COOPERATIVE LEARNING IN THE SECONDARY MATHEMATICS CLASSROOM: A CRITICAL REVIEW OF THE LITERATURE

Kyle Chappell

A Project Submitted to the Faculty of

The Evergreen State College

In Partial Fulfillment of the Requirements

For the degree

Master in Teaching

2013

This Project for the Master in Teaching Degree

by

Kyle S. Chappell

has been approved for

The Evergreen State College

by

Lester Krupp, Member of the Faculty

June 2013

ACKNOWLEDGEMENTS

My deepest and sincerest thanks go to my parents, my sister and my family and friends.

To my faculty and my fellow future teachers I offer my deepest gratitude and respect.

Bhargavi – Thank you for your endless kindness, generosity and support.

Phil Z. – My mentor and teacher in many ways. I could not have been here without your help and guidance. Thank you.

Phil, Annette and Allison – You were a second family. Thank you so much for your acceptance and support.

ABSTRACT

No Child Left Behind has increased the prevalence of standardized testing in public schools. The requirement that teachers and administrators use research-based methods of teaching leaves teachers little room for experimentation in the classroom. Due to this fact, methods of teaching that have a positive impact on student achievement as measured by testing must have a strong history of research before they can be used in the classroom. Cooperative learning has the potential to have a significant positive impact on student achievement, but the methods and strategies that have the most positive effects are unclear. The purpose of this paper is to review the research addressing cooperative learning and its effect on student achievement in mathematics.

TABLE OF CONTENTS

TITLE PAGE	i
APPROVAL PAGE	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
CHAPTER 1: INTRODUCTION	1
Introduction	1
Rationale	2
Historical Background	4
Definitions	8
Limitations	11
Summary	12
CHAPTER 2: CRITICAL REVIEW OF THE LITERATURE	13
Cooperative Learning and Achievement	14
Cooperative Learning Focused on Help-Seeking and Help-Giving and Student Achievement	53
Cooperative Learning and Metacognitive Training and Student Achievement	69
CHAPTER 3: CONCLUSION	78
Introduction	78
Summary of Findings	79
Classroom Implications	89
Further Research	93
Conclusion	96
REFERENCES	98

CHAPTER 1: INTRODUCTION

Introduction

The use of cooperative learning is one of the "great success stories of social and educational psychology" (Johnson & Johnson, 2009, p. 365). The label of cooperative learning encompasses a wide range of models based on vastly different views of learning theory and student and teacher roles (Davidson & Kroll, 1991). Today, high-stakes, standardized testing and the focus on teacher and school accountability require teachers and administrators to seek out new ways of increasing student achievement. In mathematics, this improvement is sometimes sought through new, standardized curricula that are adopted school- or even district-wide (Remillard, 2005). Standardized curricula can have a serious impact on teacher practice, so knowledge of the different models of cooperative learning, of the underlying assumptions about student learning that influenced the development of those models, and of the effect on student learning that those models demonstrated, is beneficial to any practicing teacher.

Raising student test scores and choosing curriculum that has a positive impact on student achievement are of particular importance to teachers of secondary mathematics. Many states are not on track to meet the 100% student proficiency in mathematics that is required by No Child Left Behind (Caillier, 2007). This paper will explore many of the different models of cooperative learning and critically review the available research to determine what models of cooperative learning have the most positive impact on student achievement in the secondary mathematics classroom.

Rationale

The No Child Left Behind (NCLB) act requires that 100% of students attain proficiency in reading and mathematics by 2014 and that strategies used to improve student academic proficiency be grounded in scientifically based research (Shaul & Ganson, 2005). At this point, the goal of NCLB is not being met and, if this trend continues, the goal of 100% proficiency in mathematics and reading will not be met by 2014 (Caillier, 2007). For students to achieve proficiency, teachers must “provide sound pedagogical methods in the classroom” (Caillier, 2007, p. 593).

In this era of NCLB and high-stakes testing, teachers will be expected, as stated above, to provide sound pedagogical methods that are grounded in scientifically based research. For the mathematics teacher, finding methods that can improve the academic achievement of students will be particularly important as mathematics is a subject in which 100% of students are expected to be proficient. Understanding the effects of different pedagogical methods is also important to the mathematics teacher due to current trends in relation to curriculum.

Remillard (2005) argues that mathematics education is at an intersection of two distinct trends in relation to mathematics curriculum. One trend is the availability of new designs of curriculum material accompanied by research to support their use. These new curricula emphasize the mathematical thinking, reasoning and conceptual understanding that is the focus of the Curriculum and Evaluation Standards for School Mathematics published by the National Council for Teachers of Mathematics. As a result, the curricula require the teacher to assume a very different role in the classroom than was typical previously). The second trend, according to Remillard, is the regulation of teaching practices by the standardization of a single curriculum in a school district.

The need for scientifically researched methods of teaching combined with the variety of available curriculum and the drastic effect that curriculum choice can have on teacher practice necessitates that mathematics teachers know about scientifically proven methods that have a positive effect on student achievement and understand how to effectively implement them in their classroom. One method that has been heavily researched and successfully implemented over the last 60 years is cooperative learning (Johnson & Johnson, 2009). Despite compelling successes in the classroom, traditional methods still dominate in classroom practice, particularly at the secondary level (Staples, 2007). Staples further argues that part of the continued use of traditional methods can be attributed to teacher's lack of understanding of the nature of cooperative practices and the roles required for teachers and students in a cooperative classroom.

The purpose of this paper is to help teachers in understanding the research supporting cooperative learning in the mathematics classroom and, more specifically, the requirements for successful implementation of cooperative methods for improving student academic achievement. This paper will provide a critical review of the research on cooperative learning in the mathematics classroom and requirements for the successful implementation of cooperative practices to increase student academic achievement.

Historical Background

Cooperation is an integral part of the human species. Humans seek out opportunities to work with others in pursuit of a common goal (Johnson & Johnson, 1991). The roots of cooperative learning begin with this idea of instinctual mutual interdependence. The work of John Dewey emphasized the social aspect of learning and the importance of learning to live and act cooperatively (Dewey, 1938). To Dewey, students learn to act cooperatively by experiencing cooperative action. While cooperation and collaboration are a part of any social environment, structured cooperative learning, as practiced in classrooms today, was not widely accepted until the 1980's (Johnson & Johnson, 2009).

Though cooperative learning was not widely accepted until the 1980's, research on this practice began as early as 1898 (Johnson, Johnson, & Holubec, 1994). In the United States, the theoretical foundations of cooperative learning were laid in the 1940's and 1950's in the work of Kurt Lewin and Morton Deutsch. Lewin's work was based on the idea that a state of tension within an individual motivates that individual to accomplish desired goals (Johnson & Johnson, 1974). These goals can take the forms of cooperative goal interdependence, competitive goal interdependence and individualistic goals unrelated to anyone else. In the first two goal structures, goals are dependent on at least one other individual; to achieve the goal an individual must cooperate, or work with, someone else. In the latter, goals can be accomplished individually, without interaction or cooperation with others. Deutsch built on the work of Lewin while developing a field theory of cooperation and competition, defining a cooperative social situation as one where the goals of the individual are interrelated and mutually dependent on the goals of other members of the group. The result of this definition is that if any individual increases his or her chance of goal attainment, the chance of other individuals attaining their goal increases as well (Johnson & Johnson, 1974).

The effects of cooperation on cognitive development were studied by Piaget, Vygotsky and other constructivist and social-constructivist theorists. Piaget focused on the individual construction of knowledge in relation to cooperation, whereas Vygotsky focused explicitly on the necessity of social interaction to construct knowledge. Piaget believed that when individuals act on the environment, conflict occurs and the subsequent cognitive disequilibrium stimulates the development of new cognitive structures to adapt to the new information and re-acquire equilibrium. Piaget considered the social experience as a fundamental variable in development (Miller, 2011). Vygotsky, emphasizing the necessity of social interaction to construct knowledge, believed that cooperation was necessary to learn and solve problems (Miller, 2011).

The work of B. F. Skinner also dealt with the impact of cooperation on learning. In Skinner's Operant Conditioning theory, the occurrence of desired behaviors is increased through the use of positive or negative reinforcement (Miller, 2011). The social aspect is that a child is reinforced by another peer or an adult, increasing the occurrence of desired behavior. The social interaction is necessary for the behavior to become more common. The work of Albert Bandura and the social learning theorists also had an impact on the development of cooperative learning. Bandura's theory of imitation, where a peer models a behavior that a child imitates, requires social interaction to develop some complex behaviors through imitation (Miller, 2011). Bandura broadened his theory of imitation into observational learning. Observational learning is "a more general process of acquiring information from other people, books, and electronic media." (Miller, 2011, p. 235). Clearly, observational learning requires social interaction, at some level, to occur.

From the 1940s to the 1970s, cooperative learning was relatively unknown and unused in classroom practice. This was due, in part, to the prominent cultural belief in social Darwinism: the belief that interpersonal competition drove success (Johnson & Johnson, 2009). In the 1960s the prominent cultural paradigm shifted to a belief in rugged individualism: the view that isolated, individual learning fostered strength in students. This cultural belief continued to cause resistance in society towards cooperative learning (Johnson & Johnson, 2009).

The beginning of the modern use of cooperative learning can be traced to 1966, when a group of teachers at the University of Minnesota were trained in the effective instructional use of small groups (Johnson & Johnson, 2009). This training led to the formation of The Cooperative Learning Center, formed as a result of David and Roger Johnson's efforts. The Center's four goals were: to synthesize existing knowledge regarding cooperative, competitive, and individualistic learning efforts; to develop models concerning the nature of cooperation and the essential components of it; to conduct research in cooperation; and to create concrete strategies for implementing cooperative learning in the classroom (Johnson et al., 1994). The work of the Johnsons and The Cooperative Learning Center led to the development, in the 1970s, of Learning Together, a method of facilitating group interactions that focused on teaching effective group interaction and processes (Slavin et al., 1985).

Where the work of Johnson and Johnson focused on training students to be effective group members and on continuous evaluation of group functioning, other methods arose that emphasized the importance of other aspects of cooperative learning. The work of Robert Slavin in the 1970s and 1980s led to the development of Student Team Learning, which consists of Slavin's Student-Teams Achievement Divisions (STAD), DeVries' Teams-Games Tournament (TGT), and numerous modifications including Jigsaw II. The work of Slavin emphasized the importance of competition to increase academic achievement (Slavin et al., 1985).

Another contributor to the field of cooperative learning was Elliot Aronson who, in the 1990s, developed the Jigsaw method. The Jigsaw method was similar to Learning Together in that it emphasized the successful process and interaction of the group (Aronson, 1997).

Shlomo Sharan began his research in the 1970s but was particularly active in the 1980s and 1990s. His research led to the development of the Group Investigation model of cooperative learning. In Group Investigation, a class is broken up into small groups, but students are given much more responsibility for organizing and goal-setting of the group. Students are then given an open-ended investigation to complete, requiring a variety of resources and allowing for multiple sub-tasks to be assigned within the major task (Slavin et al., 1985).

Elizabeth Cohen was also active in the 1970s and 1980s. Her research into cooperative learning led to the development of Complex Instruction at Stanford University. Complex Instruction is a more complex model of cooperative learning focused on group process and with a more rigid structure of group roles. Complex Instruction emphasizes open-ended, investigational tasks (Cohen & Lotan, 1990).

Definitions

Throughout this paper there are a number of terms that are specific to the field of cooperative learning and to the effects studied in the paper. Defining these terms as they are intended is necessary to fully understand the research reviewed and the findings of the studies. This section lists the major terms that require definition and their meaning as used in this review of literature.

Student achievement, in this paper, will refer to increased scores on tests. This is mainly due to the bulk of research on student achievement measuring performance on tests. While this may not reflect student learning as a whole, it is a valuable indicator and is applicable to teachers looking to increase standardized test scores to meet requirements of NCLB.

Collaborative Learning, as used in this paper, will reflect Kenneth Bruffee's definition: "a form of indirect teaching in which the teacher sets the problem and organizes the students to work it out collaboratively" (Stewart, 1988, p. 59).

Cooperative Learning, in this paper, will reflect the following definition: "It is a pedagogy that generates a diversified body of methods of instruction which organize students 'to work in groups toward a common goal or outcome, or share a common problem or task in such a way that they can only succeed in completing the work through behavior that demonstrates interdependence, while holding individual contributions and efforts accountable'" (Sharan, 2010, p. 301).

Learning Together, in this paper, will specifically refer to the work of David and Roger Johnson. The work of the Johnsons was more conceptual in nature than some of the other methods described. In Learning Together, the Johnsons specified three types of cooperative learning groups: formal learning groups, which stay together until task is completed and are more structured; informal learning groups, which are short-term and less structured; and base groups, which are long-term and focused on peer support and long-term accountability (Johnson & Johnson, 1991). The Johnsons also specified five essentials for cooperative learning: (1) positive interdependence, the linking of goal structures so that individual goals are tied to group goals and success depends on all group members succeeding; (2) face-to-face promotive interaction, face-to-face interaction that promotes each student's learning and success; (3) individual accountability, the requirement that students are evaluated individually as well as within the group; (4) cooperative skill, which requires that students get to know and trust each other, communicate clearly, accept and support each other and resolve conflicts constructively; and (5) group processing, the discussion and evaluation of the achievement of group goals and effective group relationships (Johnson & Johnson, 1991). The Learning Together model is focused more on group process than many of the other models that will be examined.

Student Teams Achievement Divisions (STAD) is a more prescriptive model of cooperative learning than Learning Together. In STAD, students gather in groups of four to five students and work together to master a worksheet, or set of worksheets, on a given lesson. Each student then takes an individual quiz on the material, and the scores are adjusted to reflect the difference between the current quiz score and past averages. These adjusted scores are then used to determine the group score. The team with the highest score is then recognized in a class newsletter (Slavin et al., 1985). This model is more competitive than Learning Together and Jigsaw, and more prescriptive than Learning Together.

Teams-Games-Tournament (TGT) is another, more prescriptive, method of facilitating cooperative learning. TGT is similar to STAD but instead of individual quizzes, representative students from each group compete in academic games against students from other groups of similar achievement

(Slavin et al., 1985, p. 7). This model, again, emphasizes competition rather than group process.

Group Investigation is a more complex method of cooperative learning where students are in small groups of four or five. The groups are largely responsible for organizing themselves, delegating responsibilities and deciding what they will learn. The students are given an open-ended, investigational task that requires a variety of resources and inherently contains many subtasks within the larger task (Slavin et al., 1985).

