio-economic, highly transient population. Students who attend this school reflect the population in the school district on a number of variables, including socioeconomic status and ethnicity. None of the teachers in the sample had a gifted education endorsement. High-achieving cluster teachers had been offered professional development in various forms. One teacher had begun the endorsement course offered by the school district, but did not complete it. Three had attended a gifted education summer institute at a local university. All had attended district-sponsored professional development. None of the teachers who taught clusters other than the high-achieving cluster had received any professional development in gifted education other than the initial training all teachers had received. The majority of teachers (8 teachers) had been in the cluster grouping program for the entire study, only two had been there part of the time; one for one year and one for two years.

Table 36 depicts the teacher sample interviewed in the study.

Table 36

Characteristics of the Teachers Interviewed.

Teacher ID	Sex	Grade	Clusters Taught	Gifted Certification	Years in Program
35	Female	2	HA Cluster	None	2
36	Female	3	HA Cluster	None	3
37	Female	5	HA Cluster	None	1
38	Male	4	HA Cluster	None	3
39	Female	6	HA Cluster	None	3
40	Female	2	Other Clusters	None	3
41	Female	3	Other Clusters	None	3
43	Female	4	Other Clusters	None	3
44	Female	5	Other Clusters	None	3
45	Female	6	Other Clusters	None	3

Method of analysis. In order to triangulate data I worked with another researcher to determine initial codes. We coded four interviews together and determined what codes we would use for coding. These codes came from repeated reading of the interviews independently and together during a videoconference. My colleague then read and coded three different interviews independently while I read and coded four different interviews as well as the interviews my colleague coded. We then compared notes and coded the seven interviews together to determine whether we had coded in a similar fashion. My colleague also provided someone to argue the merits of codes - a devil's advocate process as described by Marshall & Rossman (1999). This also added to the rigor of the data analysis. Responses were analyzed by whether teachers taught the high-achieving cluster or not to see if themes differed based on this difference.

Themes emerged through multiple readings of the interview transcripts. Since this was an interview some of the themes came directly from the questions such as leadership support and school environment. Other questions led to larger themes such as differentiation. Data about differentiation techniques were gathered from responses to multiple questions, and became clearer the more transcripts were read. Different techniques were tallied as they were coded and tabulated to assess the most prevalent ones that strengthened the theme.

In the following analysis high-achieving cluster teachers are referred to as such. Teachers who taught clusters other than the high-achieving cluster are referred to as other-cluster teachers. Findings are described in terms of the experiences and perceptions of treatment teachers overall, then for high-achieving cluster teachers, then for other-cluster teachers. Interesting outliers are also included in the description of findings.

Themes. ***Grouping.*** The TSCG model implies ability grouping, thus interview questions focused on the types of grouping used in classrooms between grades two and six. It was determined that grouping was used both within the classroom as well as across each grade level. Teachers reported using flexible grouping within their classrooms. However, flexibility of across-grade grouping was not reported.

The types of within-class grouping reported by teachers included ability grouping (N=10), interest groups (N=5), peer tutoring (N=5), and ability grouping in subject areas other than mathematics and language arts (N=3). Flexible grouping was used by five teachers, those who used interest groups. These changed frequently, based on the projects that students were choosing to pursue or books they were choosing to read.

High-achieving cluster teachers spoke of routinely using grouping in their classrooms as a vehicle for differentiation. Most teachers spoke specifically about ability grouping the high achieving students, but often did not speak about specific groups for the other cluster students present in their classrooms. These groupings occurred predominantly in mathematics and language arts, but were also used in other subjects such as science. Teacher 36 described (personal communication, May 15, 2009) her use of grouping in the following way, “In the small groups when they have choice…I have a science center.” Teacher 35 reported using grouping in this way (personal communication, May 15, 2009), “Right now my kids chose different topics. We started a group research on the rainforest…They helped me develop questions that we wanted to find out.”

Other-cluster teachers spoke about routinely using different reading groups or meeting with mathematical skills groups. Grouping in reading and mathematics was more frequently reported in other-cluster classrooms than in the high-achieving cluster classrooms. Teacher 44 noted her use of ability grouping in reading thus (personal communication, May 15, 2009), “I do flexible groups based on reading levels for reading.”

Between-Class Grouping was reported more by other-cluster teachers. This type of grouping was used for reading remediation. Other-cluster teachers noted that many of their students were pulled from the classroom to attend a specific reading remediation program, which left them with a small number of students who they could work with in small groups in a targeted manner. Teacher 43 noted (personal communication, May 15, 2009), “Most of my kids go…that’s why I group a lot so I can meet their needs. I will do stations quite a bit so I can meet with the small groups that need remediated in... a certain area.” Teacher 45 noted (personal communication, May 15, 2009), “I have seven kids who leave…so, I have guided reading groups. Group them every now and then in math.”

All other-cluster teachers mentioned they used grouping as a method for peer tutoring. No high-achieving cluster teachers reported using this type of grouping. Teacher 43 described her use of peer tutoring in the following way, “I try to put a higher achievement with a lower achievement just to get that peer tutoring in.” Teacher 41 explained it this way (personal communication, May 15, 2009), “A lot of times maybe partnering them up with somebody who’s not there and letting them work together, and it’s kind of nice ‘cause they are the teacher.”

Cluster grouping and its usefulness in identification. Appropriate identification procedures are foundational to the TSCG model. Students are identified for placement in clusters based on teacher observations of classroom performance. Once teachers have made recommendations for placement, the TSCG model stipulates that test results are consulted as a means of inclusion rather than exclusion. For example, the child who does not perform in the classroom, but scores high on an achievement test must be placed in the high-achieving cluster. At the beginning of the study, teachers were taught this identification method. However, during the study teachers of the high-achieving clusters became concerned that other-cluster teachers were not identifying truly high-achieving students, but those who stood out above their peers. This became such a concern that teachers requested that the identification process be changed. The curriculum coordinator designed a referral card that contained a graph with the levels of the clusters correlated with achievement test scores. Therefore, when teachers consulted test scores they were able to determine what cluster students should be placed in according to where their score landed on the graph. This changed the identification process to a more test-based identification process and identification frequencies changed substantially at the end of the third year.

Eight teachers believed that the TSCG model resulted in more students identified as high achieving. Two teachers could not answer this question as they did not have the high achieving students in their classrooms. Of the eight who reported that identification of high-achieving students had increased due to the TSCG model, four noted that they did not feel that this identification was accurate. These teachers felt that when the high achieving students were removed from the classrooms, students identified as above average seemed to perform significantly higher than their peers. Teachers believed that this indicated high achievement and referred them to the high-achieving cluster in subsequent years, but it was found that many of these students were not truly high achieving. Teacher 35 explained it this way (personal communication, May 15, 2009):

> I think what's happening. Here's what I think's happening. I think the truly high ability kids are being pulled out into a cluster with one teacher and what happens is the other above average kids who you know are very capable kids are shining a little bit more in their classroom. And sometimes are being misidentified in subsequent years as high ability when they're truly high average kids.

Teacher 36 noted that the high-achieving cluster teachers were hopeful that this new identification method would facilitate the identification process in the future (personal communication, May 15, 2009):

> Um, but I think that will be improved because we had a…a new data card and a different way to identify that um we're trying this year and I think it's been really helpful, so hopefully that won't happen as much.

Differentiation. Differentiation is the hallmark of a teacher who uses cluster grouping. The very nature of cluster grouping requires that teachers utilize different activities or levels of activities to meet the learning needs of students. Many teachers noted that implementation of the TSCG model did not change their differentiation practices much as they felt they were already doing this. Teachers reported using different differentiation techniques such as compacting, enrichment, choice, leveled questioning, and grouping (discussed as its own theme).

High-achieving cluster teachers noted their discomfort with acceleration. Three high-achieving cluster teachers used compacting as a form of acceleration for students. However, none noted that they did it well or liked doing it. Teacher 38 described compacting the following way (personal communication, May 15, 2009), "I do compacting….but I find to compact really well you end up rewriting so much of what you have. You almost like you're a tailor and you're cutting off parts of the suit." Teacher 39 admitted (personal communication, May 15, 2009), "I'm not good at compacting." No other-cluster teachers used compacting as a technique. Teacher 40 described it thus (personal communication, May 15, 2009), "We've done some tiering of assignments….I don't really use compacting with them just because…they're so close to grade level."

Teachers were asked how they challenged students in their classrooms and were also asked if they used enrichment or interest centers in their classrooms. Four of the high-achieving cluster teachers used enrichment on a consistent basis. This was done in mathematics and language arts predominantly. High achieving students would work on enrichment activities associated with the curriculum, work on enrichment activities in centers, or work on more in-depth activities in literature circles. Teacher 39 described it (personal communication, May 15, 2009):

> I use a lot of the…the enrichment that goes with our math series where they would do like maybe a series of enrichment type things…. doing some of the literature circles where they're together with their groups and…and going on a little bit above and beyond.