Complex Instruction is similar to Group Investigation in the kinds of tasks that are assigned. Tasks are generally open-ended and investigative, located at Learning Centers that are set up around the classroom. Students are in heterogeneous groups of four to five and work together throughout a unit, generally one to two weeks. These groups are rotated so that, by the end of the year, all students have the opportunity to work with each other. Group roles are specifically assigned to students and have explicit responsibilities within the group. The tasks that are given to students are designed to require participation from all group members, due to the varying subtasks required within the larger task (Cohen & Lotan, 1990).

Limitations

To make this paper most relevant to teachers at the middle and secondary levels, the research reviewed in this paper will be confined to studies involving students in the sixth through twelfth grades. The research will also be confined to studies on mathematics classrooms, with a few exceptions when the content area researched was explicitly related, as in a physics classroom where work was mathematics-based. Also, the bulk of the research reviewed here focused on classrooms in the United States, though some studies in other countries will be reviewed. Though research has been done on the impact of cooperative learning on social interaction, increased positive attitudes towards content, and increased positive multicultural interactions, the research in this paper will be limited to studies focused on the impact of cooperative learning on student academic achievement.

Summary

By critically reviewing the available research on cooperative learning in the mathematics classroom, this paper will attempt to answer the question of what models of cooperative learning have the biggest impact on student achievement in mathematics. Using the findings of this review, secondary mathematics teachers will be able to make more informed decisions when choosing curricula for their classrooms and designing effective group tasks that are based on theories of learning that match the educational philosophies of the teachers themselves.

CHAPTER 2: CRITICAL REVIEW OF THE LITERATURE

Cooperative learning has been heavily researched over the past 30 years. While the majority of studies have found a positive effect on student achievement as a result of cooperative learning methods, there has been disparity among the results. One possible explanation for the disparity in the findings is the varying structure of cooperative learning tasks and environments. As discussed in the definitions in Chapter 1, the methods of Robert Slavin, including STAD, TGT and Jigsaw II, rely on inter-group competition to drive student achievement. Many other forms of cooperative learning, such as those devised by Robert and David Johnson, focus on teamwork and collaboration as opposed to competition. Slavin's focus in devising a cooperative learning strategy was to increase student achievement, while the purpose for Robert and David Johnson was to encourage collaboration and teamwork among students. This fundamental difference in underlying philosophy leads to varying levels of impact on achievement as a result of cooperative learning.

The purpose of this paper is to review the literature that has studied cooperative learning in general, specific features of cooperative learning, and effects on student achievement. The first section of this chapter will review literature that has studied cooperative learning in different forms and its effects on student achievement. The second section of the chapter will review studies of cooperative learning that focused on help-seeking and help-giving and the effects that these specific features had on student achievement. Finally, the third section will review studies that have focused on cooperative learning used in conjunction with metacognitive training and the effect that this additional training had on student achievement.

Cooperative Learning and Achievement

There is a wide body of research on cooperative learning and its effects on student achievement. The purpose of this section is to review studies on cooperative learning and its impact on student achievement in a mathematics classroom. The focus of this paper is on cooperative learning in the secondary classroom in the United States. Due to this, the bulk of the studies in this section were on students at the junior high and secondary level, but a couple of studies also involve students in elementary grades. The majority of the studies reviewed took place in the United States, but studies were also included that took place in China, Israel, England, Bermuda, Turkey and the Netherlands.

The studies reviewed in this section were focused on cooperative learning and its effect on student achievement. However, some studies also included additional, complementary instructional strategies with cooperative learning and evaluated the impact on student achievement. Additionally, some studies examined the effect of different grouping strategies on student achievement. As a result, the first subsection is composed of studies that focused on cooperative learning and its effect on student achievement. The second subsection is composed of studies that evaluated cooperative learning combined with additional instructional strategies and their impact on student achievement. Finally, the last subsection is composed of studies that examined the impact of different grouping strategies in cooperative learning and the subsequent effect on student achievement.

To assess the effect of cooperative learning on student achievement Johnson, Johnson and Scott (1978) performed a quantitative study on 30 sixth graders and found that students taught in a cooperative method performed better on tests when the tests were given cooperatively as well. Slavin and Karwheit (1985) performed a quantitative study on 9-12 year-olds and found that cooperative learning had a positive impact on student achievement. Sherman and Thomas (1986) performed a quantitative study on two general mathematics classes in a rural, Midwestern, predominately white town and found that cooperative learning had a more positive effect on student achievement in the classroom. Hawkins, Doueck, and Lishner (1988) performed a quantitative study on 1166 seventh-grade students in Seattle, Washington, and found that cooperative learning had no significant effect on student achievement. Henderson & Landesman (1995) performed a quantitative study at a largely Latino(a) middle school with 102 students during one year and 103 students during the subsequent year and found that cooperative learning in a thematically organized mathematics class had a positive effect on test scores that focused on concepts and applications. Vaughan (2002) performed a quantitative study on fifth-grade students in Bermuda and found that cooperative learning had a positive impact on student achievement. Harskamp and Ding (2006) performed a quantitative study on 99 students in Shanghai and found that collaborative learning had a positive effect on student test scores. Tarim & Akdeniz (2008) performed a quanititative study on 248 fourth grade students in Turkey and found that Team Assisted Individualization (TAI) and STAD both had a positive effect on student achievement, with TAI having a higher positive effect than STAD. Galton, Hargreaves and Pell (2009) performed a quantitative study on middle school students in the Netherlands and found larger effect on student achievement in cooperative groups than in whole-class teaching methods. Boaler (2001) performed a qualitative study at a large, urban high school in California and found that students performed at a higher level after the school instituted a curriculum that emphasized cooperative learning strategies.

In addition to investigating the effects of cooperative learning alone on student achievement, studies also investigated the effect of cooperative learning combined with other educational tools or techniques to evaluate the effect on student achievement. Slavin and Karweit (1984) performed a quantitative study on 1092 ninth-grade general mathematics students in Philadelphia to determine the effects of STAD, Mastery Learning and STAD combined with Mastery Learning and found that STAD had a significant effect on student test scores. Mevarech (1991) performed a quantitative study on 117 third grade students and found that cooperative learning combined with mastery learning had a significant, positive effect on student test scores. Brush (1997) performed a quantitative study on fifth graders in the Midwest and found that cooperative learning had a positive impact on student achievement when using technology in a mathematics class. Ke and Grabowski (2007) performed a quantitative study on 125 fifth graders in Pennsylvania and found that playing math games had a positive effect on student achievement but did not find any difference in achievement between cooperative and competitive game playing students. Souvignier & Kronenberger (2007) peformed a quantitative study on 208 third grade students and found that Jigsaw combined with questioning training had no significant effect on student test scores.

Student grouping was also a concern in cooperative learning and its effect on student achievement. To evaluate this impact, Linchevski & Kutscher (1998) performed two quantitative studies on 1629 Israeli students at the end of seventh grade and found that heterogeneous grouping had no negative effect on student achievement and that heterogeneous grouping had a positive impact on middle-achieving students. Also, Leonard (2001) performed a quantitative study on 177 sixth graders and found that lower-achieving students performed better on tests in a heterogeneous group setting. Whicker and Bohl (1997) performed a quantitative study on 31 students in a rural high school and found that the length of time that students worked in a consistent cooperative group had a positive impact on their achievement.

Cooperative Learning Alone and its Effect on Student Achievement

One of the early studies on cooperative learning and achievement was performed by Johnson, Johnson and Scott. Johnson, Johnson & Scott (1978), studied the effect of cooperative learning on student achievement in mathematics. The study was quantitative in nature and performed on a group 30 fifth- and sixth-grade students in an upper-middle-class, suburban school. There were 12 males and 18 females and there was no ethnic or racial profile given for the students.

The study was a pretest/posttest control-group design. The groups were randomly selected and the control group consisted of 11 females and three males. The experimental group consisted of seven females and nine males. The students in the control group were instructed to work individually, avoid interaction with other students, and seek help from the teacher. Students in the experimental (cooperative) group were assigned to heterogeneous groups with high-, middle-, and low-achieving students. The students in the cooperative group were instructed to complete worksheets together and to seek help from each other rather than from the teacher. The study lasted for 50 class sessions and covered units on set theory, number theory and geometry. The students were given a pretest on each unit prior to the beginning of the unit and a posttest after completion of the unit. No information was provided on the reliability of the tests.

Student achievement was determined by the test results and the daily work of the students and was analyzed with an ANOVA. The students in the cooperative group completed the set theory problems more accurately ($F = 9.10$, $p<.01$) and in a shorter time ($F = 5.25$, $p<.05$) than the students in the individualized condition. The analysis of the posttest for set theory showed that the individualized students performed somewhat better ($F = 3.79$, $p<.10$). Interestingly, when males and females were analyzed separately, females from the cooperative group performed better on the posttest while males from the individualized group performed better ($F = 2.85$, $p < .10$). In number theory, the students in the cooperative group were more accurate in their daily problems ($F = 10.54$, $p < .01$) and on the posttest ($F = 3.89$, $p < .05$). Males from the cooperative condition performed slightly better on the posttest than females, and females in the individualized group performed better than males ($F = 2.78$, $p < .10$). In geometry, the students in the cooperative group were far more accurate ($F = 54.74$, $p < .001$) and faster ($F = 11.52$, $p < .01$). Overall, students in the cooperative groups performed significantly better on the posttests than the individualized students ($F = 49.68$, $p < .001$). However, a retention test was given two months after the completion of the study and the individualized students performed significantly better than the cooperative students ($F = 4.92$, $p < .05$).

The results indicated that the students were more accurate and faster in their daily work when in a cooperative group. Further, the advantage increased as the difficulty increased. The results of the retention test indicated that the students in the individualized condition retained the material for a longer period of time. The authors explicitly stated that the results of male-female differences in the study need to be replicated before they could be taken seriously.

This study had some strong statistical data. Specifically, the p-values for accuracy in daily tasks were less than .01. These values supported the validity of the results and indicated that the difference in accuracy was not due to random chance.

The external validity of this study is somewhat weak. The sample was small (n = 30), and the ethnic/racial profile of the sample was not included. The students were all from an “upper-middle-class, suburban” school, indicating that the results could only be valid for other, upper-middle-class, suburban schools. Finally, there was no examination of the reliability of the instrumentation, which could have a significant effect on the results.

Robert Slavin and Nancy Karwheit, two early researchers in cooperative learning, performed a general study on the effect of cooperative learning in mathematics on student achievement. Slavin and Karwheit (1985) performed a quantitative study on students in grades four through six. The purpose of the study was to determine the effects on mathematics achievement of three specific programs: Team Assisted Individualization (TAI), a form of individualized instruction; Ability Grouped Active Teaching (AGAT), an in-class ability grouped cooperative learning strategy; and the Missouri Mathematics Program (MMP), a whole-class instructional strategy.

The study was conducted on 345 students in 15 fourth-, fifth- and sixth-grade classes in a single school district in Wilmington, Delaware. The classes were somewhat diverse: 71% of the students were white, 26% were African American and 3% were Asian American.

The design of the study was a pretest/posttest design with three groups (TAI, AGAT, and MMP). The students were given the California Achievement Test (CAT) as a pretest and the Comprehensive Test of Basic Skills (CTBS) as a posttest. Classes and their teachers were randomly assigned one of the three methods to serve as treatments. The experiment took place over the course of 18 weeks. During the experiment, implementation checks were performed to ensure that teachers were correctly implementing their assigned experimental method.

The results of the pre- and posttests were analyzed using random-effects nested analysis of variance (ANOVA). The results of the tests were adjusted so that the mean was 50 and the standard deviation was 10. The results for the TAI group for computations were 51.88 and 53.32 for the pre- and posttest. The results for the TAI group on concepts and applications were 54.03 and 52.83. The mean scores for the AGAT group on computations were 49.57 and 52.27 and on concepts and applications it was 48.88 and 49.60. The MMPT group scored 48.56 and 45.44 on computations and 47.02 and 47.66 on concepts and applications. After analyzing the results via ANOVA, the treatments showed no significant effect on student achievement (significance defined as $p<.01$).

This study had some significant weaknesses. There were no control groups, so comparisons were not available to test internal validity. Additionally, the course of the study was somewhat short (18 weeks). The pre- and posttests were also different, which may have influenced the results.

The study's main strengths were the size ($n = 345$) and the relatively high p-value ($p<.01$). The level of significance for the p-value was stated early in the study, and the results showed one p-value less than .018. The transparency of the researchers in what they felt demonstrated statistical significance indicates that the researchers adhered to their own standards during the study. The overall impact of these strengths and weaknesses indicates that the results of this study are somewhat generalizable to classrooms with a majority of white students in grades four through six.

Building on the work of Slavin and Karwheit, Sherman and Thomas (1986) performed a quantitative study to determine whether student achievement was higher in an individualistic or cooperative goal-structured classroom. The participants in the study were students in two general mathematics classes in a rural, Midwestern, predominately white town. The students were generally low achievers in ninth and tenth grades. There was an equal distribution of sexes and the median age was 15.

The design of the experiment was a two-group pretest/posttest design. The experimental group was taught using STAD and TGT methods in groups of four or five. The control group was taught through lectures, individual drills, and homework assignments. The pretest and posttest were the same and had a KR-20 coefficient of .86, which the researchers felt was “highly reliable.”

The scores from the pretest and posttest were analyzed using a MANOVA. The cooperative group had a pretest mean of 3.1 with a standard deviation of 2.75. The same group had a posttest mean of 19.85 with a standard deviation of 5.77. The control group had a pretest mean of 3.33 with a standard deviation of 3.65 and a posttest mean of 12.89 with a standard deviation of 5.96.

The MANOVA analysis yielded a score F (1, 36) = 18.62 which was highly significant ($p<.001$). This study supports the effectiveness of STAD and TGT models of cooperative learning in regards to student achievement.

The research design is quasi-experimental and has inherent threats to internal validity such as selection bias and its interaction with maturation and instrumentation (Sherman & Thomas, 1986). Additionally, the sample size of only two classes was small and the study is only generalizable to white, rural students due to the sample. However, the p-values ($p<.001$) were quite low and the instrumentation was constant for both the pre- and posttest. The tests also had a strong reliability coefficient. Overall, this study has limited generalizability, but for white rural students there is an indication that the results could be reliable.