One other-cluster teacher used enrichment in her classroom. Other teachers who did not teach the high-achieving cluster noted that there was no need for enrichment in their classes. Teacher 43 explained it this way (personal communication, May 15, 2009), "I haven't used enrichment just because there's not a need for it this year."

There was substantial variation in views expressed by teachers regarding choice. Opinion seemed to be largely based on teaching style. Only two (20%) teachers offered complete choice in reading content. However, nine (90%) teachers offered choice in the type of reading assignment students could complete at the end of a book. Four (40%) teachers offered choice in writing topic on a consistent basis, while the rest of the teachers assigned students a topic write about. Choice in mathematics occurred even less. Only two (20%) teachers offered choice in math assignments or the mode of producing mathematical understanding, while three (30%) teachers offered enrichment activities to their students in mathematics.

Teacher 38 described himself as a highly organized teacher who runs a structured classroom. Students are given choice in all content areas with different choices for literature response, math pentathlon games in addition to or in lieu of the assignment. Teacher 38 also provided choice in projects for students. He noted (personal communication, May 15, 2009):

> [I]n literacy groups [I give students a] text that is leveled to their ability, but [I] give them 20 different choices a week to review the text, and so rather than, for instance, having to do a book report they can write a play or create a diorama or create a picture book or um write a letter to the author or make a radio drama or write a crossword puzzle or write 10 facts about the story or an additional chapter to the end of the story or um create a fictional character to insert during the climax of the story, but different choices where they can be creative and where they can show off their different learning styles.

Teacher 37 reported that she did not provide much choice. She allocated students to reading groups and provided them with the books to read. However, students were given the choice of literature-response assignment. She noted (personal communication, May 15, 2009), “The choice with the novels comes at the end when they get to pick which project they get to do.”

Teacher 39 (personal communication, May 15, 2009) reported similar practices. She did not provide much choice to students.

> Literature…projects. There's a limited…you know a certain amount of choice with writing….We do different writing things, and I may give them a certain genre or a format that I want them to use, but the topic that they use is their choice….Sometimes with the enrichment in the math, you know they'll have some choice, not…not total choice all the time but…but sometimes they can choose.

Teacher 36 reported providing choice throughout the curriculum and encouraged students to create their own projects. She found that students tended to reveal their ability when activities were open-ended. She explained (personal communication, May 15, 2009): "I think my kids tend to shine most in their individual…projects that they create themselves."

Only one other-cluster teacher reported offering choice in homework assignments. Only one other-cluster teacher reported offering students a different way of demonstrating understanding. Teacher 44 noted (personal communication, May 15, 2009), "There's lots of choice in things that they get to do….If they like to write, they get to write. If they like to draw, they get to draw, act out, or make a song, whatever."

Four (80%) high-achieving cluster teachers reported using leveled questioning in their classrooms. Only one high-achieving cluster teacher mentioned the importance of wait time. She noted that different types of students would answer at different lengths of wait time. She explained it thus (personal communication, May 15, 2009):

> I think…an important strategy is wait time, and if you wait long

> enough for an open-ended response, then those students who really are thinkers and really are creative are more willing to share what's going on in their mind. If you want a quick flippant response, you'll get lots of average kids who jump on the bandwagon for that response, but if you want something that is insightful, then you need to wait for it as a teacher and not be so giddy and impatient that you cheat the kids who are really thinking in the classroom.

Three other-cluster teachers noted the importance of wait time when questioning students with higher levels of questions. Teacher 40 explained it as follows (personal communication, May 15, 2009):

> I try to ask questions at all types of different levels, try to give them wait time before I ask for answers…. I try to call on somebody that might/might not have the right answer and then we try to work through.

Meeting the Needs of Learners. Teachers were asked if the TSCG program had helped them meet the needs of individual students. Two high-achieving cluster teachers agreed that the program had helped them meet students' needs by knowing who the high achieving students were right at the beginning of the year. Three (60%) high achieving teachers reported that they did not think that the TSCG model had helped them meet student needs. These teachers felt that they were already differentiating in their classrooms before the model was implemented. Teacher 38 noted that any proficient teacher would identify the high achieving students very quickly without knowing they were grouped in the classroom. He noted (personal communication, May 15, 2009):

> No. I think that a teacher who has been in the field for any length of time can spot kids that are quite obviously higher than their peers, and that teacher by the very

> nature of having a student like that needs to meet their needs by separating them, by giving them more challenging work that enriches their education and just having a slip of paper that says, “Be forewarned this kid is gifted”. I don’t really think is necessary for most teachers because they can spot a gifted kid within the first 30 minutes of the day.

All other-cluster teachers reported their belief that the implementation of the TSCG program did assist them in meeting the needs of their students. The use of cluster grouping allowed them to better focus remediation for students in their classrooms. This was done through within-class and between-class grouping. The grouping of students facilitated the between class grouping for reading remediation. When students left the room for this program teachers were left with specific small groups of children who they felt they could better target for interventions. Teacher 40 noted (personal communication, May 15, 2009):

> We have guided reading groups for instruction…where…each individual group meets with me. This year, we’ve also worked as a grade level so we grouped the students and then I have a whole class that are all at about the same level…of 18 students that come to me.

Teacher 40 further noted that the restriction of achievement range in her classroom also facilitated three leveled groups in mathematics (personal communication, May 15, 2009).

Influence of teachers. Teachers reported that classroom climates were positive even if some were more formal than others. Classrooms ranged from places where children could work anywhere they chose and sit on a variety of furniture to more formal environments where students were expected to be more business-like about their learning.

During the interviews, high-achieving cluster teachers' descriptions of their classrooms included (personal communication, May 15, 2009):

> I think there's a serious atmosphere in the classroom, but I also think that there's an underlying freedom where kids can ask questions and be treated in a way that makes them feel older than they are….I like to take questions seriously in my classroom, and I like kids feeling like they have a voice and that there's no stupid question…but at the same time, I don't really feel like my class is a silly place or a goofy place either, so I think that… there is a seriousness…in the atmosphere of my classroom. (Teacher 38)
>
> I work really hard to try to make it more like…what we call it our class family….So we kind of talk about how family should treat each other and… that's kind of our theme I guess you could say. (Teacher 36)

Other-cluster teachers described their classrooms more as busy, sometimes noisy places. These teachers reported offering students alternative workspaces or exercise balls to sit on. Routine was also an important aspect of these classrooms. Descriptions of other-cluster classrooms included (personal communication, May 15, 2009): "I would describe it as a caring atmosphere. The kids are pretty free to do what they need to do. They know what supplies they can get. It's comfortable." (Teacher 43) and "It's always moving. Always busy." (Teacher 40), and "It's very calm.…I have to have a schedule and a routine". (Teacher 40).

All teachers reported having high expectations for their students regardless of cluster. One sixth grade teacher noted that her expectations were that all students would leave her classroom academically and socially ready for middle school. All teachers reported that they believed that all students could learn and that mastery of grade level standards was necessary.

High-achieving cluster teachers noted that although their academic expectations were high they had different expectations for different children. It was clear from their responses that the high achieving students were expected to perform at a higher standard than their peers. One high-achieving cluster teacher noted that sometimes students who were not identified as high achieving also met these higher expectations and surprised her and themselves. Responses from these teachers included (personal communication, May 15, 2009): “extremely high.” (Teacher 38). “I think they are different….I expect that they can eventually do those things….I feel like it’s my job to help move them along the path to get there.” (Teacher 35), and “You set that bar high for everybody and the expectation with the high cluster kids is that they’re going to meet that and the other kids sometimes surprise you and do it too.” (Teacher 39).

Other-cluster teachers all responded that their expectations were that all students master the grade level standards. Two (40%) noted that this may take longer than usual, but was still a realistic goal. One teacher noted that that students should do their personal best and that this should mean that they could move beyond grade level standards. Responses included (personal communication, May 15, 2009): “I expect them to master the fifth grade standards just like everybody else…sometimes it takes us longer to get there.” (Teacher 44). “It’s really to do their personal best but yet I want to challenge them to beyond grade level…to do better things.” (Teacher 41).

School environment. Teachers were asked if the administration had been supportive of the TSCG model implementation. No teachers reported that the administration had not been supportive, but several teachers reported that the support was more passive or reactive, rather than overt. This was reportedly due to the change in leadership during the study period. The principal who was present at the beginning of the study was very proactive in the implementation of the model. However, she left the school at the end of the first year. The principal who was present for the remainder of the study was very positive about the program and had expectations that teachers would continue to implement the model, but did not have much knowledge about the specifics of the model and did not actively engage in dialogue with teachers about it. This may have resulted in some inconsistencies in implementation.