Seeing some success with cooperative learning in mathematics in other research, Hawkins, Doueck and Lishner (1988) conducted a quantitative study on the effects of proactive classroom management, interactive teaching, and cooperative learning on academic achievement for 150 seventh-grade students in five Seattle middle schools. The students were selected as low-achievers due to low scores (23rd percentile) on the California Achievement Test (CAT) the prior year. The sample was 37.5% white, 53% male, and 48.1% from low-income families.

The experimental design was a static-group comparison. The treatment consisted of teacher-training in methods of proactive classroom management, interactive teaching, and cooperative learning. The cooperative learning techniques used in the study were STAD and TGT. The control group teachers received no training. The study lasted for one year. Teachers' adherence to the techniques was monitored over 3 day periods in the fall, winter and spring. The spring CAT scores from two years were compared; the year before the study and the year of the study.

The authors analyzed the CAT scores with an ANOVA. There was no significant difference between pretest and posttest in mathematics for the low-achieving students ($F = 21.62$, $p < .05$). The authors asserted that these findings replicated findings in another study (Natriello & Dornbusch, 1984) hat indicated that cooperative learning and other interventions have little effect on the achievement of low-achieving students.

This study was strong in its sample size and selection. It was a large sample ($n = 130$) and the selection was from a variety of schools in Seattle. The sample was diverse in race and sex. The length of the study was also a strength, as it was a year-long study. The authors also stated that their student achievement effect was not significant, even though the p-value was less than .05. This indicates that the authors had set a predetermined p-value higher than .05 and stuck with the predetermined value.

The study also had weaknesses in its sample. Because the sample was composed of all low-achieving students, the findings are only generalizable to low-achieving students. There is also significant question as to adherence of teachers to the experimental techniques. Teachers were only observed three times during the study, which makes the replication of this study difficult.

Henderson and Landesman (1995) used prior research in cooperative learning and devised a study to focus on the effect that cooperative learning would have on Latino(a) student achievement. Henderson and Landesman (1995) performed a quantitative study on middle school students to determine the effects of thematically integrated mathematics inistruction that relied on small, collaborative learning groups on student achievement. The study participants were from a middle school with a population that was 90% of Mexican descent. Sixty percent of these students were identified as limited English speakers. The study design was a pretest/posttest design. The tests were designed to measure computational skills, concepts, and applications. The alpha reliability coefficient of the tests was .84 to .86.

In the first year of the study, the students were randomly assigned to either the treatment group or a traditional control group, while in the second year administrators assigned students to one of the two groups. In the first year, the sample was 102 students. In the second year, another 103 students participated. The thematic treatment group was taught mathematics through subsequent themes. The themes were contextual topics explored in the mathematics. For example, the ocean and crime were two themes used during the study. Students would learn required mathematical content while exploring the thematic topic to provide context. During each theme they were taught through the use of small, collaborative learning groups. The traditional control group was taught in a traditional method of lecture and individual work.

The data from the two years was pooled and a MANOVA was used to analyze results. The treatment was found to have an effect ($p<.05$). The results were then analyzed with an ANCOVA to determine the effect of the treatment on posttest scores. The computation subset of the test showed no significant effect, but the mean scores on the concepts and applications subsets showed significant effect, $F(1, 74) = 5.15$, $p<.03$ and $F(1, 72) = 6.53$, $p<.01$ for Years 1 and 2 respectively. The study did not examine the performance of students randomly assigned to the experimental treatment versus those assigned administratively.

The results of this study indicate that cooperative learning has a positive effect on student test scores for concepts and applications. The sample size is moderate, and the subjects were of Mexican descent, leading to limited generalizability to populations that are of other ethnicities/races. The reliability coefficient of the tests is sufficient and the length of the study (2 years) leads to an overall strong study.

The success of cooperative learning in the United States led researchers to examine the effects of cooperative learning on the student achievement of students in other countries as well. Vaughan (2002) performed a quantitative study on 21 fifth-grade students in Bermuda. The study examined the effect of cooperative learning on the achievement of students of color outside of the

United States. This choice was due to the relatively high number of studies on cooperative learning and children in the United States and the lack of studies exploring the effects on students outside of the United States. The class was composed of 21 students, 10 boys and 11 girls, all students of color. Students were in a self-contained, fifth-grade classroom in Bermuda.

The study was designed in the single-group pretest/posttest method. The California Achievement Test (CAT) was used as both the pre- and posttest. Students were trained for one hour per day for five days on the expectations and procedures of Student Teams Achievement Divisions. The students were taught the skills necessary for cooperative learning: social skills, leadership skills, and basic group skills. After the one-week training, the students were taught mathematics through a STAD structure. Through the week, students were taught concepts in groups and on each Friday the groups were given individual quizzes. Additionally, students were given the CAT at the end of Weeks 5, 9, and 13.

The scores from the pre- and posttests were analyzed using a one-factor ANOVA for multiple measures, allowing for multiple posttests to be compared to one pretest. The CAT scores for computation and application were both analyzed. The mean scores for the pretest were 31.7 for computation and 27.4 for application, with a standard deviation of 8.1 and 8.2 respectively. The Week 5 posttest yielded scores of 31.8 for computation and 34.5 for application, with standard deviation of 9.7 and 6.9 respectively. The posttest for Week 9 had a mean score of 37 for computation and 36.5 for application, with standard deviation of 4.6 and 6.4 respectively. The final posttest had a mean score of 38.1

for computation and 39.3 for application, with a standard deviation of 5.1 and 6.4 respectively. These scores were analyzed using a one-factor ANOVA and the results yielded significant differences of $F(3, 60) = 7.509$ ($p<.0002$) for computation and $F(3, 60) = 26.06$ ($p<.0001$) for application. Vaughn concluded that these findings indicate that cooperative learning has a positive effect on the achievement of students of color.

While the p-values of this study showed strong statistical significance ($p<.0001$ and $p<.0002$), the sample size was very small (21 students), so the results do not necessarily have a high degree of generalizability. Additionally, the length of the study was somewhat short, only 12 weeks, which again leaves room for error due to random chance. Another significant weakness in this study was the lack of a control group.

However, as mentioned above, the p-values were very high. Statistically, these results are not likely to have occurred due to random chance. There were also multiple posttests given, allowing for more data to be tested. Finally, the same pre- and posttest was used in analysis, giving the study more internal validity. Overall, these results do show that cooperative learning is significantly correlated to student achievement, but the results are not necessarily generalizable to secondary classrooms or classrooms in the United States.

Harskamp and Ding (2006) performed a quantitative study on 99 students in a secondary physics classroom to attempt to determine how cooperative learning can be structured to improve students' problem-solving in physics. The participants of the study were 99 students at a secondary school in Shanghai. The class was composed of 54 females and 45 males. The design of the study was a pretest/posttest design. There were four conditions in the experiment: cooperative learning with hints provided, cooperative learning without hints provided, individual work with hints provided, and strictly individual work. The students in each class were randomly assigned to the four conditions and worked in pairs to solve multi-step physics problems. The study took place over one class period.

The posttests were analyzed using an ANCOVA. The results showed that cooperative learning with hints, cooperative learning without hints and individual learning with hints all had a positive effect on student achievement on the posttest ($F = 2.7$, $p = .048$). Only the individual condition did not have an effect on student achievement.

The sample size of this study was moderate, and the p-value was a little high ($p = .048$). The length of this study was quite short, occurring over a class period. This leads to little strength in the study and the researchers admit that more research needs to be done. However, it does appear that in this case cooperative learning and learning individually with hints led to greater student achievement on tests.

Tarim and Akdeniz (2008) performed a quantitative study to compare the effects of Team Assisted Individualization (TAI) and Student-Teams Achievement Divisions (STAD) on student achievement. The participants of the study were

248 fourth-grade students from a primary school in Turkey. There were eight classes of fourth graders in the school. The TAI and STAD groups were formed by four classes randomly selected from the total group of eight classes. The TAI group had 73 students and the STAD group had 71 students. The control group was comprised of the remaining four classes and had a total of 144 students. The samples were homogenous regarding academic achievement and family background. The achievement level at the school was above average.

The design of the study was a pretest/posttest design with a control group and two treatment groups. The three groups were taught in separate classes. The TAI and STAD groups received whole-class instruction, worked in small groups on worksheets and took individual tests to assess their readiness to progress further in the material. The control group was taught through lecture and individual work followed by tests to assess their readiness to progress further in the material.

To measure mathematics achievement, the researchers developed a test that was separated into two parts; achievement test A and achievement test B. The test was broken into two, 25 question parts because it was assumed that students could get bored answering all 50 questions at once, skewing the results. There was no difference between parts. The reliability coefficient was .84 for the first part and .81 for the second part. The study took place over 14 weeks. The pretest was given prior to the study and the posttest at the conclusion of the study. The results of the tests were analyzed using an ANCOVA. The results indicated that students in the TAI group performed significantly better than those in the control group (mean difference = 7.057, $p = 0.000$). Students in the STAD group also performed significantly better than the control group (mean difference = 2.714, $p = .018$). Also, there was a significant difference between TAI and STAD posttests in favor of TAI (mean difference = 4.343, $p = 0.000$).

This study has some significant strengths. The sample size is moderately large ($n = 248$) and the p-values are very low. However, the duration is only fourteen weeks, which makes the effect of the treatment somewhat questionable. Additionally, the students are in fourth grade in Turkey, so generalizability is limited in secondary schools in the United States. Overall, findings from this study indicate that cooperative learning in mathematics through the STAD and TAI methods was more effective than traditional methods, but generalizability is limited.

Galton, Hargreaves and Pell (2009), working in England, performed a quantitative study to assess the effects of cooperative learning on student achievement. The participants in the study were 12 mathematics classes of students aged 11 to 14 years. The specific number of participants was not given in the study. These classes were each taught by different teachers, and the study extended over the course of a school year. The study involved a mixture of English, science and mathematics lessons, but the results of the mathematics lessons were discussed in this paper.

The study was a variation of a pretest/posttest control-group design. Due to the reluctance of teachers to structure their classes using only individual seat-work, the researchers allowed teachers to use collaborative and cooperative learning groups for the experimental groups for all tasks, both low cognitive-demand and high cognitive-demand. The control group was changed from strictly individual seat work to whole-class discussion for high-cognitive-demand tasks and individual work with answer-checking allowed among students for low-cognitive-demand tasks. The essential difference was the use of small groups for all tasks in the experimental groups and whole-class work for the control groups.

Students were tested prior to each unit in the mathematics lessons and again at the conclusion of the unit. The same test was used as a pretest and posttest in all cases. For the number patterns in algebra, the test was composed of problems from the Diversity in Mathematics Education (DIME) resources. Other tests were created using the Secondary Mathematics Independent Learning Exercises (SMILE). Some questions were used from the Qualifications and Curriculum Authority (QCA) tests and were mainly focused on knowledge application. The DIME and SMILE test questions were generally higher in cognitive demand.

The teachers were trained in the use of cooperative learning strategies and were observed on a regular basis to ensure that they were using the methods during the teaching of their lessons.

The results of the pre- and posttests were separated into tasks of low-level cognitive demand and tasks of high-level cognitive demand, and then the results were analyzed. Due to students moving and changing classes throughout the year the sample populations changed in number resulting in the different n-values that will follow.

The experimental group mean for the pretest was 34.64 with a standard deviation of 4.37 while the posttest mean was 35.08 with a standard deviation of 5.64 (n = 167). The control group mean for the pretest was 30.35 with a standard deviation of 4.14 while the posttest mean was 30.53 with a standard deviation of 4.62 (n = 133). In these cases the experimental treatment did not have a statistically significant effect. However, in the high-level demand category, the experimental group had a pretest mean score of 33.54 with a standard deviation of 4.56 and a posttest score of 34.98 with a standard deviation of 5.22 (n = 185). The control group had a pretest mean score of 29.78 with a standard deviation of 3.78 and a posttest mean score of 30.45 with a standard deviation of 4.26 (n = 178). There was a statistically significant progress between the tests of high-level demand, and the effect size was larger (.7 vs. .48) for the small-group cooperative learning treatment ($p<.01$).

This study does appear to show some correlation between academic achievement and cooperative learning groups. However, there are some weaknesses in the study. First, the teachers were not randomly assigned to control or experimental groups, but rather opted to be a part of either the whole-class or small-group treatment. Additionally, the pretest did reveal some inconsistencies between the abilities of the whole-class and small-group samples. These inconsistencies could have led to the progress that was found to be statistically significant.

The study had some strength in that it was a long-term study. Also, the study had a large sample size. Overall, the results of this study could be generalized to certain populations, but not all. The inconsistencies revealed in the pretest lead to some serious doubts about internal validity.

In a large, comprehensive and qualitative study, Boaler (2006) examioned 3 large, urban high schools in the United States to determine how cooperative learning affected students. The study was a four-year longitudinal case study involving Railside High School, the experimental treatment school, and two other large, urban high schools. The students were racially and ethnically diverse at Railside: 38% were Latino/a, 23% African American, 20% white, 16% Asian American, and 3% from other groups. The researcher used classroom observations, questionnaires, assessments and interviews to collect data. Railside adopted Elizabeth Cohen's Complex Instruction method of cooperative learning in mathematics courses.

To measure the impact on academic achievement, the students all took an algebra test at the beginning of the study. Railside students began with relatively low test scores when compared with the other two schools ($t = -9.141$, $p < .001$, $n = 658$), but within two years they were achieving significantly higher levels ($t = -8.304$, $p < .001$, $n = 512$). Students were also given questionnaires on a yearly basis. The results of these questionnaires showed that by year 3 students looked more positively at mathematics in Railside than in the other schools (54 % enjoyed math 'all the time' at Railside versus 29 % at the other schools). Enrollment records showed that students at Railside took more mathematics courses. In Year 4, 41% of seniors were enrolled in calculus, compared with about 27% in the other two schools. Eight factors were found to be important to increasing equity and decreasing academic status differences.

- Multidimensionality: “Group worthy” problems were used whose solution required multiple abilities. In this way no one student had all the necessary abilities for a problem, and each student had at least one of these abilities.
- Specific Role Assignments: Students were assigned roles that changed frequently and had clearly assigned, useful tasks. Teachers frequently reminded and reinforced students about the expectations of their roles.
- Assigning Competence: This practice of raised the status of lower-status students through praise and acknowledgment to the group.
- Student Responsibility: Individual and group assessments were performed, creating individual accountability. Groups were given time to help each other further if an individual could not explain something.
- Justification and Reasoning: At Railside, students were required to justify their answers and explain their reasons to other students and to teachers. Right answers were not sought; justification and explanations of reasoning were required.
- High Expectations: Students were kept aware of high expectations. When groups could not figure out a problem, they were left to work on it.

Teachers consciously did not give small, leading questions to guide; they left the students to work problems out on their own.

- Effort over Ability: Teachers consistently made the point that success in mathematics is related to hard work and tenacity, not innate ability.
- Learning Practices: Teachers would point out effective learning practices that students were engaging in.
- Equity: At Railside, students learned through group problem-solving that all students have something to contribute, and all students have important thoughts and ideas. This awareness in math class led to more overall student equity. Also, as differences in achievement began to become more equalized, different groups were not as singled out

The length of this study and the transparent methods of data collection lend credibility to this study. The study has transferability to other diverse, urban schools.

Cooperative Learning Combined with Additional Instructional Strategies

The previous studies all focused on cooperative learning strategies and their effect on student achievement in mathematics. The following studies were focused on cooperative learning combined with additional, complementary instructional strategies and their effect on student achievement.

In an early study that combined cooperative learning with another educational tool to evaluate its effect on student achievement, Slavin and

Karwheit (1984) examined the effect of Student-Teams Achievement Division (STAD), mastery learning, and mastery learning combined with STAD to determine the effects of each of these variables on student achievement. The researchers studied 1092 ninth-grade mathematics students in inner-city Philadelphia. The sample was 76% African American, 19% White, 6% Latino/a, and 1% Asian American. Due to mortality in the study, only 588 (n=588) students had pre- and posttest data used in the study.

The experimental design was a four-group pretest/posttest model. All students were given an abridged version of the Mathematics Computations, Concepts and Applications section of the Comprehensive Test of Basic Skills, Level 2, Form S as a pretest. As a treatment, four different methods were used to teach the students. The mastery group was given a formative test, corrective instruction, and a summative test. The teams group worked in small groups with incentives. The mastery teams group combined these two methods; they worked in groups, took a formative quiz, had corrective instruction within teams, and took a summative quiz. Finally, the focused instruction group, which served as a control, was taught via lecture and given individual worksheets and a quiz. Following treatment, the students were given the same test as a posttest. The posttests were analyzed using a 2 X 2 nested ANCOVA with the pretest as a covariate.

The results of the tests were as follows. The mastery teams groups (n=125) had a pretest mean of 14.28 and standard deviation of 4.74. The posttest score was 17.11 with a standard deviation of 5.2. The teams group (n=138) had a pretest mean of 14.90 and standard deviation of 6.12. The posttest score was 17.54 with a standard deviation of 6.31. The mastery group (n=165) had a pretest mean of 14.50 and standard deviation of 5.57. The posttest score was 16.19 with a standard deviation of 5.83. The focused instruction group (n=160) had a pretest mean of 14.49 and standard deviation of 4.3. The posttest score was 16.10 with a standard deviation of 5.32.

The results of the analysis indicated that the classes that used teams (either alone or with mastery learning) achieved significantly higher scores than those that did not: $F(1, 39) = 5.18$, ($p<.05$). After performing an individual-level ANOVA, the result was highly significant: $F(1,543) = 10.40$, $p<.001$. This study supports the effectiveness of the teamwork and team-rewards structure of STAD in increasing student achievement.

This study was strong overall. The sample size was quite large and the instrumentation was the same for both the pre- and posttest. The p-values for the analysis performed at the classroom level were somewhat high ($p<.05$) and the p-values on the individual level were very high ($p<.001$). All of these factors indicate a strong study. However, generalizability is somewhat limited, in that the sample was largely African-American (76%). This indicates that the findings are most readily applicable to classes or schools with a high percentage of African-American students.

Another study that looked at cooperative learning and mastery learning was performed in 1991 and looked at the effect on the achievement of younger students. Mevarech (1991) performed a quantitative study on 117 third-grade students to determine the effect of cooperative learning on student achievement. The participants were 117 third-grade students in four classrooms. The study took place over the course of one quarter, and each of the classes learned from the same textbook and were taught the same content.

The design of the study was a pretest/posttest design. The students were given a pretest at the beginning of the quarter and a test at the end of the quarter. The four intact classrooms were randomly assigned to one of four treatment groups: cooperative learning (CL), mastery learning (ML), cooperative mastery learning (CML), and a traditional control group. The CL group ($n = 32$) was taught in small, heterogeneous groups of four. The teacher presented material to the whole class, and then small groups discussed and practiced mathematical concepts. The ML group ($n = 31$) was taught similarly, but the ML students worked individually on practice. At the end of each unit a test was administered and students that did not attain mastery were given corrective activities. The CML group ($n = 32$) was a combination of the above two methods; students worked in cooperative groups, took a unit test and were given corrective activites if they did not attain mastery. Finally, the traditional control group was taught with lecture and individual work followed by a unit test with no corrective activities for students that did not attain mastery.

The tests that were analyzed had a KR-21 reliability coefficient of .86 and .89 for the pre- and posttest, respectively. The analysis of the results of the test showed that the CML treatment had a significant effect ($F = 7.31$, $p<.05$). The analysis showed that the students in the CML group had higher mean scores and less variation in scores than the other groups.

The sample size of this study was moderate. The duration of the study, a full quarter, was a strength as well. The p-values were a little high ($p<.05$), but acceptable. The students were third graders, leading to some limited generalizability to secondary math classrooms, but the results still indicate that cooperative learning combined with mastery learning has a larger effect on student achievement than cooperative learning alone.

Another educational tool that was combined with cooperative learning was computers. Brush (1997) performed a quantitative study on the effect of cooperative learning with instruction delivered via Integrated Learning Systems on the academic gains of students. The participants of the study were 65 fifth-grade students in an elementary school in the upper Midwest. The students were from low- to lower-middle-class families. The students were 60% White, 30% African American and 10% categorized as *other*. The sample was 37% female and 63% male.

The study was designed as a posttest-only design. The control group worked with Individual Learning Systems individually while the experimental group worked with Individual Learning Systems in cooperative groups. Pretests showed similar scores between the groups, so only posttests were used to analyze differences. The posttest scores were analyzed using an ANOVA, and the mean scores for the cooperative group were found to be significantly higher than the scores for the individual group, $F(1, 62) = 4.53$, $p<.05$.

The sample size of this study was somewhat small, but acceptable for some validity. The p-value was also a little high ($p<.05$). The math instruction was delivered via Integrated Learning Systems, so the generalizability to a teacher-led classroom is limited. However, the results do indicate that cooperative learning has some positive impact on student test scores.

Along a similar vein as Brush (1997), Ke and Grabowski (2007) performed a quantitative study that evaluated the effect of Teams-Games-Tournament (TGT) on student achievement in mathematics during math games. The subjects were 125 fifth graders from six public schools in Pennsylvania. The sample was composed of 46% females and 54% males. In the subject group, 45% were economically disadvantaged and 8% were classified as minorities. The study did not provide a breakdown of the race and/or ethnicity of the minority students.

The study was a three-group pretest/posttest design. Students were given the Game Skills Arithmetic Test (GSAT), which had a K-R 20 reliability coefficient of .8, prior to the treatment. The students were then randomly assigned to three groups; TGT cooperative game-playing, interpersonal competitive game-playing, and no game-playing. The students in both game-playing groups were taught to play a web-based game designed to train the students in arithmetic. The students in the TGT group followed a traditional TGT structure of collaborative work and competitive tournaments. The students in the no-game-playing group worked individually on math drills. Following the treatment, the GSAT was administered to the students.

The mean scores on the pre- and posttest for the students in the group that had no game-playing were 61.2 and 58.2 respectively, with a standard deviation of approximately 15. The mean scores for the competitive game-playing students were 56.6 and 57.7 respectively, with a standard deviation of about 13.5. The mean scores of the TGT students were 56.7 and 59.5 respectively, with a standard deviation of about 14.4. The results were analyzed with a MANCOVA, and game-playing in general was found to have a significant effect compared to no game-playing ($F=3.81$, $p = .025$). However, there was no significant difference between competitive and cooperative game-playing ($p = .543$).

The authors found that, in the case of computer-based game-playing, cooperative learning had no significant effect on student achievement based on test scores. The p-value of .543 indicated that this was a reasonable deduction and that the null hypothesis should be assumed.

In terms of internal validity, the study was relatively short, at only four weeks. The external validity of this study is questionable, as the sample is predominately white, indicating that results can only be generalized to other predominately white classrooms. Additionally, the high p-value ($p=.543$) indicates that in game-playing, cooperative groups do not have an effect on student achievement compared to competitive groups.

Additional training in questioning was researched by Souvignier and Kronenberger (2007), who performed a quantitative study to determine the effects of cooperative methods on elementary school student achievement. The participants in the study were nine third-grade classes from three schools. A total of 208 students participated in the study. The design of the study was a pretest/posttest design with three conditions. The first condition was cooperative learning using the Jigsaw method. The second condition was Jigsaw with questioning training. The questioning training included instruction in the use of the five question stems from King's Guided Peer Questioning (1995). Essentially, students were taught to ask clarifying and explanatory questions of their classmates. The last treatment was traditional, teacher-guided instruction. The classes each studied three units on geometry.

The results of the posttest were analyzed and none of the treatments had any significant effect on student learning. The sample size of this study was relatively large, though the duration was not particularly long. The results could be generalized, with caution, to similar third-grade classrooms.

Grouping Strategies in Cooperative Learning and Their Effect on Student Achievement

Grouping was another aspect of cooperative learning that was researched. The question of whether students would benefit more from heterogeneous or homogeneous groups was research by Linchevski and Kutscher (1998), who performed two quantitative studies on Israeli junior high school students to determine the effects of ability grouping in cooperative learning groups on students' mathematics achievement.

The participants of the studies differed due to the different questions being asked. In the first study, the question was whether the achievement gap between students would change over time in a heterogeneous grouped setting. The unit of analysis for this question was a school and, as such, the sample of students was very large. The participants were 1629 students at 12 Israeli junior high schools at the end of seventh grade. The fact that 12 schools were used constituted independent replication of the study. Also, using multiple schools analyzed separately allowed for variability between schools to be seen and analyzed as well as variability at the student level.

The second study aimed to determine which of the two systems – heterogeneous or homogeneous classes – would lead to greater student achievement. Due to the scope of the question, the unit of analysis was a smaller group of 150 Israeli junior high school students from grade six through grade eight.

The first study was designed as a longitudinal, pretest/posttest design. Students were assigned different ability levels based on placement scores, and the students were taught in heterogeneous classrooms. The students and teachers were unaware of the ability designations. The students were taught for two years during the Together and APart (TAP) project. The major principle of the TAP project was to keep classes together as one learning unit. To respond to the different needs of students, TAP emphasizes cooperative learning in heterogeneous settings. The placement scores that were used to assign ability levels were also used as pretest scores. At the end of Years 1 and 2 (seventh and eighth grades), students were given achievement tests to serve as posttests.

The assumption of the researchers was that no gaps would be created that would not be expected due to differences in the students' original ability differences. The results of the regression showed no statistical effect of the grouping of students. They hypothesis of the researchers was shown to be correct. In essence, there was no negative effect of heterogeneous grouping on student achievement. In fact, two of the twelve schools showed a significant ($p<.05$) effect where higher-ability students gained more on average than lower-ability students. However, this result only occurred in two of the twelve schools.

This study had a large sample ($n = 1629$) which gave the statistics more validity. The sample was composed of students from different schools that were kept separate, creating more internal validity, as differences due to school could be determined. The main weakness of this study was the instrumentation. While the tests were validated by experts and the inspector for mathematics teaching in the Israeli Ministry of Education, the pretest and posttest differed. This could skew the results and the differences between pre- and posttest scores.

The second study compared the effects of heterogeneous and homogeneous groupings on the grades of students. The researchers conjectured that the grades of low- and middle-achieving students would be higher in heterogeneous, cooperative learning classes than the grades of low- and middle-achieving students in homogenous, individualized classes.

The participants were 150 eighth-grade Israeli students. The students were assigned to either homogeneous classrooms of low, medium or high ability or to heterogeneous classrooms. The teachers of the classes attended weekly workshops, some of which were taught by a TAP counselor and others by the school mathematics coordinator. Discussions and activities were structured differently according to whether the settings were mixed ability or homogeneous.

The classes remained together for two years and, at the end, were given a posttest. The scores of the tests for mixed ability and homogeneous classrooms did show a significant positive effect for the middle ability students ($p<.05$), but no statistically significant effect for any of the other groups.

The strengths of this study were a moderately large sample ($n = 150$) and a long duration (2 years). However, the p-values were relatively high ($p<.05$) and the tests were not the same for pre- and posttests. The high p-value gives more room for the results to be due to random chance rather than the to treatment, and the different instrumentation used for evaluation could be the cause of changes in student achievement. Overall, this study has some strengths, but results should be viewed with caution.

Continuing the research into grouping, Leonard (2001) performed a quantitative study on 177 sixth graders to determine how heterogeneous and homogeneous small groups differed in academic achievement. The setting of the study was an elementary school in suburban Maryland. The populations were fairly homogeneous in terms of socio-economic status, as most students came from working-class families. The study took place over two academic years with two separate cohorts of students. The first cohort was grouped heterogeneously by ability and the second cohort was grouped homogeneously. The racial and ethnic diversity for the first cohort was 78% white, 19% African American, and 3% Asian American with 57 girls and 38 boys. The second cohort was 70% white, 27% African American and 2% Asian American, with 40 girls and 55 boys. Due to changes of the sample population over the two-year study, the final results were analyzed for only those students retained throughout the full study. Therefore, while the total size of the two cohorts was 190, the study results only include scores for the 177 retained students.

The design of the study was a pretest/posttest design. The students were given the Maryland Functional Mathematics Test Level 1 (MFMT-I) in the fall as a pretest and the MFMT-II in the spring as a posttest. The KR 20 coefficient ranged from .59-.86 depending on the domain of the subtest. Pretest scores were used to group students according to ability and, in the first year, students were grouped heterogeneously in table clusters in the classroom. Students were grouped homogeneously in the second year. Teaching in the classrooms was heavily oriented toward small groups, and students worked collaboratively to accomplish complex tasks. At the end of the year the students took a posttest which served, with the pretest, as the data for analysis.