All high-achieving cluster teachers reported that there was administrative support, but only one teacher described it as strong administrative support. Teacher 39 noted (personal communication, May 15, 2009), “I believe the administration has been very supportive….I think that’s what our principals have wanted us to do…they’ve been supportive of it.” Other high achieving teachers described administrative support as (personal communication, May 15, 2009):

> I feel like in the beginning it was very well supported….I’m not positive that our administration has total buy-in, so you know they’re still taking things like parent requests….You know it kind of defeats the purpose when you have a couple of random kids that aren’t in a cluster. (Teacher 36)
>
> I certainly think that it’s, that it is supported, but I can’t give you, I can’t think of an example of how it is, it’s just, It’s sort of the expectation and, and it’s not even, um. I don’t think it’s really been a situation where there was…another option. (Teacher 35)

The district was supportive of the program. The curriculum director had monthly trainings for high-achieving cluster teachers and provided professional development and collaboration time for these teachers. However, no model-specific professional development was provided for other-cluster teachers. Teacher 35 noted (personal communication, May 15, 2009):

> A district coordinator has had six different times this year we've had staff development for half day…. The other things that I really appreciated this year is that we've been given time in the context of those half day trainings to do…some sharing with teachers in our grade level at other schools.

Other-cluster teachers noted that little sharing had been done between teachers at their grade level as grade-level planning time had been removed from their schedules. This prevented much transfer of knowledge gained at training from the high-achieving cluster teacher to his/her colleagues. One other-cluster teacher referred to the high-achieving cluster trainings as secret meetings.

> As far as cluster grouping though, because I've not done…because most of our cluster grouping um professional development has been with just the high cluster people, and they leave the building you know a couple times a month and have their own little secret meeting. It's a secret society. (Teacher 44, personal communication, May 15, 2009)

Qualitative findings provided context for some of the quantitative findings. Analyses of the qualitative data produced six core themes: grouping, differentiation, cluster grouping and its usefulness in identification, meeting the needs of learners, influence of teachers, and administrative support. Various forms of grouping were used by treatment teachers. Within-class grouping was most prevalent with between-class grouping used most for remediation. Differentiation was a hallmark of all teachers regardless of cluster taught and was practiced in several forms such as enrichment, compacting, leveled questions, and choice in assignment and mode of learning. Most teachers reported that they believed that the TSCG model resulted in increased identification of students as high achieving, but most teachers did not believe that identification was accurate, which resulted in a change in the identification process at the end of year three. Teachers ran positive classrooms and held students to high standards of performance. With these high expectations came the understanding that students work in different ways and at different rates. Teachers reported using a variety of classroom practices to meet these diverse needs. Administrative support was reportedly strong at the beginning of the study; however, a change in leadership resulted in continued support of the model, but a lack of knowledge about it. This resulted in teachers feeling that support was in the form of expectation to implement the model, but not much support in the implementation. Professional development was available for high-achieving cluster teachers, but not for other-cluster teachers. This was a point of contention for some teachers. All teachers involved in the program responded that they felt the TSCG model was beneficial to all students and helped them meet students' individual needs.

CHAPTER FIVE: DISCUSSION

Findings for Research Question One

The primary purpose of this study was to investigate the effects of TSCG on the achievement and identification of urban elementary students. This study focused on the influence of ethnicity and socioeconomic status on achievement and identification, as this study was a partial replication of a study conducted in a rural school district that enrolled only Caucasian students, many from low-income families. This difference was purposeful to investigate if the TSCG model was useful in a different context. Modeling of the data included ethnicity and socioeconomic status as covariates for both identification and achievement growth.

Differences in growth rate by ethnicity and socioeconomic status. Research question one investigated the differences in growth trajectories between the treatment and comparison school based on four groups of students grouped by ethnicity and socioeconomic status. Findings indicated that socioeconomic status was less of a predictor of achievement growth than ethnicity. This was not surprising as the school district served students from predominantly lower-income families; therefore ethnicity was the covariate that better distinguished students from one another. Students from ethnic groups traditionally represented in gifted education programs achieved at a higher level than their underrepresented peers. Therefore, students from represented groups were more likely to be identified for the high-achieving cluster than their underrepresented peers, at least at the beginning of the study. This finding supported Yoon and Gentry's (2009) finding regarding representation of students in gifted programs by ethnicity favors those ethnic groups included in the represented group in this study (Caucasian, Asian, and other). These students would have had to perform at a higher level than their underrepresented peers to be identified.

Differences in achievement growth by condition. Although differences were found between students from different ethnic and socioeconomic groups, differences were also found between the treatment and comparison schools. Students at the comparison school began and ended the study at higher levels of achievement in both reading and mathematics. However, their average growth rate was lower than for treatment students. This was a marginal difference and may be due to several factors. This study was a small study using only one treatment and comparison school pair. Therefore, differences in schools may account for differences in growth despite the attempt to match schools as closely as possible. However, since the identification of students for the high-achieving cluster increased over time in the treatment school and proportional representation, by ethnicity and socioeconomic status, of students in the high-achieving cluster became more representative throughout the study, differences are more likely due to the implementation of the TSCG model. This supports Gentry and Owen's (1999) finding that student achievement increased over time for students in schools where the TSCG model was implemented.

Marginal differences in achievement gains may also be due to lack of complete implementation of the TSCG model. Gentry and Mann (2008) noted that all teachers, regardless of cluster taught, should be thought of as cluster teachers. All teachers should receive professional development in differentiation techniques and use the same pedagogy with leveled materials appropriate for the students in each classroom. This did not occur in the treatment school. Teachers of the high-achieving clusters were perceived as the cluster teachers while other teachers did not see themselves, nor were perceived by the district, as cluster teachers. No model-specific professional development was given to other-cluster teachers after the initial training prior to implementation. However, high-achieving cluster teachers received model-specific training at least six times per year and were offered the opportunity to participate in gifted education courses and summer institutes. This may account for the increase in achievement of high-achieving students seen in the treatment school, as high achieving cluster teachers were better able to serve their high achieving students, but may well have been evident at all levels if every teachers had received the same training. Increases in achievement in other clusters may well be accounted for by the reduction in range of student ability due to the use of the cluster grouping model. This was confirmed by other-cluster teachers who noted that remediation was easier and facilitated between-class grouping due to the grouping of students that occurred in the implementation of the model. Overall, gains were greater in reading than in mathematics for treatment students. This may be due to the fact that implementation of the TSCG model facilitated between-grade grouping for reading remediation for students in the lower-achieving clusters. There was also a focus on reading remediation in the treatment school, which may also account for greater gains in reading achievement.

Differences in achievement growth within the treatment school based on initial identification. Students initially identified as low achieving made greater gains than their high-achieving peers, which supports Gentry and Owen's (1999) finding that implementation of the TSCG model influenced the achievement growth of all students. These gains may have been even greater had the other-cluster teachers received professional development in gifted education strategies. This finding is contrary to Oakes (1985; 1995; Slavin, 1987) contention that teachers of low-achieving students are poor teachers. Indeed, these teachers continued to have high expectations for their students, which contributed to gains in their achievement. These expectations are evidenced in the following teacher comments (personal communication, May 15, 2009):

> I expect them to master the fifth grade standards just like everybody else…sometimes it takes us longer to get there (Teacher 44)
>
> It's really to do their personal best but yet I want to challenge them to beyond grade level…to do better things." (Teacher 41)

Differences in achievement growth within the treatment school based on ethnicity and socioeconomic status. This finding was non-significant. With the exception of students initially identified as low-achieving in mathematics. There was no relationship between achievement growth trajectories for treatment students when socio-economic status and ethnicity were taken into account. Students in this group achieved at a marginally higher level than their higher-achieving peers at the end of the study. However, their rate of growth over the course of the study was not significantly different. There is no clear reason for this distinction in final achievement level, except that this group may have received the most remediation given that they were the lowest achievers.

Research Question Two

Length of time in study as a predictor of higher identification. For students present for three identification periods the number of students identified as high achieving increased over time. Students who were only present for two identification periods were less likely to be identified as high achieving. This supports the findings of Gentry and Owen (1999). Grouping high-achieving cluster students in one classroom per grade allowed other students the chance to perform academically and resulted in more students identified as high achieving. Although high-achieving cluster teachers asserted that some students were misidentified and changed the identification process, more students were identified over the three identification periods and achievement increased overall at greater rate than for comparison students thus indicating the positive effects of the model on student identification and achievement.