The tests were analyzed using an ANCOVA to compare the change in test scores by ability grouping and by cohort. The results showed significant effects for cohort, $F(1,153) = 6.083$, $p<.05$, and ability, $F(2,153) = 31.097$, $p<.05$. The scores of the high-ability students were relatively unaffected by the two cohorts, but the mean scores of the middle- and low-ability students showed significant improvement in the heterogeneous grouping cohort. The mean scores for the pretest and posttest for middle-achieving students were 307.26 to 341.81 for the heterogeneous cohort and 301.75 to 333.79 for the homogeneous cohort. The low-achieving students' mean scores were 290.43 to 323.60 for the heterogeneous cohort and 284.49 to 310.40 for the homogeneous cohort. In summary, the low-achieving students in the homogeneous setting scored significantly lower than their counterparts in the heterogeneous setting.

The results of this study are generalizable to a population of largely working-class students. The racial breakdown of the sample was predominately white, which also should be taken into account. The KR 20 coefficient is somewhat low for some subtests (.59) which makes the reliability of the instrumentation questionable. However, the sample was fairly large ($n = 177$) and the study took place over a long period of time (2 years). The results of this study are generalizable to populations of largely working-class, white students, but the internal validity of the study is questionable. While the results have validity, they are not particularly strong.

Whicker and Bhol (1997) performed a quantitative study on the effects of cooperative learning on the achievement of secondary mathematics students.

They focused on the length of time that a student spent with a group and whether that length of time would have an effect on student achievement. The sample consisted of 31 eleventh- and twelfth-grade students in a lower-middle-class area of the mid-South. The sample contained 18 males and 13 females in a precalculus course.

The study was a static-group comparison. Students were randomly assigned to the experimental (STAD) group and the control group. The STAD group contained ten males and five females while the control group contained eight males and eight females. The study lasted six weeks and consisted of units covering polynomial functions, inequalities and linear programming, and exponential and logarithmic functions. The STAD group was instructed on small-group work and told to seek help from one another rather than the teacher. The STAD group was taught the lesson, students studied in groups, and the teacher administered tests. The control group used the same process, but students worked individually. A test was given after each unit. The tests were teacher-created and the reliability was checked by having another teacher evaluate a sample of completed tests. The scores given by the two teachers were compared and were the same. This indicated that the scoring of the tests was reliable.

The researchers used the scores from the students' previous semester to analyze subsequent achievement. An independent t-test was used to compare prior semester scores and no significant difference was found between groups. Previous achievement and subsequent test scores were analyzed with a MANCOVA. The results showed a significant interaction between Group and

Time, $F(3, 86) = 29.41$, $p<.001$. The mean scores plotted over time indicated a slight advantage to students in the cooperative group. Analysis of the first, second, and third unit tests showed no difference in mean scores on the first test, no significant difference in mean scores on the second test, and significantly different mean scores on the third test, $F(1, 28) = 4.57$, $p = .04$. The explained variance attributed to prior scores and grouping changed from 96% attributed to prior scores on the first unit test to 93% attributed to grouping on the third unit test.

The study had some strengths. The sample had generalizability for classrooms in lower-middle-class areas with fairly even distributions of males and females. It showed a strong interaction between group and time, with a p-value less than .001. Also, the authors used another teacher to check reliability of the testing instrumentation. However, the sample size was quite small ($n = 31$). This could significantly affect results. Also, the racial/ethnic profile of the class was not provided, which decreases the generalizability of the findings. The researchers determined the slight advantage in mean scores between the STAD group and the control group to be significant, but the p-value of .04 is questionable, as one out of 25 times these results would be due to random chance.

The findings of the studies reviewed here showed that cooperative learning has an effect on student achievement, especially when using STAD, TGT or other, competitive goal structures. Slavin and Karwheit (1984) determined that classes using a form of STAD cooperative learning achieved significantly higher scores on tests. Sherwin and Thomas (1986) determined that classes composed of white, rural students using STAD and TGT scored higher on tests than students taught with traditional lecture and individual work. Whicker and Bhol (1997) determined that STAD cooperative learning had a positive effect on test scores of lower-middle-class, Southern students. These findings were replicated in other countries as well. Tarim and Akdeniz (2008) found that TAI and STAD both had a significant effect on fourth-grade student test scores in Turkey. Vaughan (2002) determined that STAD had a positive effect on student test scores of fifth-grade students in Bermuda and finally, Boaler (2006) determined that cooperative learning had a positive effect on student test scores, increased equity, and decreased status difference in a diverse, urban high school.

In contrast, Hawkins, Doueck and Lishner (1988) determined that STAD and TGT had no effect on the test scores of seventh-grade, low-achieving students in Seattle. Slavin and Karwheit (1985) found that neither TAI, AGAT nor MMP had a significant effect on the scores of students in fourth, fifth, and sixth grades in Delaware and Souvignier and Kronenberger (2007) found that Jigsaw had no significant effect on the test scores of third-grade students.

Additionally, the studies showed that cooperative learning has a positive effect on conceptually difficult and abstract mathematics. Johnson et al. (1978) determined that students were faster and more accurate in their daily work, and this advantage increased as the difficulty of the work increased. The students were upper-middle class. Galton et al. (2009) found that cooperative learning had no significant effect on low-level demand test questions, but did have a positive effect on high-level demand test questions. The participants were 11-14 year-olds in England.

The studies also showed that cooperative learning with technology can be effective, but more research is needed. Brush (1997) found that cooperative learning during instruction delivered via Integrated Learning Systems had a significant effect on the test scores of fifth-grade students in the Midwestern United States. In contrast, Ke and Grabowski (2007) found that cooperative learning had no significant effect on student achievement when students were taught using computer based game-playing.

Harskamp and Ding (2006) determined that cooperative learning with hints, cooperative learning without hints, and individual learning with hints were all positively correlated with student test scores. The students were secondary students in Shanghai. Mevarech (1991) found that cooperative learning combined with mastery learning had a significant effect on third-grade student test scores. Henderson and Landesman (1995) found that thematically integrated mathematics with cooperative learning had a significant effect on Latino/a student test scores on concepts and applications, but no effect on test scores on computations.

The studies determined that heterogeneous groupings are advantageous, or at least not disadvantageous, in cooperative learning. Linchevski and Kutscher (1998) found that heterogeneous grouping did not have a negative effect on seventh-grade Israeli students. Additionally, the results showed a positive effect on posttest scores for middle-ability eighth-grade Israeli students. Leonard (2001) found that heterogeneous grouping had a positive effect on test scores for middle-ability and low-ability sixth-grade, working-class students in Maryland.

Cooperative Learning Focused on Help-Seeking and Help-Giving and Achievement

The purpose of this section is to review studies that focused on help-seeking and help-giving behaviors in cooperative learning and the effect that these behaviors had on student achievement. The focus of this paper is on secondary students in the United States. Due to this focus, the studies reviewed were conducted on students at the junior high and secondary levels. The majority of the studies were conducted in the United States, though studies from Israel and the Netherlands are included as well.

Webb and Cullian (1983) performed a quantitative study on 105 students in Grades 7-9 and found that, for students, asking questions during group work and not receiving an answer had a significant negative correlation with achievement. Webb & Farivar (1994) performed a quantitative student on 166 seventh-grade students in six general mathematics classes and found that African American and Latino students' test scores were positively affected by helping behavior in one class, while in another class no significant effect was found. Leikin and Zaslavsky (1997) performed a qualitative study on four low-level, ninth-grade math classes in Israel and found that, though scores were similar on posttests, students were more active, gave more help to each other during class, and received more help in a cooperative learning environment.

Webb & Mastergeorge (2003) performed a quantitative study on classes in an urban US middle school and found that, with training, student requests for specific explanations during group work were found to be significantly correlated with achievement, and the number of requests for help was also found to have a significant effect on student achievement. Veenman, Denessen, van den Akker and van der Rijt (2005) performed a quantitative study on 48 sixth grade students and found that training in help-seeking and help-giving behavior had no significant effect on student test scores. Ross (1995) performed a qualitative study on 18 seventh-grade mathematics students in one class and found that teacher feedback strengthened the frequency and quality of help-seeking and help-giving. Dekker (2004) performed a quantitative study on 35 students ages 16-17 and found that teacher interventions that were aimed towards group process had a significant positive effect on student test scores. Finally, Esmonde (2009) performed a qualitative study on three high school mathematics classes at a diverse, urban high school and found that presentation activities, rather than group quizzes, caused an increase in helping behavior and collaborative interactions in a cooperative learning classroom.

One of the earliest studies in help-seeking and help-giving behavior was done by Webb and Cullian (1983). They performed a quantitative study on 105 students to determine the effect that helping behavior in small groups had on student achievement. The participants in this study were 105 students in seventh, eighth, and ninth grades. The students were from two average and two above-average general mathematics classes at a junior high school in Los

Angeles, California. Approximately 40% of the students were minorities. The average-level classes participated in a study on area and perimeter ($n=51$), and the above-average classes participated in a study on probability. The design of the studies was the same, but the content taught differed.

For one-week periods the classes attended a special classroom designed for interactive group work for one-week periods. A posttest was given and analysed for achievement. The posttest had an internal consistency alpha of .81. In addition to the posttests, observations were conducted to assess interaction among group members. The observations were recorded by two observers who underwent extensive training, and only one set of observations was chosen, at random, to be analyzed. After comparing the posttest score to the observations, the interaction variable that best predicted achievement was if a student asked a question and received no answer on a regular basis. The correlation was $r = -.41$, $p<.001$ in the classes of average students and $r = -.42$, $p<.001$ for the above average classes.

This study had credibility as the methods of data-collection and coding were clear and explicit. However, no triangulation was used. The internal validity of the instruments was high and the p-values for the correlation were low. The sample size was not ideal, but it was somewhat large. These conditions lead to this being a strong study. The implication is that engaging in non-helping behavior in a group is negatively correlated to student achievement.

In a study that focused on how to train students is helping behavior, Webb and Farivar (1994) performed a quantitative study on 166 students to determine whether training students in academic helping skills had an effect on student achievement. The participants of the study were 166 students in six seventh-grade general mathematics classes at an urban middle school in the United States. Two teachers each taught three classes. The classes had comparable student achievement levels and similar mixtures of students of differing gender and ethnic background. The sample was 55% Latino, 26% European American, 15% African American, 2% Asian American and 1% Middle Eastern.

The study was a pretest/posttest design with a control group. The study had two instructional conditions: (a) cooperative learning with instruction and practice in basic communications and helping skills (experimental condition), and (b) cooperative learning with instruction and practice in basic communication skills only. Two classes of each teacher were randomly assigned the experimental condition and the remaining classes were assigned the comparison condition. Students were given a pretest on general mathematics. The pretest had an internal consistency alpha of .67 to .73. The scores were then broken into high (top 25% of the sample), medium (middle 50%), and low (bottom 25% of the sample). Students were separated into small groups. Each group comprised one high, one low and two medium students. Classes were conducted with some whole-class, direct instruction followed by group work. The reward structure was 80% individual and 20% group. Quizzes were graded individually and classwork and homework were graded as a group, with one member's work being selected for scoring that the whole group would receive. The students worked in groups for four weeks, and the students were given a posttest at the end of the unit.

After comparing the verbal interactions of students and determining that the experimental group did, in fact, participate in more helping behavior, researchers analyzed the tests. European American, Asian American, and Middle Eastern students' test scores were not significantly affected. African American and Latino students' scores had a significant positive effect for one teacher ($p<.01$) and a positive effect that was not significant ($p<.06$) for the other teacher.

The sample of this study was relatively large and quite diverse. The p-value was low, indicating that the results were not due to random chance. Additionally, the internal reliability coefficient of the instrumentation was fairly high. However, the study was short term. Overall, this was a strong study. The researchers were quite transparent about differences in teacher interactions and instruction that could affect the results. The results could be generalized to other urban, diverse mathematics classrooms. The implications are that minority students could benefit academically from more cooperative learning opportunities.

In another study that continued in the same research vein, Webb and Mastergeorge (2003) performed a quantitative study on seventh-grade students to determine the effect of help-giving and help-receiving on student achievement. The participants were four seventh-grade classes in an urban middle school in

Los Angeles, California. The ethnic breakdown of the sample was 57% Latino, 24% European American, 12% African American, 3% Asian American and 4% other. The gender breakdown was 52% female, 48% male.

The design of the study was a pretest/posttest design. A cooperative learning program was instituted in four seventh-grade math classes. The program took place over a semester and students were instructed to work with the goal of helping each other learn how to solve the problems. It was conducted in three 4-week phases. A pretest and posttest were given to measure achievement. The pretest was also used in separating the class into heterogeneous groups based on ethnicity, gender and ability level. The pretest had an internal reliability coefficient of .81 and the posttest had an internal reliability of .89.

In addition to the pretest and posttest, group work was observed and various behaviors related to help-giving and -receiving were recorded and coded for analysis. The researchers reviewed their observations and focused the study on the students that had some problem requiring help from their classmates. This created a smaller sample of 48 students. The researchers were very transparent about coding and observation techniques. After analyzing the posttest scores in relation to observations of help-seeking and help-giving behaviors, the researchers found that asking for specific explanations was a behavior that was correlated to posttest scores ($r = .48$, $p<.001$).

The sample size in this study was small ($n = 48$), leading to limited external validity of the results. However, the length of the study was a semester, leading to higher internal validity of the effect of the treatment. The p-value of the correlation was quite low ($p<.001$), indicating that the results were most likely not the result of random chance. The observation and coding of the qualitative aspect of the study are transparent, giving the study credibility, though there was no triangulation. This study is a strong study and indicates that engaging in help-seeking behavior on the process of solving a problem has a positive effect on student achievement.

Another group of researchers that studied this same topic were Veenman, Denessen, van den Akker and van der Rijt (2005). These researchers performed a quantitative study on the effect of training in help-seeking and help-giving behavior on student achievement. The participants of the study were 24 pairs of sixth-grade students in the Netherlands. The design of the study was a pretest/posttest design with three conditions: cooperative learning accompanied by training in help seeking and help giving, cooperative learning alone, and no cooperative learning. The teachers were trained in help-seeking and help-giving behaviors and cooperative learning strategies. Teacher fidelity to the model was checked by observing the teachers in their classrooms on three occasions in an 11-month interval prior to the study. The study took place over the course of a year. Students were instructed on mathematics for one hour, three times each week.

The results of the pretest and posttest, in addition to analysis of student group interactions, were analyzed using an ANCOVA. The analysis revealed no significant differences in posttest scores between the treatment group and the control groups, but there was a moderate effect size (non-significant) in favor of the treatment group ($d = .76$).

This study had a moderate sample size and the duration was adequate. The observation methods and coding were transparent, giving credibility to the study. However, the sample was a group of students from the Netherlands, a sample that may not be transferrable to the United States.

To determine how cooperative learning led to help giving behavior, Leikin and Zaslavsky (1997) performed a qualitative study on the effect of cooperative, small-group settings on student activeness and help-giving behavior. While these are not direct indicators of student achievement, help-giving and help-receiving are some of the beneficial aspects of cooperative learning environments. Indeed, cooperation involves working together, so helping one another is a key element of cooperative learning.

The participants in the study were four low-level, ninth-grade classes in Israel. The design was a quasi-experimental design, with one class being taught in the experimental, cooperative style. Another class was taught traditionally, and the other two classes were taught in alternating experimental/control styles. The students were observed during group work at regular, two-minute intervals. Pearson's correlation coefficients were between .97 and .79 ($p<.01$), indicating that the study had credibility. The students were also given a questionnaire to record what kind of help they received from group mates during the study.

The observations and questionnaires were analyzed after the study concluded. To answer the question of student activeness, the researchers analyzed one group of four students. This group was from a class that alternated methods. The two most active, on-task interactions (giving an explanation; posing a question) had substantial increases (46.3-68.3 %) and, in general, passive behaviors decreased (53.6-31.6 %). However, one student showed no decrease in passive behaviors, which the researchers attributed to behavior problems.

To answer the question about help received, researchers analyzed the data from the self-reports and classified the data in two types:

a. Help was received: Student self-reports indicated that help was received 166 times (72% of the time) during group interactions.
b. Help was not received: Student self-reports indicated that help was not received (and not requested) 65 times (28% of the time) during group interactions

The majority (69 %) of help received by students in the cooperative learning setting was explanatory in nature as opposed to simply the correct answer.

An additional note in the study was included regarding student achievement. While it was not part of the study, it was noted that students performed similarly on tests in both methods. This indicates that student achievement was not positively affected by training in help-seeking and help-giving behavior.

The results of this study indicate that student activeness and on-task behaviors increased in the cooperative setting, while off-task behaviors decreased. Additionally, the help given in the cooperative learning setting was largely explanatory in nature. While these behaviors are positive, they did not have a positive affect on student achievement, as noted in the article.

The methods of observation give this study credibility, though the study is not necessarily transferable to all settings, as the student were Israeli and there was no demographic information given about the race or ethnicity of the students. Additionally, the instrumentation had a high reliability coefficient ($.79 < r < .97$), indicating that the self-report questionnaires were reliable for data gathering.

In an attempt to determine how teacher interventions affected helping behavior, Dekker (2004) performed a quantitative study on 35 students to determine if teacher interventions were more effective if they focused on group process or group product. The benefit of cooperative learning is most prevalent when students work together cooperatively. The purpose of this study was to evaluate whether students would work together more effectively when the teacher's interventions focused on their group process or on their product.

The participants of the study were 35 students (ages 16-17) from two classes in a high school. The students were used to working in small groups, and the teacher was used to teaching in a cooperative learning environment. The teacher was trained in the kinds of help he was to offer to students, which was strictly content-related. The group process help was provided by the researchers, and focused on encouraging students to show, explain, justify and reconstruct their work for one another.

The design of the experiment was a pretest/posttest design. The students were given a teacher-designed pretest on geometric transformation. The results of the pretest were used to divide the students into comparable subgroups that formed the process-help group and the product-help group. Following the pretest, students were given five days of 65-minute lessons on geometric transformation and then given a posttest at the end of the week.

The analysis of the tests revealed that students in the process-help condition raised their mathematical level more than students in the product-help condition ($p<.05$). Moreover, product-help tended to lead to divergence rather than convergence of students' gain scores, while process-help had the opposite effect. The mean scores of the product-help group ($n = 15$) on the pretest and posttest were 12.2 and 13.13 respectively. The mean scores of the process-help group ($n = 20$) were 11.9 and 14.45 respectively. While these findings appeared to indicate that process-help interventions had a positive effect on student achievement, the researchers admitted that more research is necessary before the results should be widely accepted.

The sample of this study is quite small (n = 35) and the p-values are somewhat high ($p<.05$). The small sample size could result in skewed statistics, and the high p-value could lead to a false assumption of significance. The duration of the study was also quite short, at only one week. While this leaves little room for maturation as an explanation for effect, it also brings into question the effects of the interventions, because the study covered such a short period of time. Overall, the findings of this study are suggestive, but the small scale and short duration do not lend themselves to a high degree of generalizability. More research is certainly required in this area.

Another study that examined the effects of teacher intervention on helping behavior was Ross (1995), who performed a qualitative study on the effects of teacher feedback on attitudes toward seeking and giving help. The study was performed on 18 Grade 7 math students from a single class. The students were taught over a 16-week period and recorded on four occasions while working in five cooperative groups. After the second and third recordings, the students were given edited transcripts of their discussions and trained in how to interpret them. Students used an instrument to assess their group process once or twice each week thereafter. The instrument used was a self-assessment worksheet that asked students to quantify the number of times that students in their group asked each other for help, gave each other help, and stayed on task. The students were given teacher feedback in the first weeks, but were trained through skits and demonstrations on how to use the worksheet to assess their group. Students were also given feedback from the teacher describing their group interactions and assessing their behaviors during group work.

The researchers used audiotapes to record group interactions. Audiotapes were analyzed and transcribed and two researchers listened to each tape. The coding methodology was described in detail and appears to be confirmable. Analysis of the tapes demonstrated that the teacher assessment and self-assessment increased the frequency and the quality of help-seeking and help-giving and improved students' attitudes about asking for help. This positive effect was attributed to three factors: the teacher feedback strengthened helpfulness norms built up in the classroom, increased students' skill in asking for and giving help, and the self-assessment enhanced student self-efficacy.

This study had credibility in the transparent data-collection and coding procedures. However, there was no evidence of triangulation. The study appears to be confirmable, as the methods are clearly outlined in the study. The results could be transferrable to other seventh-grade classrooms, but without demographic data, the full transferability of this study is questionable.

With a slightly different focus, Esmonde (2009) performed a qualitative study on three different mathematics classes all taught by the same teacher. The focus of this study was to see what activities led to more helping behavior in cooperative learning. The mathematics classes were all taught at a large, urban high school. The classes were diverse with respect to race, gender, prior achievement, and grade level.

The study was designed as an ethnographic study. The study was intended to provide rich, ethnographic data of cooperative group interactions and mathematics learning over the academic year. As such, the study took place over a year of instruction. Video recordings and ethnographic methods were used in addition to supplementary data in the form of student work, classroom artifacts, interviews, and questionnaires. During the study, two main activity structures were investigated. One was group quizzes. Quizzes were given to a group and students received two grades. One grade was for mathematical correctness. For this grade the teacher randomly selected one quiz from the group and the whole group received the grade of that quiz. The second grade was for group participation. There was a rubric posted, and students were evaluated on whether they included everyone in the group, worked together, asked questions, and explained their thinking. This activity had a group-reward structure and individual accountability, elements necessary for positive effects on student achievement according to Slavin (Slavin et al., 1985).

The second activity structure investigated was presentation preparation. Students were required to present problem solutions to their classes at least once each week. All students in a group had to be prepared to present and one group member would be chosen to present, sometimes by the group and sometimes by the teacher. There was no reward for this task, but students were still somewhat interdependent in that they had to work together and all be prepared to present.

Samples of group quizzes and presentations were selected. Quality of video-recording was a factor in deciding what to use and due to this, one quiz from the first-period class, one from sixth period, and four quizzes from fourth period were used, as fourth period had the best footage. A similar sample was taken from presentation preparation.

The interaction of the groups during group quizzes took three main forms: helping, collaborating, and individualistic. The groups that interacted in a helping way were found to be inequitable. In some cases, an expert student did all the work and told the other students what to write down with no conceptual explanation. In another case, there was no expert student, so the group was stuck and simply waited for the teacher to explain. The collaborating groups were more equitable; all group members contributed, though at times conceptual explanation was not offered. However, all group members were expected to stay togeteher as they worked through the problems on the quiz.

In summary, the group quiz had some strengths and weaknesses. The task did encourage students to work together and to persevere. However, mathematical discussions focused on correct answers, not on conceptual understanding, and multiple strategies for solving problems were not encouraged. Therefore, some students were silenced.

The interaction of the groups took the same main three forms as the group-quiz interactions. In general, presentation preparation allowed for more interactive work styles to occur. More individualistic work occurred, though collaborating and helping styles were represented as well. The strength of this was that its structure encouraged multiple people to get involved. The weakness was that students were able to opt out of discussing their ideas, listening to peers or participating in writing solutions.

This study had some aspects of credibility; the researcher was transparent in his methods of data-collection and coding. There was not, however, any triangulation or member-checking. The results are transferable to similar urban populations of students with diverse ethnic and socioeconomic backgrounds. The study was generally strong and transferrable to other high school math classes in the United States.

These studies showed that cooperative learning can be affected by group activities and dynamics. Dekker (2004) found that cooperative learning accompanied by teacher interventions in group process had a significant effect on 16- and 17-year-old students' test scores as opposed to teacher interventions in product. Leikin and Zaslavsky (1997) found that student activeness and on-task behavior of low-level, ninth-grade Israeli students increased during cooperative learning. Esmonde (2009) determined that presentation preparation allowed for more interactive work and helping behavior than group quizzes for students in a large, diverse, urban high school in the United States. Webb and Farivar (1994) determined that training in helping behavior while engaging in cooperative learning had a significant effect on some African American and Latino students but no effect on other students in urban, seventh-grade classes in the United States. Ross (1995) determined that teacher feedback on help-seeking and help-giving during cooperative learning had a positive effect on student attitudes towards help-seeking and help-giving. Webb and Cullian (1983) found that average and above-average seventh-, eighth-, and ninth-grade students in California performed worse on a test if they asked questions and received no help. Webb and Mastergeorge (2003) found that asking for specific explanations was a help-seeking behavior that was positively correlated to test scores of diverse, seventh-grade, urban students in California. Veenman et al. (2005) found that there was a non-significant positive effect on the test scores of students from the Netherlands that were trained in help-seeking and help-giving during cooperative learning.

Cooperative Learning with Metacognitive Training and Achievement

The purpose of this section is to review studies that focused on the effects on student achievement of cooperative learning conducted along with training in metacognitive thinking. The focus of this paper is on secondary students in the United States, but due to lack of research in the United States, the studies in this section were conducted on secondary students in Israel and Australia.

Mevarech and Kramarski (1997) performed a quantitative study on 247 seventh-grade students in Israel and found that the IMPROVE method, which uses metacognitive questioning and cooperative learning, had a significant positive effect on student achievement. Kramarski, Mevarech and Arami (2002) performed a quantitative study on 91 seventh graders in Israel and found that metacognitive training improved student achievement in cooperative groups.

Kramarski and Mevarech (2003) performed a quantitative study on 384 students in Israel and found that cooperative learning combined with metacognitive training had a significant positive effect on student achievement. Goos, Galbraith and Renshaw (2002) performed a qualitative study on senior secondary students in Australia and found that metacognitive, transactive comments on problem-solving were helpful for students.

The first large-scale study focused on metacognitive training combined with cooperative learning was Mevarech and Kramarski (1997). This was a quantitative study to determine the effects of IMPROVE on students' mathematics achievement. The IMPROVE is based on cooperative learning that includes metacognitive questioning about comprehension and strategy. The participants of the study were 247 seventh-grade students in four junior high schools in Israel. Three classes (n=99) used IMPROVE and five classes (n=148) served as control groups. Classes were similar in mathematics level and socioeconomic status.

The study used a pretest/posttest design. A 36-item pretest was used to assess students' basic skills and reasoning. Twenty-five items were multiple choice designed to assess basic skills, and 11 items were designed to assess reasoning. The KR reliability coefficient for the assessment was .91. All classes used the same mathematics book and studied mathematics five times per week. The treatment group used the IMPROVE method of instruction, and these students read problems aloud and solved them by using the metacognitive questions. If there was no consensus, the group discussed the disagreement until resolved. Finally, IMPROVE used feedback-corrective-enrichment, which consisted of unit tests and corrective activities for students that did not attain mastery (80% correct) on these tests. Those that did attain mastery were given enrichment activities. The control group received traditional instruction based on lecture and independent work. Unit test scores were analyzed as a posttest.

To determine the effect of the treatment on student achievement, the *treatment*, *prior ability* and *treatment + prior ability* interactions were all evaluated using an ANOVA. Treatment and prior ability were found to be significant ($p<.01$), but the interaction between them was not found to be significant ($p<.05$). The treatment condition alone was significant as well, $F(2, 12) = 154...85$, $p<.01$. At the end of one semester, IMPROVE students significantly outperformed traditional students on the posttest. The significant effect on achievement was found for each lower, middle and higher achieving students when compared with their counterparts in the non-treatment group.

The sample size of this study was relatively large, at 247 students. Additionally, the p-values were quite low ($p<.01$). However, the instruments were not evaluated with a reliability correlation, and different tests were used for the pre- and posttest. The students were also all from Israel, leading to limited generalizability to the United States. While this study does have valuable data regarding the effect of metacognitive training in cooperative learning on student achievement, more research is necessary.

A continuation of this research was Kramarski, Mevarech and Arami (2002), a quantitative study to determine the effect of cooperative learning and metacognitive training on students' ability to solve authentic and standard tasks. Additionally, the study sought to determine the effect of these two conditions on lower and higher achievers.

The participants of the study were 91 seventh graders in three heterogeneous classrooms in two junior high schools in Israel. Forty-four of the students were male and 47 were female. The schools were heterogeneous with respect to socioeconomic background and the classes were similar in size, mean age, and achievement.

The study was a pretest/posttest design. First, a pretest composed of authentic and standard tasks was administered. Then the groups were taught in the different methods for one year: one group with the cooperative learning and one group with cooperative learning and metacognitive training. The added metacognitive training was focused on training in questioning and answering related to comprehension, connection, strategy, and reflections on the task. A posttest composed of authentic and standard tasks was administered after the year.