Length of time in study as a predictor of proportional representation in clusters. Students who were present in the study for longer periods of time were more likely to be represented in a proportional manner than those who were more transient. This provided evidence that teachers are more able to note achievement gains or strengths in students if they are able to evaluate them over time. Consistent with the findings of Donovan and Cross (2002) and Yoon and Gentry (2009) the study began with issues of proportionality of representation in clusters by both ethnicity and socioeconomic status. Students from underrepresented ethnic groups and those from lower-income families were overrepresented in the lower-achieving clusters and underrepresented in the higher-achieving clusters. The opposite was true for students from represented groups and not low-income families. However, as students spent more time in the cluster-grouping program issues of proportionality resolved and all clusters became more representative of the population. This finding is supported by the answers given by teachers when asked if they viewed students from lower-income families and underrepresented groups differently from their peers. All teachers answered that they did not see underrepresented students any differently from their represented peers. They also noted that their perceptions did not change throughout the course of the study. Therefore it is not clear if proportionality increased as a result of the TSCG model or as a result of teachers who were acculturated to their environment and did not see ethnicity or socioeconomic status as issues that affected any one child more than another. Teacher beliefs are represented in the following teacher comments (personal communication, May 15, 2009):

> How I perceive the child? I see them as a child. It doesn't matter to me their achievement. I know that they need love and they need education just like every other student, so I really honestly see a child. Most the time I don't even see the color of their skin. I just see the child and what they need and I know where I need to meet them. (Teacher 43)
>
> I don't change my expectations for my students. (Teacher 40)

The cause of over-representation in lower-achieving clusters for underrepresented students is difficult to discern. This could be due to environmental or biological factors that affect a child's background knowledge and/or ability to retain learning. Continued overrepresentation of this group in lower-achieving clusters compared to underrepresentation of represented groups is cause for concern as there may be issues of bias or misdiagnosis. Underrepresentation of underrepresented groups in high-achieving clusters is also of concern. This could also be an issue of bias or of a lack of differentiation to build background knowledge necessary for high performance. Over time this issue should resolve itself if true differentiation is used in the classroom and underrepresented students are given opportunities to perform at higher levels. Better adherence to the TSCG model or more time using the model may have remedied some of the disproportionality seen in cluster group identification. Further research is needed to evaluate this finding.

Question Three

Practices and perceptions of classroom teachers. Qualitative findings in this study yielded interesting results that helped explain the trends in both identification and achievement in the treatment school over the course of the study. High-achieving cluster teachers believed that the increase in the number of students identified during the three identification periods was due to the fact that the high-achieving students had been removed from most classrooms at the beginning of the study and above-average students were more visible to teachers. However, teachers of the high-achieving clusters expressed that students who were subsequently identified were higher achieving than their peers, but not truly high achieving. This resulted in the changes to the identification procedures. This and other themes are discussed below.

Identification. This increase in the frequency of students identified as high achieving throughout the study supports previous findings by Gentry and Owen (1999) that removal of the highest achieving students from all but one classroom per grade level provided the opportunity for other students to achieve at higher levels. This might not have occurred had the high-achieving students remained in the classroom.

High-achieving cluster teachers' lack of support for the identification process in subsequent study years is a different finding from previous studies and is of concern. Initial training on identification was provided to all teachers, regardless of cluster. However, ongoing training was only provided to high-achieving cluster teachers. This may have affected the identification process as teachers did not have a clear understanding of what skills were necessary to qualify for the high-achieving cluster. Gentry and Mann (2008) argued that students in the high-achieving cluster should be high-achieving in both reading and mathematics. Students who are high-achieving in one of those areas are above-average achievers. Teacher 38 noted, "I've had kids that were identified…high achieving this year who are not gifted. They were above average, but they really should not have been in my class." This was supported by 80% of high-achieving cluster teachers as well.

A combination of identification procedures and grouping may have been responsible for changes in identification in this study. Teachers were better able to see the achievement of their students as they remained in the program over time. This resulted in an increase in students identified as higher achieving over the course of the study. As student achievement increased over time or was recognized by the teacher, they were identified as higher achieving. The TSCG program offered students the opportunity to perform by removing the highest achievers from the classroom.

These results suggest that cluster grouping may improve how teachers view their students with respect to achievement and ability. When the highest achieving students are grouped in one classroom, teachers in other classrooms have the opportunity to recognize the talents and achievements of others. In addition, students who may not have been recognized as achievers are given the opportunity to perform when the traditional high achievers are removed from the room. These findings refute the findings of previous researchers who noted that grouping is damaging to low-achieving students (Black, 1983; Oakes, 1985; Slavin, 1987). These findings contribute to the field of research on ability grouping (Gentry & Owen, 1999; Gentry & Mann, 2008; Kulik & Kulik, 1992; Rogers, 1991), and therefore, should be considered when making decisions regarding identification and placement of students in classrooms. This model may provide students with opportunities for their achievement to be recognized by teachers as they grow as learners.

The TSCG model stipulated that cluster grouping should reduce the range of student achievement levels present in the classroom, which allows the teacher to be more intentional in planning instruction (Gentry & Mann, 2008). In this study several (30%) teachers noted that the range of students was broader than expected with the cluster grouping. Other teachers (30%) felt that just knowing that a student was a high-achiever did not make much difference as any good teacher should be able to spot the high achievers almost immediately. Teacher 38 noted (personal communication, May 15, 2009):

> No. I think that a teacher who has been in the field for any length of time can spot kids that are quite obviously higher than their peers, and that teacher by the very nature of having a student like that needs to meet their needs by separating them, by giving them more challenging work that enriches their education and just having a slip of paper that says, “Be forewarned this kid is gifted” I don’t really think is necessary for most teachers because they can spot a gifted kid within the

> first 30 minutes of the day. The very first 30 minutes of the first day of school, you start seeing it. It's just obvious. So, I don't really think that identifying them necessarily is as important as knowing how to meet their needs.

These findings suggest that cluster grouping and/or identification are not sufficient to improve how teachers perceive student achievement or to optimally increase student achievement. All teachers must receive professional development in the pedagogy necessary to teach students in a differentiated manner in order ensure that all students are receiving appropriate instruction. Hansen and Feldhusen (1994) found that teachers trained in gifted education pedagogy evidenced superior teaching skills and created a more positive classroom environment than their counterparts who did not receive gifted education training. Further, Robinson, Shore, and Enerson (2007) found that professional development for teachers is critical to a successful education for all students. Robinson (2008) also noted that teacher content-knowledge, gifted education in this context, is critical for engaging and motivating students in the classroom. Therefore, these research foundations suggest that student achievement may have been affected by the lack of professional development provided to all teachers. Misconceptions such as those of Teacher 38 may have been altered. In addition, all teachers may have understood the need for differentiation if targeted professional development had been offered as stipulated by the TSCG model.

Consistent with the identification findings, quantitative analyses of student achievement in reading and mathematics revealed that achievement scores improved slightly in the treatment school. Over the three years of the study, treatment school student achievement increased at a greater rate than student achievement in the comparison school. However, the rate of growth was only marginally greater for the treatment school.

Qualitative findings indicated that students in the treatment school were grouped into the five clusters, plus special education, stipulated by the model. Some students (from two teachers) were regrouped for interventions in reading and mathematics while high-achieving students were offered enrichment in these content areas. However, few other changes were made. Teachers of all clusters did not receive professional development in differentiation strategies that are crucial to the success of the TSCG model. High-achieving cluster teachers did receive some training in this area, but many (80%) cluster teachers mentioned that this was not new training as they differentiated before becoming the high-achieving cluster teacher. When asked if the implementation of the model helped teachers meet the needs of learners. High-achieving cluster teachers noted that their practice had not changed much, but other-cluster teachers (100%) said clustering had helped in remediation efforts with lower-achieving students.

Teachers appeared to be more comfortable with remediation strategies for lower achieving students. Teacher 40 noted (personal communication, May 15, 2009):

> I have guided reading groups…about five different guided reading groups, and then we also have Read 180, and so I have seven kids who leave and go to Read 180. So, it's nice because I only have 19 kids left when those seven leave, so it's nice. I have a nice small group.

These strategies, along with targeted cross-grade grouping for intervention may partially explain the increase in achievement for lower achieving students.

The TSCG model suggests that identification of students in one of five clusters is done to reduce the range of achievement levels that the teacher must teach as not all clusters are placed in each classroom. This requires teachers to accurately identify students for the clusters based on classroom performance and test performance as a means of inclusion. Many teachers in the treatment school felt that the identification process was not accurate as they ended up with wide ranges of student achievement levels, or students were placed in their classroom during the year and did not fit into the clusters that were assigned to them. This reduced the efficacy of the cluster grouping and may have influenced the rate of achievement growth for students.

Teacher effects. Researchers have repeatedly noted that grouping alone does not contribute to changes in student achievement. Rather, it is what occurs within the groups and the teacher interactions with students that effects achievement. All teachers noted that they had high expectations for students regardless of cluster taught. Some (60%) noted that their expectations differed for different groups of students based on their level of achievement; however, all were expected to work diligently and achieve at a higher level by the end of the year. All teachers claimed to use a variety of strategies to meet the needs of all learners in their classrooms and a few (30%) attempted to work collaboratively with their colleagues to meet these needs. Collaboration was an issue though, as it was available to teachers as grade-level planning time at the beginning of the study, but was not by the end. This finding is contrary to Oakes' (1995) findings that teachers of lower achieving students were poor teachers and had lower expectations of their students. Teachers in this study, regardless of cluster taught, did not have lower expectations for their students and continued to expect academic growth in all students.