Prior to the study, an ANOVA showed that only prior knowledge had any significant effect on scores ($p<.05$). To determine the effect of the training on student achievement, the researchers analyzed the posttest scores. The analysis of the posttest revealed a significant main effect for the treatment, $F(4, 84) = 5.01$, $p<.05$ and prior knowledge, $F(4, 84) = 5.01$, $p<.01$. The researchers then broke the scores down by ability. Both groups benefited from metacognitive training on standard tasks. An ANOVA showed that prior to the study, only prior knowledge had any effect on score ($p<.05$). After the study, both prior knowledge and the treatment had a significant effect ($p<.0001$). The interaction between prior knowledge and the treatment was not significant ($p<.01$). Both higher and lower achievers in the group with cooperative learning with metacognitive training outperformed their counterparts in the cooperative learning group, but the effect was larger on lower achievers.

The sample size of this study was moderate. Additionally, the p-values were high in relation to the main effect of the treatment for student achievement ($p<.05$). However, the p-value was quite low for the benefit of metacognitive training on standard tasks ($p<.0001$). Due to the fact that the study was conducted in Israel, the results are not necessarily generalizable to all secondary classrooms in the United States. While this study does present strong data in support of metacognitive training, more research is necessary.

Kramarski and Mevarech (2003) performed a quantitative study to compare the effects on mathematical reasoning of four approaches: (a) cooperative learning, (b) cooperative learning with metacognitive training, (c) individualized learning, and (d) individualized learning with metacognitive training. The participants of the study were 384 students (181 male, 203 female) who studied in 12 eighth-grade classrooms randomly selected from four junior high schools, which were randomly selected from a pool of 15 schools in a district with heterogeneous classrooms without ability groupings or tracking. The schools were similar in size, socioeconomic status, student mean age, and levels of mathematics achievement. Student mean age was 13.3 years. There were 12 teachers, all female, and all had at least five years of experience. The researchers then randomly assigned each of the four schools one of the four approaches. Each school had 3 teachers, and all of the teachers would teach using the same approach. The researchers felt that since the teachers worked collaboratively and shared materials, this design would be best for maintaining fidelity to the approaches in each school.

The design of the study was a pretest/posttest design. All classes studied a linear graph unit and all students were taught the same skills and used the same books and problems. The instructional method was the condition of the experiment, so it varied from group to group. Three measures were used to assess students' mathematical reasoning and metacognitive knowledge: a graph interpretation test, a graph construction test, and a metacognitive questionnaire. A pretest covering this material was administered at the beginning of the unit and a different posttest covering the material was administered at the end of the unit. The test scores were analyzed and a post-treatment MANCOVA was carried out on graph interpretation and graph construction because there was a significant correlation between them.

The scores from the graph interpretation test were analyzed with an ANOVA on the pretest scores and an ANCOVA on the posttest scores. The group that experienced cooperative learning with metacognitive training outperformed the group with individualized instruction with metacognitive training.

These groups, in turn, outperformed the two groups without metacognitive training. The differences were found to be significant, F (3,371) = 7.19, p<.01. When the differences between the two groups with metacognitive training were analyzed, no significant difference was found. The findings indicated that metacognitive training had a positive effect on mathematical reasoning.

The sample size of this study was fairly large, and the p-values for significance were quite low (p<.01). This indicates that the effect of the treatment was probably not due to random chance. However, the instrumentation was different between the pretest and posttest. Additionally, the sample was taken from a group of Israeli junior high schools, which may not be largely generalizable to the United States.

Goos, Galbraith and Renshaw (2002) performed a qualitative study to determine the interactions between peers of comparable expertise that led to successful or unsuccessful problem-solving outcomes. The participants in the study were students in senior secondary school classrooms in Queensland, Australia. Within these classes, target students were chosen for observation on the basis of their metacognitive sophistication and preference for working collaboratively with their peers. The study lasted for three years.

The study was a longitudinal case study. Portions of audio and videotapes were analyzed and the transcriptions were coded into *moves*. The moves were types of metacognitive acts where new information was recognized or an assessment of a particular aspect of a solution was made. The moves were also coded to identify their contribution to the collaborative structure of the interaction. Three transcripts were chosen for detailed analysis to show the common features of collaborative metacognitive activity.

The analysis of the transcripts, combined with analysis of student work, showed that metacognitive responses dealing with the solution of the problem were not particularly related to successful problem-solving. Transactive comments, those that introduced new ideas, were particularly useful for successful problem-solving.

This study was long-term and had a fairly large sample. The results are transferrable to other populations. The study has credibility as the methods and observations are transparent and clear.

These studies showed that metacognitive training can have an added effect on student test scores when combined with cooperative learning. Kramarski et al. (2002) found that cooperative learning with metacognitive training had a significant effect on student test scores as compared with cooperative learning alone with seventh-grade students in Israel. Mevarech and Kramarski (1997) found that seventh-grade students in Israel who participated in the IMPROVE program, which used extensive cooperative learning strategies and metacognitive training, had significantly better test scores than students taught with traditional lecture and individual learning. Kramarski and Mevarech (2003) determined that eighth-grade Israeli students who were taught with cooperative learning and metacognitive training outperformed students taught with metacognitive training and individualized instruction. Goos et al. (2002) found that metacognitive, transactive comments on problem-solving were helpful for students.

CHAPTER 3: CONCLUSION

Introduction

No Child Left Behind posed a daunting task to schools: 100% proficiency in reading and mathematics must be attained by the year 2014. While the goal is laudable, the reality is that America is far from achieving this benchmark (Caillier, 2007). To achieve this goal, No Child Left Behind requires that teachers and schools use methods backed by “scientifically based research” (Shaul & Ganson, 2005). Scientifically based research must show that student test scores improve when the methods and strategies studied are used. Due to this requirement, it is important and necessary that methods and strategies used in schools are heavily researched. Additionally, the methods and strategies used must be implemented with fidelity to how they were implemented in the studies. This requires that teachers and schools are aware of the specific aspects of teaching methods that show improvement in student test scores.

Cooperative learning has shown positive results in a number of areas including cognitive development, but the impact of cooperative learning on student achievement has been an area of debate. Part of this debate centers on the implementation of the strategy. Some cooperative learning strategies focus on the development of a highly functional group. Other strategies rely on competition to engage students. Depending on the philosophical framework of the implementer, cooperative learning can look very different in practice from classroom to classroom.

The purpose of this paper is to review the literature relating to cooperative learning and student achievement as measured by test scores. This section of the paper will provide a summary of the findings from Chapter 2, implications for classroom practice, and suggestions for further research.

Summary of Findings

Chapter 2 reviewed the research on the effect of cooperative learning on student achievement. The research was divided into sections reflecting the focus area of the research. The first section was Cooperative Learning and Achievement, which reviewed research that focused on various strategies of cooperative learning and its effect on the student achievement of various student groups. The second section was Cooperative Learning Focused on Help-Seeking and Help-Giving and Student Achievement. This section reviewed literature that studied the effects on student achievement of specific help-seeking and help-giving behaviors and training combined with cooperative learning. Finally, the last section was Cooperative Learning with Metacognitive Training and Student Achievement. This section reviewed research that focused on the effect on student achievement of metacognitive training combined with cooperative learning.

Cooperative Learning and Achievement

The findings of the studies showed that cooperative learning has an effect on student achievement, especially when using Student-Teams Achievement

Divisions (STAD), Teams-Games Tournament (TGT), or other competitive goal structures.

Slavin and Karweit (1984) determined that classes using a form of STAD-based cooperative learning achieved significantly higher scores on tests. This study was performed on a large sample of students (1092), and the results were highly significant ($p<.01$). The sample was composed of ninth-grade students in the United States and was largely African American (76%). Due to these facts, the results have generalizability to other secondary classrooms in the United States, particularly those with a large percentage of African American students.

A study that provided additional strength to the general finding of the positive effect of STAD and TGT cooperative learning methods was Sherman and Thomas (1986). This study determined that classes composed of white, rural students using STAD and TGT scored higher on tests than students taught with traditional lecture and individual work. This study had a relatively small sample size, with only two classes of students. However, the students were also secondary students in the ninth and tenth grades. The sample population was mainly low-achieving, white students in a rural Midwestern town. The instrumentation was found to be “highly reliable” and the results were highly significant ($p<.001$). The results indicated that STAD and TGT had a significant positive effect on student achievement. Though this study has a weakness in the small sample size, the high level of significance and the agreement with the results found in Slavin and Karweit (1984) are indicative of the strength of the finding.

In addition to the findings of Slavin and Karheit (1984) and Sherman and Thomas (1986), Whicker and Bhol (1997) also found that STAD had a significant positive effect on the test scores of lower-middle-class, Southern students. This study focused on thirty-one 11^{th}- and 12^{th}-grade students in a precalculus course. The sample size was rather small at only 31, and the six-week duration of the study was also quite short. Additionally, the method for checking the reliability of the instrumentation was questionable. However, the p-value for difference in student achievement was acceptable ($p=.04$) and the p-value for the significance of the effect of length of time with the group was very low ($p<.001$), indicating a very high level of significance. While this study was not particularly strong overall, it does agree with the findings of Slavin and Karweit (1984) and Sherman and Thomas (1986) and adds the additional finding of the impact of length of time with a group on the effect of STAD on student achievement.

Two more studies that provide additional support to these findings are Vaughan (2002) and Tarim and Akdentz (2008). Vaughan (2002) determined that STAD had a positive effect on student test scores of fifth-grade students in Bermuda. The sample was only 21 students, which is extremely small and makes the findings of the study very questionable. Additionally, this study was conducted on fifth-grade students in Bermuda, which creates limited generalizability to secondary mathematics students in the United States. However, the p-values were very low ($p<.0001$ and $p<.0002$), indicating a very low probability of effect due to random chance. The findings also agree with those of the previous authors, lending a limited degree of support to their findings.

Tarim and Akdentz (2008) found that Team Assisted Individualization (TAI) and STAD both had a significant effect on fourth-grade students' test scores in Turkey. This study had a good sample size at 248 students and the p-values were very low ($p<.018$ and $p<.0000$). However, the students were in the 4th grade in Turkey, which lends limited generalizability to secondary mathematics students in the United States.

While the above authors found a significant positive effect of STAD and TGT on student achievement, there were also some studies that did not find a similar effect. Hawkins et al. (1988) determined that STAD and TGT had no effect on the test scores of seventh-grade, low-achieving students in Seattle. This study had a long duration of one year and a moderately large sample size of 150 students. The subjects were low-achieving, low-income students, and the population had a high percentage of minorities. The findings indicated that STAD and TGT had no effect on the test scores of seventh-grade students. However, the fidelity of the participating teachers to the STAD and TGT model is questionable, as they were observed only three times over the course of the study. Regardless of this weakness, the findings indicate a need for further research. Additionally, Slavin and Karweit (1985) found that neither TAI, AGAT nor MMP had a significant effect on the scores of students in Grades 4-6 in Delaware. The study had a large sample size of 345 students and an adequate duration of 18 weeks. However, there was no control group to test for internal validity and the instrumentation differed between the pre- and posttests. While this study is not particularly strong, it does have significant strengths that call into question the positive effect of STAD and TGT on student achievement.

Another common thread was that cooperative learning with technology can be effective, but the findings were limited. Brush (1997) found that cooperative learning during instruction delivered via Integrated Learning Systems had a significant effect on the test scores of fifth-grade students in the midwestern United States. This study had a small sample (65) and was disproportionately male (60%), but the results had adequate significance ($p<.05$). However, Ke and Grabowski (2007) found that cooperative learning had no significant effect on student achievement when students were taught using computer-based game-playing.

Additionally, the studies showed that cooperative learning has a positive effect on conceptual, abstract mathematics as opposed to computational mathematics. Johnson, Johnson & Scott (1978)determined that students were faster and more accurate in their daily work, and this advantage increased as the difficulty of the work increased. The study did have some significant weaknesses in sample size and p-value, but the findings of Galton et al. (2009) added strength when they found that cooperative learning had no significant effect on low cogitive-demand test questions, but did have a positive effect on high cognitive-demand test questions. Henderson and Landesman (1995) had similar findings when they found that thematically integrated mathematics with cooperative learning had a significant effect on Latino/a students' test scores on concepts and applications, but no effect on test scores on computations. While the sample size was not particularly large, the duration was long (2 years) and the reliability of instrumentation was strong. The p-values were low and the findings indicate that cooperative learning can have a positive effect on the student achievement of Latino/a students on conceptual understanding of mathematics as well.

Another focus of the studies was group composition. The findings of the studies indicated that heterogeneous grouping is advantageous, or at least not disadvantageous, in cooperative learning. Linchevski and Kutscher (1998) found that heterogeneous grouping did not have a negative effect on seventh-grade Israeli students. Additionally, the results showed a positive effect on posttest scores for middle-ability eighth-grade Israeli students. This study was quite large (1629 students) and the duration was quite long (2 years), lending strength to the findings. Additionally, Leonard (2001) found that heterogeneous grouping had a positive effect on test scores for middle-ability and low-ability sixth-grade, working-class students in Maryland. While not as strong as Linchevski and Kutscher (1998), Leonard (2001) had a moderate sample size (171) and a long duration (2 years). The findings are also supported by Linchevski and Kutscher (1998), and the study was performed in the United States, adding more generalizability of the results to secondary mathematics students in the United States.

The rest of the studies had findings that provided some additional information about the effects of cooperative learning. Boaler (2006) determined that cooperative learning had a positive effect on student test scores, equity, and decreased status difference in a diverse, urban high school. Souvignier and Kronenberger (2007) found that Jigsaw had no significant effect on the test scores of third-grade students. Harskamp and Ding (2006) determined that cooperative learning with hints, cooperative learning without hints, and individual learning with hints were all positively correlated with student test scores. The students were secondary students in Shanghai. Mevarech (1991) found that cooperative learning combined with mastery learning had a significant effect on third-grade students' test scores.

The strongest findings of the research indicate that cooperative learning in the form of STAD and TGT has a positive effect on student achievement. Cooperative learning has more of a positive effect on conceptual mathematics than computational mathematics. Cooperative learning can have a positive effect when used with technology, but this finding is questionable. Finally, heterogeneous grouping has a positive effect or no effect on student achievement.

Cooperative Learning Focused on Help-Seeking and Help-Giving and Achievement

These studies showed that cooperative learning can be affected by group activities and dynamics. In a general sense, Leikin and Zaslavsky (1997) found that student activeness and on-task behavior of low-level, ninth-grade Israeli students increased during cooperative learning. However, rather than focusing solely on the effect of cooperative learning on student achievement, these studies focused on behaviors within the group that led to increased student achievement. For example, Webb and Cullian (1983) found that average and above-average seventh, eighth, and ninth graders in California performed worse on a test if they asked questions and received no help. This was a strong study with low p-values ($p<.001$) and an adequate sample size (105). Since this behavior had a negative impact on student achievement, behaviors that had a positive impact were researched.