Differentiation and what has traditionally been known as gifted education pedagogy should be taught in all pre-service teacher programs to prepare all teachers to meet the needs of all students. These professional development opportunities were only offered to high-achieving cluster teachers in this study and therefore, other cluster teachers did not have the benefit of this training. When teachers know how to differentiate instruction in a meaningful way – providing challenging work with leveled materials – then students begin to think in more complex ways and achieve at higher levels. Without differentiation, remediation often becomes the focus for lower-achieving students. Remediation pedagogy is often confined to asking students lower-level questions and work consists of more rote learning than higher-order thinking. This pedagogy does not challenge or motivate students and thus does not always promote increased student achievement.

School environment and leadership support. Delcourt and Evans (1994) noted that schools that were successful in grouping students and raising student achievement exemplified the following characteristics: leadership, supportive school atmosphere, positive school environment, and flexible and appropriate curriculum and instruction. Teachers of the lower-achieving clusters were either ambivalent or positive about the TSCG program, but did not feel a part of the program. They did not receive the specific training associated with implementation of the model and did not consider themselves cluster teachers. The teachers of the high-achieving clusters were referred to as the cluster teachers. This finding is contrary to Delcourt and Evans and Gentry and Owen’s findings.

Leadership was also an issue during this study. The principal of the treatment school changed twice over the course of the study. The original principal was very supportive of the model. The second principal was supportive, but in a more reactionary manner. She did not actively provide teachers with time to plan or collaborate; however, she did expect that teachers would implement the model. When teachers had issues with the identification the principal supported changes to the process. Hallinger (2003) noted that effective leadership is context specific. Effective leaders are influenced by changes in the school. They also effect change in their schools. This did not seem to occur in the treatment school. While the administrator was supportive of the TSCG model and its implementation, she was not proactive in promoting the model or providing professional development necessary for optimal implementation. Further, May and Supovitz (2011) noted that educational leaders who concentrate on the improvement of specific teachers while keeping a broader focus can produce larger changes in instructional practices (Keck & Hallinger, 2010). All grade two through six teachers should have received professional development directly related to implementation of the TSCG model This targeted professional development may well have had a trickle-down effect for the rest of the school.

Limitations

One of the most obvious limitations of this study is the use of only one treatment and comparison school. Therefore, additional research is required on additional similar samples to increase generalizability. A second is that this study was done within the context of a school. No single variable was studied, but an entire program in a real school was studied. Schools are multifaceted places with many different processes existing therein. Therefore, the results of this study are multifaceted. Despite this limitation, the results are important in that they show the potential of the TSCG model to work for students who remain in the program over time. Much can be learned about the importance of each component of the TSCG model for successful implementation. In addition, teacher perceptions of the model offered rich information about how administrative decisions can affect the success or failure of an implementation. Although the intention was that all teachers were considered cluster teachers, the practice resulted in different perceptions.

As with any study of a model such as the Total School Cluster Groping model, a single variable examination would not be sufficient to show growth. All of the components that were implemented or not implemented within the school must be considered as they contribute to the success or failure of grouping. The intention of this study was to understand the model within the context of the school to see if previously observed increases in achievement using this model could be replicated in a very different environment. To this end, it is clear that TSCG must be implemented with fidelity to ensure student success. Furthermore, students should remain in the program for longer period of time to see greater success.

This study was a quasi-experimental, matched comparison study, which inherently incorporates areas of potential measurement error. However, this type of study is powerful in other ways. This study examined achievement in a real-life school setting rather than in a controlled, clinical environment. Second, it investigated an implementation initiated by the school district rather than an implementation requested by a researcher or demanded by other external sources. This resulted in a more vested interest in the implementation of the model, which is evidenced in how the high-achieving cluster teachers worked and the quality of the professional development offered to them. Although this was a transient population, the cohort nature of this study afforded insight into a group of students over time, which allowed for comparisons to be made for the same students. This study adds to a small but growing body of research about cluster grouping. It partially replicates the study completed by Gentry and Owen (1999) and provides important information to the field of gifted education.

Qualitative findings were limited in the study in several ways. A small number of teachers were interviewed, which limited the amount of information available for analysis. In addition, all high-achieving cluster teachers were interviewed, but only one other-cluster teacher per grade was interviewed. These other-cluster teachers were given the choice to participate, which influenced the type of teacher who was interviewed, rather than a cross-section of teachers.

Directions for Future Research

This study was limited by the use of only one treatment and comparison school. This limited the amount of causality that can be associated with findings. Addition of more treatment and comparison schools would increase the generalizability of findings. This increase would also offer researchers a larger pool of teachers for qualitative data gathering, which would provide more in-depth understanding of the beliefs and practices of teachers who implement the model. Although random sampling of teachers reduced some of the bias associated with the gathering of qualitative data, the fact that teachers could refuse to participate in the interviews resulted in a somewhat biased sample of teachers who were willing to talk about their practices and beliefs about the TSCG model. Offering teachers the chance to complete an anonymous survey might reduce the feeling of vulnerability teachers may feel during the interview process and increase the number of respondents. Additionally, increased fidelity in the implementation of TSCG model would provide further insight into the efficacy of this model in an elementary school setting. These additions to the research design will provide researchers with greater ability to ascertain the causes of change in student achievement in a cluster grouping setting.

REFERENCES

REFERENCES

Anastasi, A. (1988). *Psychological testing*. Prentice-Hall, Inc.

Anastasi, A. (1996). *Psychological testing* (7th ed.). New York: McMillan.

Anastasiow, N. J. (1968). A comparison of two approaches in upgrading reading instruction. *Elementary English, 45*, 495-499.

Barr, R. (1992). Teachers, materials, and group composition in literacy instruction. In M. J. Dreher & W.H.Slater (Eds.), *Elementary school literacy: Critical issues* (pp. 27-50). Norwood, MA: Christopher-Gordon.

Barr, R., & Dreeben, R. (1991). *How schools work*. Chicago: University of Chicago Press.

Bear, C. (1998). An evaluation of the effects of cluster grouping on the academic achievement of elementary students in the regular classroom. Unpublished Dissertation. Saint Louis University.

Black, S. (1993). Derailing tracking. *Executive Educator, 15(1)*, 27-30.

Bollen, K.S., & Curran, P.J. (2006). *Latent curve models: A structural equation perspective.* Hoboken, NJ: Wiley.

Bowen, G.A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research, 8*(1), 137-152.

Brulles, D. (2005). An examination and critical analysis of cluster grouping gifted students in an elementary school. (Unpublished Doctoral Dissertation, Arizona State University).

Bryant, M. A. (1987). Meeting the needs of gifted first grade children in a heterogeneous classroom. *Roeper Review, 9*, 214-216.

Chorzempa, B., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. *Journal of Educational Psychology, 98*(3), 529-541.

Collins, J. (1986). Differential instruction in reading groups. In J. Cook-Gumperz (Ed.), *The social construction of literacy* (pp. 117-137). Cambridge: Cambridge University Press.

Creighton, J.L. (2005). *The public participation handbook.* Hoboken, NJ: John Wiley and Sons.

Davies, J., Hallam, S., & Ireson, J. (2003). Ability groupings in the primary school: Issues arising from practice. *Research Papers in Education, 18*(1), 45-60.

Delcourt, M. A. B., & Evans, K. (1994). *Qualitative extension of the learning outcomes study.* Storrs, CT: The National Research Center on the Gifted and Talented.

Delcourt, M. A. B., Loyd, B. H., Cornell, D. G., & Goldberg, M.D. (1994). *Evaluation of the effects of programming arrangements on student learning outcomes.* Storrs, CT: The National Research Center on the Gifted and Talented.

Donovan, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education.* Washington, DC: National Academy Press.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education, 54*, 151-161.

Eder, D., & Felmlee, D. (1984). The development of attention norms in ability groups. In P. Peterson, L. Wilkinson & N. Hallinan (Eds.), *The social context of instruction* (pp. 189-208). San Diego, CA: Academic Press.

Elbaum, B., Vaughn, S., Hughes, M., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children, 65*(3), 399-416.

Floyd, C. (1954). Meeting children's reading needs in the middle grades: A preliminary report. *The Elementary School Journal, 55*(2), 99-103.

Ford, D. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *The Journal of Special Education, 32*(1), 4-14.

Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Another look at the achievement gap: Learning from the experiences of gifted black students. *Urban Education, 43*(2), 216-239.

Gallagher, J. (1993). Ability grouping: A tool for educational excellence. *The College Board Review, 168* , 21-27.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools. *Review of Educational Research*, 57, 415-435.

Gentry, M. L. (1999). *Promoting student achievement and exemplary classroom practices through cluster grouping: A research-based alternative to heterogeneous elementary classrooms.* Storrs, CT: University of Connecticut.

Gentry, M. L. (2008). *Total School Cluster Grouping: Preliminary findings from year 1 of longitudinal replication study.* Paper presented at the American Educational Research Association, New York.

Gentry, M. & Owen, S.V. (1999). An investigation of total school flexible cluster grouping on identification, achievement, and classroom practices. *Gifted Child Quarterly, 43*, 224-243.

Gentry, M. & Mann, R. L. (2008). Total school cluster grouping: A comprehensive, research based plan for *raising student achievement and improving teacher practices.* Mansfield Center, CT: Creative Learning Press.

Gentry, M., Gable, R. K., & Springer, P. (2000). Gifted and nongifted middle school students: Are their attitudes toward school different as measured by the new affective instrument, My Class Activities ...? *Journal for the Education of the Gifted,* 24(1), 74-96.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Grant, K., & Rothenberg, J. (1981). *Charting educational futures: Interaction patterns in first and second grade reading groups.* Paper presented at the American Educational Research Association, Los Angeles.

Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupil's experiences of different types of grouping in school. British Educational Research Journal, 30(4), 515-533.

Hedeker, D., & Gibbons, R.D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*(1). 64-78.

Hoover, S., Sayler, M., & Feldhusen, J.F. (1993). Cluster grouping of elementary students at the elementary level. *Roeper Review*, 16, 13-15.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hurley, E.A., Boykin, A.W. & Allen, B.A. (2005). Communal versus individual learning of a math-estimation task: African American children and the culture of learning tasks. *The Journal of Psychology: Interdisciplinary and Applied, 139*(6), 513-527.

Indiana Department of Education. (2008). *ISTEP+ program manual*. Indianapolis, IN: Author.

Indiana Department of Education. (2008) Corporation snapshot. Retrieved November 27, 2008.

Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education, 25*(3), 343-358.

Kanevsky, L., & Keighley, T. (2003, Fall). To produce or not to produce? Understanding boredom and the honor in underachievement. *Roeper Review*, *26*(1), 20-28.

Kennedy, D. M. (1989*). Classroom interactions of gifted and non-gifted fifth graders.* Unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Kenny, D.A., Archambault, F.X., & Hallmark, B.W. (1995). *The effect of group composition on gifted and non-gifted elementary groups in cooperative learning groups*. (95116). Storrs, CT: National Research Center on Gifted and Talented.

Kennedy, D. M. (1995). Teaching gifted in regular classrooms: Plain talk about creating a gifted friendly classroom. *Roeper Review,* 17, 232-234.

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*, 842-858.

Kleinbaum, D. G., Kupper, L. L., Muller, K. E., & Nizam, A. (1998). *Applied regression analysis and other multivariable methods.* Pacific Grove, CA: Brooks.

Kulik, C.-L. C., & Kulik, J. A. (1984). *Effects of ability grouping on elementary school pupils: A meta-analysis.* Paper presented at the annual meeting of the American Psychological Association, Toronto. (ERIC Document Reproduction Service No. ED 255 329)

Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives.* Storrs, CT: The National Research Center on the Gifted and Talented.

Kulik, J. A., & Kulik, C-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36,* 73-77.

Leonard, J. (2001). How group composition influenced the achievement of sixth-grade mathematics students. *Mathematical Thinking and Learning, 3*(2), 175-200.

Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198-1202.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90*, 1112-1121.

Lou, Y., Abrami, P., Spence, J., Poulsen, C., Chambers, B. & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423-458.

Lou, Y. Abrami, P., & Spence, J. (2000). Effects of within-class grouping on student achievement: An exploratory model. *Journal of Educational Research, 94*(2), 101-112.

Loveless, T. (1998). The tracking and ability grouping debate. Retrieved online (11/01/08) from: http://www.edexcellence.net/foundation/publication /publication.cfm?id=127.

Mastropieri, M.A., & Scruggs, T.E. (2000). *The inclusive classroom: Strategies for effective instruction.* Upper Saddle River, NJ: Prentice Hall.

McCoach, D.B., & Siegle, D. (2003). Factors that differentiate underachieving gifted students from high-achieving gifted students. *Gifted Child Quarterly*, 47, 144-154.

McDermott, R. (1976). Kids make sense: An ethnographic account of the interactional management of success and failure in one first-grade classroom. (Unpublished Doctoral Dissertation, Stanford University).

McKnight, P.E., McKnight, K.M., Sidani, S., & Figueredo, A.J. (2007). *Missing data.* New York, NY: The Guildford Press.

McSherry, K. & Ollerton, M. (2002, January). Grouping patterns in primary schools. *Mathematics in School, 31*(1), 2-7.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority Children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly,47,* 155-160.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

Nomi, T. (2010). The effects of within-class ability grouping on academic achievement in early elementary years. *Journal of Research on Educational Effectiveness, 3*(1), 56 – 92.

Northwest Education Association (2004). *Reliability and validity estimates: NWEA achievement level tests and measures of academic progress*. Lake Oswego, OR: Northwest Education Association.

Northwest Education Association. (2005). *NWEA program manual*. Lake Oswego, OR: Author.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Ory, M.G., Jordan, P.J. & Bazzarre, T. (2002). The Behavior Change Consortium: setting the stage for a new century of health behavior-change research. Health Education Research, 17(5), 500-511.

Patton, M. Q. (2002). *Qualitative research & evaluation methods (3rd ed.)*. Thousand Oaks, CA: Sage Publications.

Peklaj, C., & Vodopivec, B. (1999). Effects of cooperative versus individualistic learning on cognitive, affective, metacognitive and social processes in students. *European Journal of Psychology of Education, 14*(3), 359-373.

Plata, M., Masters, W., & Trusty, J. (1999). Teacher's perception and nomination of fifth-grade Hispanic and Anglo students. *Journal of Research and Development in Education, 32*, 113-123.

Plucker, J. A., & McIntire, J. (1996). Academic survivability in high-potential, middle school students. *Gifted Child Quarterly*, 40(1), 7-14.

Poole, D. (2008). Interactional differentiation in the mixed-ability group: A situated view of two struggling readers. *Reading Research Quarterly, 43*(3), 228-251.

Prais, S.J. (1998). Raising schooling attainments by grouping pupils *within* each class. *National Institute Economic Review*, *165*, 83–88.

Prais, S.J. (1999). Within-class grouping: A rejoinder. *National Institute Economic Review*, *169*, 109–110.

Purcell, J. (1994). *The status of programs for high ability students* (CRS94306).Storrs, CT: University of Connecticut, The National Research Center on the Gifted and Talented.

Raudenbush, S.W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501-25.

Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models (Second Edition).* Thousand Oaks: Sage Publications.

Reis, S.M., McCoach, D.B., Coyne, M., Schreiber, F.J., Eckert, R.D., & Gubbins, E.J. (2007). Using planned enrichment strategies with direct instruction to improve reading fluency, comprehension, and attitude toward reading: An evidence−based study. *The Elementary School Journal, 108*(1), 3-23.

Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, *60*(3), 180-184, 261.

Renzulli, J. S. (2005, May). A quiet crisis is clouding the future of R & D. *Education Week, 24(38),* 32-33. 40.

Renzulli, J. (2005). A practical system for identifying gifted and talented students. Retrieved November 4, 2008, from http://www.gifted.uconn.edu/sem/semart04.html

Richard B. Russell National School Lunch Act §2507 (2004).

Robinson, A., Shore, B. M., & Enerson, D. L. (2007). *Best practices in gifted education.* Waco, TX: Prufrock Press.

Rogers, T. (1987). Exploring the socio-cognitive perspective on the interpretive process of junior high school students. *English Quarterly, 20*(3), 218-230.

Rogers, K. (1991). *The relationship of grouping practices to the education of the gifted and talented learner*. Storrs, CT: The National Research Center on the Gifted and Talented.

Ruthven, K. (1987). Ability stereotyping in mathematics. *Educational Studies in Mathematics 18*(3), 243-253.

Sayler, M.F., & Feldhusen, J.F. (1990). *Special classes for academically gifted youth.* Roeper Review, *12*(4), 244-256.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.

Slavin, R.E. (1987). Ability grouping: A best-evidence synthesis. *Review of Educational Research, 57*, 293-336.

Slavin, R.E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60,* 471-499.

Smith, J., LaRose, B., & Clasen, R. R. (1991). Underrepresentation of minority students in gifted programs: Yes! It matters! *Gifted Child Quarterly, 35*, 81-83.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Stambaugh, T. (2007). Next steps: An impetus for future directions in research, policy, and practice for low-income promising learners. In J. V. Tassel-Baska & T. Stambaugh (Eds.), *Overlooked gems: A national perspective on low-income promising learners* (pp. 83-88). Washington, DC: National Association for Gifted Children.

State of the States (2005). State of the States. Washington, D.C.: National Association for Gifted Children, author.

Statistics, N. C. F. E. (2006). Public elementary and secondary students, staff, schools, and school districts: School year 2003-04. Retrieved December 1, 2008, from http://nces.ed.gov/fastfacts/display.asp?id=96

Stevens, R.J. & Slavin, R.E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal, 32*(2), 321-351.

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Grounded theory, procedures and techniques. Newbury Park, CA: Sage.

Tierney, R. J., & Rogers, T. (1989). Exploring the cognitive consequences of variations in the social fabric of classroom literacy events. In D. Bloomie (Ed.), *Classrooms and literacy* (Vol. 250-263). Norwood, NJ: Ablex.

Tieso, C. (2003). Ability grouping is not just tracking anymore. *Roeper Review, 26*(1), 29-36.

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, *29*(1), 60–89.

Trochim, W.M.K. (2006). *Quasi-experimental knowledge base.* San Francisco: Jossey-Bass, Inc.

Tu, X.M., Zhang, J., Kowalski, J., Shults, J., Feng, C., Sung, W., & Tang, W. (2007). Power analyses for longitudinal study designs with missing data. *Statistics in Medicine, 26*, 2958-2981.

U.S. Office for Civil Rights. (2002). *2002 Office for Civil Rights Elementary and Secondary School Survey Projections and Documentation.* Washington, DC: U.S. Department of Education, Office for Civil Rights.

U.S. Office for Civil Rights. (2004). *2004 Office for Civil Rights Elementary and Secondary School Survey Projections and Documentation.* Washington, DC: U.S. Department of Education, Office for Civil Rights.

U.S. Office for Civil Rights. (2006). *The Civil Rights Data Collection.* Washington, DC: U.S. Department of Education, Office for Civil Rights.

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students.* Thousand Oaks, CA: Sage Publications.

Vaughn, S., Moody, S.W., & Schumm, J.S. & Klingner (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly, 21*(1), 57-74.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10*, 395-426.

Wilkinson, I. A. G., & Fung, I. Y. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research, 37*, 425-447.

Wyner, J. S., Bridgeland, J. M., & John J. Diiulio, J. (n.d.). *Achievementrap: How America is failing millions of high-achieving students from low-income families*. Lansdowne, VA: Jack Kent Cooke Foundation.

Yoon, S., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*.

Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods, 41*(4), 1083-1094.

APPENDIX

APPENDIX

SEMI-STRUCTURED INTERVIEW PROTOCOL

Date:____________________________ Time: _______________-_______________

Name: _________________________ School: BP NW Grade taught: ______

Number of years involved in the cluster grouping program:

Cluster classroom taught: _____________________________

What gifted and talented training or coursework have you had? ______________________

____

Do you have Gifted and Talented Licensure? YES NO

Prior to the interview:

- Give a brief overview of the project.
- Remind interviewee that we are attempting to reconstruct what happened in the classrooms for the students who will graduate in the years 2014, 2015, 2016, and 2017 when they were in grades 3, 4, 5, and 6.

- Explain the importance of the person's insights into his/her classroom practices and how the cluster grouping program works.

- Tell interviewee that the interview will be brief – but that you want to hear all he/she has to say on the subject.

- Inform interviewee that the interview will be taped and transcribed; an opportunity will be provided for review of the transcription if desired; notes will be taken during the interview and the audio recorder may be turned off at any time by request of the interviewee.

- Assure the person that confidentiality will be maintained and that only the researchers will have access to the audio files.

As you know I am interested in finding out about your classroom practices and how the cluster grouping program operates here at ___________________. This interview will focus on the practices of classroom teachers that might have influenced student achievement as well as on the elements of successful programming for gifted and talented students. I am interested in these areas as related to all of your students, not just those identified as "above average" or "high achieving." Any questions before we begin?

Teachers only:

1. Tell me about the types of grouping that you used within your classroom?
2. In previous applications of the cluster grouping program, over the years that the students were in the program, more and more students were identified as high achieving or above average and fewer students were

identified as low or low average. Did you see this? Why do you or do you not think this was the case?

3. In what ways did you recognize talent in your classroom?
4. Did the cluster grouping program help you in your efforts to meet individual student's needs? If so how?
5. What strategies did you use within your classroom with regard to questioning and thinking?
6. In what ways did you provide students with challenges?
7. You taught the _______cluster. Describe your expectations for the students.
8. In what ways did you provide students with choice?
9. Describe the types of reading and written assignments you gave your students.
10. What curriculum modifications did you make in order to meet the needs of the high achievers in your room (compacting, tiered assignments, independent study, etc.)
11. In what ways did you modify your instruction and curriculum in order to meet the needs of individual students?
12. Did you use enrichment or interest centers in your classroom? Describe the centers and how they were used.
13. To what extent did you use seatwork in your classroom?
14. How would you describe the atmosphere in your classroom?

15. Describe the professional development experiences that you have had that have influenced what you do in your classroom? Specifically, which are related to cluster grouping and how?

Teachers and Administrators:

Consider the cluster grouping program and answer the following questions relative to your school.

16. How has leadership in the school supported (or not) the cluster grouping program?
17. What is the general atmosphere and environment of the school? Was there support for the program?
18. What communication existed regarding the cluster grouping program among teachers, administrators, students, and parents?
19. Do you feel that, overall, teachers in the program were flexible in matching curriculum and instruction to individual students' needs? Why or why not?
20. To what degree were students from underrepresented groups represented in this program?
21. How do you perceive the ability of students from underrepresented groups?
22. What factors influenced your perceptions?
23. Have these perceptions changed over time since the implementation of cluster grouping?

24. What else would you like to tell me about the cluster grouping and your role as _____________?

Thank you very much for taking the time to discuss you classroom practices/role as administrator and the cluster grouping program with me. As I stated earlier all of your comments are confidential and your name will not be used on any documents.

VITA

VITA

Jillian C. Gates

EDUCATION

2006-2011	Purdue University, Ph.D. Candidate in Gifted and Talented Development Department of Educational Studies Related Areas: Measurement, Reading
2005	University of Alaska, Anchorage. Master's of Education in Special Education Related Area: Gifted Education
1996	Trinity Western University. Bachelor's of Education in Elementary Education Related Areas: Geography, Psychology

PROFESSIONAL EXPERIENCE

Teaching

2009-Present	Teacher, Anchorage School District, Anchorage, AK Grade 4/5 Optional Program
2006-2007	Teacher, Lafayette School Corporation, Lafayette, IN Special education resource room K-5
2005-2006	Teacher, Benton Community School Corporation, Fowler, IN Special education self-contained classroom K-5
2001-2005	Teacher, Anchorage School District, Anchorage, AK Grade 3, 3/4, gifted cluster teacher
1998-1999	Teacher, Fairhill Elementary, Fairbanks, AK Grade 6, all subjects
1996-1998	Substitute Teacher, Surrey School District, Surrey, B.C. Canada K-12, music, art, psychology, physical education

Academic

2006-2007 Graduate Assistant
Gifted Education Resource Institute (GERI)
Purdue University, West Lafayette, IN

◊ **Teaching Assistant** – EDPS 430 Creating and Managing Learning Environments
◊ **Research Assistant** – *Longitudinal Study of Former GERI Kids* - Survey coordination, data collection, and analyses, *Mathematics Self-Efficacy* - Data collection and analyses.
◊ **Teacher** – Summer Residential (Grade 5/6) – Engineering and Mythbusters.2007-2008 Graduate Assistant
Gifted Education Resource Institute (GERI)
Purdue University, West Lafayette, IN
◊ **Research Assistant** - Project HOPE (Having Opportunities Promotes Excellence) – HOPE Nomination Form development, data collection, HOPE Scholar program coordination, research coordination.
◊ **Grant Writing PRF** – locate potential funding sources, assist in grant writing.
◊ **Teaching Assistant** - EDPS 540Y Introduction to Gifted Education. Wrote curriculum for online delivery and taught class.
◊ **Teacher** – Summer Residential (Grade 5/6) – Engineering and Mythbusters.

2008-2009 Graduate Assistant
Gifted Education Resource Institute (GERI)
Purdue University, West Lafayette, IN

◊ **Research Assistant** - Project HOPE (Having Opportunities Promotes Excellence) – HOPE Nomination Form development, data collection, HOPE Scholar program coordination, research coordination.
◊ **Grant Writing PRF** – locate potential funding sources, assist in grant writing.
◊ **Teacher** – Summer Residential (Grade 5/6) – Engineering and Mythbusters.

PROFESSIONAL AFFILIATION

2008-2009 Graduate Student Education Council (GSEC)
2007-2009 Graduate Organization of Educational Studies (GOEDS)
2007-Present American Educational Research Association
- Research on Giftedness and Talent SIG

2002-Present National Association of Gifted Children
- Research and Evaluation Network member
- Counseling and Guidance Network Chair-Elect

2000-Present National Education Association

SERVICE

National

2008-2010 National Association for Gifted Children Counseling and Guidance Network Convention Program Co-Chair

2009-2011 National Association for Gifted Children Counseling and Guidance Network Network Chair -Elect

2007- Present American Educational Research Association reviewer

2007-Present National Association for Gifted Children Graduate Student Committee member

2007-Present National Association for Gifted Children Research and Evaluation Network reviewer

2007-Present National Association for Gifted Children Counseling and Guidance Network reviewer

College

2008 Graduate Student Education Council – President

2008 Graduate Studies Leadership Team member

2008 Strategic Planning Taskforce – Discovery Team

2008 Teacher Education Council member

Department

2008 Graduate Organization of Educational Studies – Vice-president

PUBLICATIONS

Peters, S. J., & **Gates, J. C.** (2010). The *Teacher Observation Form*: Revisions and Updates. *Gifted Child Quarterly,* 54, 179-188.

Gates, J. C. (2010). Children with gifts and talents: Looking beyond traditional labels. *Roeper Review*, 32, 200-206.

Gates, J. (September, 2009). Mistaking giftedness for ADHD. *Twice-Exceptional Newsletter,* pp.8-10.

Gates, J. (2008) Bridging the gap for gifted students: Advocacy strategies to help get what your child needs. *Nautilus.*

Gates, J. (2007, Fall). Infusing thinking skills into the classroom. *Teaching for High Potential*.

Gates, J., Gentry, M., Mann, R., & Peterson, J. (2007, Fall). Project HOPE (Having Opportunities Promotes Excellence). *Gifted Children: An Electronic Journal of the AERA SIG Research on Giftedness and Talent.*

Gates, J. (2005). *Social, emotional, behavioral, and cognitive phenotypes of a student with ADHD and academic giftedness.* Unpublished master's thesis, University of Alaska, Anchorage, AK.

Peters, S. J., Gates, J.C., Gentry, M. (under review). Exploratory and confirmatory validation of the HOPE Scale: Instrumentation to identify low-income K-5 students.

Gates, J.C. & Pereira, N. P. (in progress). Perceived classroom management needs of pre-service teachers.

Evaluation Reports

Gates, J. & Gentry, M. (2008). DISCOVER! 2008 annual report.

Gates, J. & Gentry, M. (2007). DISCOVER! 2007 annual report.

Gates, J. & Moon, S. (2006). Science bound evaluation summer 2006: Summer camp component.

PRESENTATIONS

National and International Peer Refereed Presentations

Gates, J., & Pereira, N. (2009, November). Perceived needs of pre-service teachers regarding gifted learners:

Miller, R. & Gates, J. (2009, November). Out-of-school enrichment for high potential students: A blueprint for successful programming. National Association for Gifted Children Conference, St. Louis, MO.

Mann, R., Gentry, M., Gates, J., Yang, Y., Miller R., Pereira, N. (2009, August). Recognizing and nurturing giftedness among underserved populations. Biennial Conference on Gifted Children, Vancouver, BC.

Peters, S. J., Gates, J.C., Gentry, M., Peterson, J. S., & Mann, R. L. (2009, April). Exploratory and confirmatory validation of the HOPE Scale: Instrumentation to identify low-income K-5 students. Annual Meeting of the American Educational Research Association, San Diego, CA.

Gates, J.C. & Gentry, M. (2009, April). Empirical evidence to support the possibility of misdiagnosis of giftedness as Attention Deficit Hyperactive Disorder. Annual Meeting of the American Educational Research Association, San Diego, CA.

Gates, J.C. & Pereira, N. P. (2009, April). Perceived classroom management needs of pre-service teachers. Annual Meeting of the American Educational Research Association, San Diego, CA.

Gates, J.C. (2008, November). Further research support for the idea of misdiagnosis of giftedness as ADHD. National Association for Gifted Children Conference, Tampa, FL.

Gentry, M., Peters, S. J., & Gates, J. C. (2008, November). Initial validity evidence for the HOPE scale: An instrument designed to find talent among underserved populations. National Association for Gifted Children Conference, Tampa, FL.

Gates, J.C., Gentry, M., & Peterson, J. (2008, November). *Having opportunities promotes excellence: Project HOPE*. Poster session presented at the annual convention of the National Association of Gifted Children, Tampa, FL.

Gates, J.C. (2007, November). The possibility of misdiagnosing giftedness as Attention Deficit Hyperactivity Disorder (ADHD). National Association for Gifted Children Conference, Minneapolis, MN.

Peters, S., Gates, J.C., & MacDougall, J. (2007, November). Observing and evaluating teachers of the gifted: What's out there? National Association for Gifted Children Conference, Minneapolis, MN.

Jeffrey, T., Gates, J.C., Mann, R., & Gentry, M. (2007, November). Engineering and Socratic dialogue: Using the engineering design process in the elementary and middle school classroom. National Association for Gifted Children Annual Conference, Minneapolis, MN.

Gates, J.C. (2007, August). Gifted and/or ADHD? The possibility of misdiagnosis. World Council for Gifted and Talented Children, 17th Biennial World Conference, Warwick, UK.

Gates, J.C. (2007, August). Gifted and ADHD: The Pain and the Promise – A Teacher Workshop. World Council for Gifted and Talented Children, 17th Biennial World Conference, Warwick, UK.

Mann, R., & Gates, J.C. (2007, June). Engineering and Socratic thought in elementary classrooms. DISCOVER! 2007 Institute, West Lafayette, IN.

Gates, J.C. (2007, March). Gifted and ADHD: The pain and the promise. Indiana Association for the Gifted conference, Indianapolis, IN.

Other Presentations and Professional Development

Ongoing consultation Rochester Community School Corporation, Rochester, IN. January-Present (9 days in 2008).

"Meeting the needs of all learners: Differentiation 101." Session presented at the Metropolitan School District of Washington Township, Indianapolis, IN. February 8, 2008.

"Student-Based Differentiation." Session presented at Blue River Valley School Corporation, New Castle, IN. March 19, 2008.

"Renzulli Learning System." Session presented at Blue River Valley School Corporation, New Castle, IN. April 18, 2008.

"Differentiation: What is it all about?" Session presented at the Logansport Community School Corporation, Logansport, IN. December 12, 2007.

"Curriculum mapping and writing." Session presented at the Rochester Community School Corporation, Rochester, IN. December 17-19, 2007.

"Acceleration and compacting in the cluster grouped classroom." Session presented at the Blue River Valley Community School Corporation, New Castle, IN. November 29-30, 2007.

"Meeting the needs of all learners in the general education classroom." Session presented at the West Central Community School Corporation, Francesville, IN. November 17, 2007.

"Differentiation strategies 6-8." Session presented at the Rochester Community School Corporation, Rochester, IN. November 14, 2007.

"Differentiation strategies K-5." Session presented at the Rochester Community School Corporation, Rochester, IN. November 13, 2007.

"Differentiation for all students" Session presented at the Rochester Community School Corporation, Rochester, IN. November 6, 2007.

"Gifted and ADHD: The Pain and the Promise." Session presented at Purdue University undergraduate class, West Lafayette, IN. October 3, 2007.

"Differentiation within cluster grouped classrooms." Session presented at the Rochester Community School Corporation, Rochester, IN. September 20, 2007.

"Differentiation for elementary school classrooms." Session presented at the Rush County Schools, Rushville, IN. February 2007.

"Curriculum compacting and tiered questioning." Session presented at the Metropolitan School District of Wayne Township, Indianapolis, IN. November, 2006.

"Differentiation strategies and planning for teaching gifted students." Session presented at School Corporation of Southern Hancock County, New Palestine, IN. October, 2006.

GRANTS

2011	Mickelson Exxon-Mobil Teachers Academy All expenses paid national academy for teachers of mathematics and science.
2008	Purdue University College of Education Dean's Graduate Student Research Support Program Project Title: Hierarchical Investigation of Parent and Student Patterns of Overexcitabilities. Amount: $250
2007	Purdue University College of Education Dean's Graduate Student Research Support Program Project Title: Examination of the Purdue Teacher Observation Form as an Instrument for Assessment of Quality Teachers in the Gifted Classroom. Amount: $300.
2007	Purdue University College of Education Dean's Research Mentoring Fellowship Project Title: Possibility of Misdiagnosis of Children Who Experience Giftedness as Having Attention Deficit Hyperactivity Disorder (ADHD). Amount: $625.
2007	Purdue University College of Education Dean's Travel Grant World Council for Gifted and Talented Children, World Conference, Amount: $150.
2007	Purdue University College of Education Educational Studies Department Travel Grant World Council for Gifted and Talented Children, World Conference, Amount: $150.

2007	World Council for Gifted and Talented Children Bursary World Council for Gifted and Talented Children, World Conference, Amount: £500 sterling.
2007	College of Education Gifted Education Resource Institute Travel Grant National Association for Gifted Children Annual Conference Amount: $500.

AWARDS

2008	Second Place, Completed Research Project Division National Association of Gifted Children Conference Graduate Student Research Gala
2007	First Place, Completed Research Project Division National Association of Gifted Children Conference Graduate Student Research Gala
2007	Most Outstanding Paper National Association of Gifted Children Conference Graduate Student Research Gala (Conference registration)
2006	Andrews Fellowship, Purdue University Graduate School (4 years tuition and stipend).
2004	British Petroleum (BP) Teacher of Excellence Finalist