Webb and Mastergeorge (2003) found that asking for specific explanations was a help-seeking behavior that was positively correlated to test scores of diverse, 7^{th} grade, urban students in California. While the sample size was small (48), the duration of this study was adequate (one semester) and the p-value ($p<.001$) indicated significant effect. Since this behavior was correlated to student achievement, methods for teaching and training in this behavior were researched.

Webb and Farivar (1994) determined that training in helping behavior while engaging in cooperative learning had a significant effect on some African American and Latino students but no effect on European American, Asian American, or Middle Eastern students in urban, seventh-grade classes in the United States. This study was short-term and the training was not very long-term, so the results are questionable. Additionally, Veenman et al. (2005) found that there was a non-significant positive effect on the test scores of students from the Netherlands that were trained in help-seeking and help-giving during cooperative learning. While this study was not the strongest, the results still indicate that training alone may not be adequate to increase helping behavior in cooperative learning groups.

The role of the teacher was another focus in the research. Dekker (2004) found that cooperative learning accompanied by teacher interventions in group process had a significant effect on 16- and 17-year-old students' test scores as opposed to teacher interventions in product. This study was small (35) and had a short duration (one week), but the findings are similar to those in other studies. For example, Ross (1995) determined that teacher feedback on help-seeking and help-giving during cooperative learning had a positive effect on student attitudes towards help-seeking and help-giving, and Esmonde (2009) determined that presentation preparation allowed for more interactive work and helping behavior than group quizzes for students in a large, diverse, urban high school in the United States. Both the Ross (1995) and Esmonde (2009) studies had small sample sizes, but the durations were long and the findings are similar to those of Dekker.

The findings of the research in this section indicate that cooperative learning groups are more effective if students engage in helping behaviors. Training in helping behaviors may be effective, and teacher intervention on group process, rather than product, also increases student achievement.

Cooperative Learning with Metacognitive Training and Achievement

Helping behavior is a group process that has a positive effect on student achievement, but there are also skills that can have a positive impact on student achievement when used in cooperative learning. The studies reviewed in this section showed that metacognitive training can have an added effect on student test scores when combined with cooperative learning. Working with seventh-grade students in Israel, Kramarski et al. (2002) found that cooperative learning with metacognitive training had a significant effect on student test scores as compared with cooperative learning alonel. This study had a moderate sample size (91) and p-value ($p<.05$), which is not particularly strong on its own. However, Mevarech and Kramarski (1997) also found that seventh-grade students in Israel that participated in the IMPROVE program, which used extensive cooperative learning strategies and metacognitive training, had significantly better test scores than students taught with traditional lecture and individual learning. This study was large (247) and the p-values ($p<.01$) indicated that the metacognitive training and cooperative learning strategies did have a positive effect. Kramarski and Mevarech (2003) determined that eighth-grade Israeli students that were taught with cooperative learning and metacognitive training outperformed students taught with metacognitive training and individualized instruction. Goos et al. (2002) found that metacognitive, transactive comments on problem-solving were helpful for students.

The findings of these studies indicate that cooperative learning used with metacognitive training has a positive effect on student test scores. Additionally, this effect is larger when both cooperative learning and metacognitive training are used together rather than separately.

Classroom Implications

The research on cooperative learning and student achievement found that cooperative learning structures like STAD and TGT can have a positive effect on student test scores. Additionally, cooperative learning has more effect on conceptual topics and applications than on computation. Finally, cooperative learning in heterogeneous groups has a positive effect or no effect on student achievement.

The findings of the research reviewed in Chapter 2 indicated that STAD and TGT have a generally positive effect on student achievement. Students from the fourth grade to the twelfth grade were studied, with the bulk of the studies focusing on the Grades 7-9. Additionally, the studies used samples that were composed of various races and ethnicities, with samples that were mainly European American (Whicker and Bhol, 1997) and mostly African American (Slavin and Karweit, 1984). Due to the wide range of ages, the classroom implications would be to include cooperative learning through STAD and TGT in secondary classrooms in the United States that are composed of European American and African American students. Using these structures could have a positive effect on student achievement as measured through standardized tests such as the CAT and the CTBS.

The findings reviewed in Chapter 2 also indicate that cooperative learning has a positive effect on conceptual mathematics more than computational mathematics. Galton et al. (2009) found that cooperative learning had a significant positive effect on test questions with high-level cognitive demand, and Johnson et al. (1978) found that the positive effect of cooperative learning on high cognitive-demand questions increased as the difficulty of the questions increased. The ages of the students studied were lower, from 11 to 14, and the racial composition of the samples was not included in the studies. However, Henderson and Landesman (1995) also found that cooperative learning had a positive effect on the conceptual and application problems for Latino/a students. These all indicate that teachers could use cooperative learning in lower grades (eighth and ninth) and with students that are Latino/a and European American as a means to increase student achievement in conceptual, high-level mathematical questions.

The research also indicated that group composition was a factor in student achievement, and that heterogeneous groups have a positive or, at the very least, no negative impact on student achievement. Both Linchevski and Kutscher (1998) and Leonard (2001) found that heterogeneous group composition had a positive effect on student achievement as measured by test scores. Linchevski and Kutscher (1998) studied seventh graders in Israel, which has limited generalizability to the United States, but Leonard (2001) studied sixth graders in Maryland. These studies indicate that a teacher could use heterogeneous groupings in cooperative learning environments as a means for improving student achievement.

For cooperative learning to have a positive effect on student achievement, group activities and process must be conducive to learning. As Webb and Cullian (1983) found, students performed worse on a test if they asked questions and received no help. Due to this finding, teachers should ensure that group process is optimal for increased student achievement.

The research reviewed in Chapter 2 indicates that help seeking behavior where specific explanations were requested were positively correlated to test scores of seventh-grade students in California (Webb and Mastergeorge, 2003). Teachers would benefit from being able to encourage this kind of behavior, and one classroom implication as indicated by the research would be training in help-seeking and help-giving behavior. Webb and Farivar (1994) found that training in helping behaviors while engaging in cooperative learning had a significant effect on some African American and Latino/a students in urban, seventh-grade classes in the United States. Veenman et al. (2005) also found that there was a positive, but not significantly positive, effect on test scores of students from the Netherlands who were trained in help-seeking and help-giving behavior. These findings indicate that student achievement could be increased through training in help-seeking and help-giving behavior. Due to this, teachers should engage in this training in the classroom.

Teachers should also intervene and give feedback on group process during cooperative learning. Dekker (2004) found that interventions in group process had a significant effect on 16- and 17-year-old students' test scores. This finding coincides with Ross's (1995) finding that teacher feedback on help-seeking and help-giving had a positive impact on increasing this behavior. These findings indicate that teachers should focus on process-oriented feedback rather than product-oriented feedback during cooperative learning.

In addition to encouraging help-giving and help-seeking behavior through training and process-oriented feedback, teachers can also improve cooperative learning in their classroom through metacognitive training. Kramarski et al. (2002) found that cooperative learning with metacognitive training had a significant positive effect on student test scores. Additionally, Mevarech and Kramarski (1997) found that metacognitive training with cooperative learning was more effective than individual, traditional instruction. Kramarski and Mevarech (2003) also found that metacognitive training was more effective when used with cooperative learning than with individual instruction. These results indicate that teachers should train students in metacognitive skills when working in cooperative learning groups.

In the current educational environment where there is significant pressure on teachers to increase student test scores, teachers need specific strategies to use in the classroom that can have a positive effect on student achievement as measured by test scores. The research reviewed in Chapter 2 and further summarized earlier in Chapter 3 indicates that there are a variety of specific applications of cooperative learning and strategies that can be used to improve cooperative learning and to have a positive effect on student test scores. Teachers can use STAD and TGT to increase student achievement. Teachers using cooperative learning for applications and high-level cognitive-demand tasks will find a greater effect than on low-level or computational tasks. When using cooperative learning, teachers should use heterogeneous groupings and should train students in help-seeking and help-giving behavior. Additionally, teachers should focus interventions and reinforcement on group process rather than on product. Finally, teachers should train students in metacognitive skills to improve the positive effect of cooperative learning on student achievement.

Suggestions for Further Research

A teacher in the United States attempting to incorporate methods into his or her classroom that meet the criteria of “research-based” as defined by No Child Left Behind must be able to show that the research is generalizable to the classroom. The research reviewed in Chapter 2 varied in its generalizability to the United States, to secondary mathematics classrooms, and to a wide range of races and ethnicities. Due to the varying sample populations, ages, races, and ethnicities that were studied, there is a variety of possible avenues for further research in the topic.

The effect of STAD and TGT on student achievement has been widely researched and the studies reviewed in Chapter 2 focused on a wide range of ages, races, ethnicities, and countries. However, the studies that focused on the effect of cooperative learning on student achievement on high-level cognitive-demand tasks and conceptual understanding had some large gaps. Two of the studies did not have a racial/ethnic composition of the sample population and one study was almost entirely Latino/a. Further research could be done in this area using populations in the United States of varying races and ethnicities. Studies that focused on classrooms with a heterogeneous population, an African American population and a European American population could all be beneficial to the body of research. Additionally, more explanation and breakdown of socioeconomic class of the sample could be beneficial as well. Teachers must be able to show that the research is applicable to their classroom, and if a teacher does not teach a largely Latino/a class, this research may not meet the necessary criteria to be useable.

Group composition was another area reviewed that was lacking in studies focused on the United States. Linchevski and Kutscher (1998) was done in Israel and Leonard (2001) was done in Maryland with a small, sixth-grade sample population. This area could use significantly more research focusing on a variety of ages, specifically anything at the secondary level. Additionally, research into the effects of heterogeneous ability grouping with race as a dependent factor and also heterogeneous racial/ethnic grouping and student achievement would be another area of research that could prove insightful.

The research on help-seeking and help-giving behavior in cooperative learning classrooms also had significant gaps in sample populations analyzed that could provide additional strength or contradictory findings. Webb and Farivar

(1994) focused on seventh-grade students in heterogeneous classrooms, but there were no studies about the effectiveness of training on help-giving and help-seeking behavior of secondary students. More research focusing on secondary students and studies that were longer in length, perhaps over a semester or an entire school year, would be helpful in determining the effectiveness of training. Training in a new behavior or skill will take a while to be used regularly in interactions, so more time for the study would be important to gain better understanding of the effectiveness.

In general, the research on help-seeking and help-giving was dominated by Noreen Webb and her counterparts, which demonstrates a need for further research by other educators.

The research reviewed that focused on metacognitive training was almost entirely done in Israel. Further research in this area should be focused on secondary mathematics students in the United States. The sample populations should include a variety of races, ethnicities, and socioeconomic classes. Additionally, the specific methods for teaching metacognitive skills and the methods for continuing an environment that uses metacognitive skills regularly would be useful for teachers. Without these concrete examples, putting metacognitive training into place in the classroom would be difficult. Replication of a set of training procedures that is included in a study would provide a baseline for teachers to use when implementing these ideas in their own classes.

Conclusion

Cooperative learning is one of the “great success stories of social and educational psychology” (Johnson & Johnson, 2009, p. 365). However traditional, individualized instruction tends to be dominant in many classrooms in the United States. With the advent of No Child Left Behind and the need for research-based practices in the classroom, many teachers are reluctant to try new methods for teaching. While cooperative learning is not new, some of the specific strategies and additional skills that can improve student achievement are.

Cooperative learning can have a positive effect on student achievement as measured by test scores. STAD and TGT in particular are effective cooperative learning strategies that have been heavily researched and can have a significant, positive effect on student achievement. Cooperative learning can also positively impact student achievement on tests that focus on high-level cognitive-demand questions and conceptual understanding. Also, when using cooperative learning, heterogeneous groupings can have a positive effect on student achievement and test scores.

Cooperative learning encompasses a broad range of strategies and methods. However, the one key component that all of the strategies and methods have in common is students working together in a group, cooperatively. Working in a group can pose some significant challenges. The work can be inequitably distributed and some students may be lost in the shuffle. However, groups that exhibit consistent help-seeking and help-giving behavior tend to have a positive effect on student achievement. Through training in help-seeking and help-giving behavior and teacher intervention and reinforcement focused on the process as opposed to the product, cooperative learning can have a positive effect on student achievement.

Finally, cooperative learning combined with metacognitive training can have a significant positive effect on student achievement, though more research is necessary to fully inform this finding.

REFERENCES

Aronson, E. (1997). *The jigsaw classroom: Building cooperation in the classroom*. New York, NY: Addison Wesley Educational Publishers, Inc.

Boaler, J. (2006). How a detracked mathematics approach promoted respect, responsibility and high achievement. *Theory into Practice, 45*(1), 40-46.

Brush, T. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Technology Research & Development,* 45(1), 51-64.

Caillier, J. G. (2007). The no child left behind act: "Are states on target to make their goals?" *The Journal of Negro Education*, *76*(4), 582-596.

Cohen, E. G., & Lotan, R. A. (1990). Teacher as supervisor of complex technology. *Theory into Practice*, *29*(2), 78-84.

Davidson, N., & Kroll, D. L. (1991). An overview of research on cooperative learning related to mathematics. *Journal for Research in Mathematics Education*, *22*(5), 362-365.

Dekker, R. (2004). Teacher interventions aimed at mathematical level raising during collaborative learning. *Educational Studies in Mathematics,* 56(1), 39-65.

Dewey, J. (1938). *Experience and education*. New York, NY: Macmillan Publishing Co., Inc.

Esmonde, I. (2009). Mathematics learning in groups: Analyzing equity in two cooperative activity structures. *The Journal of the Learning Sciences, 18,* 247-284.

Galton, M., Hargreaves, L., & Pell, T. (2009). Group work and whole-class teaching with 11- to 14-year olds compared. *Cambridge Journal of Education, 39*(1), 119-140.

Goos, M., Galbraith, P. & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics, 49*(2)*,* 193-223.

Harskamp, E., & Ding, N. (2006). Structured collaboration versus individual learning in solving physics problems. *International Journal of Science Education*, *28*(14), 1669-1688.

Hawkins, J. & Doueck, H. & Lishner, D. (1988). Changing teaching practices in mainstream classrooms to improve bonding and behavior of low achievers. *American Educational Research Journal*, 25(1), 31-50.

Henderson, R., & Landesman, E. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *The Journal of Educational Research, 88*(5), 290-300.

Ifamuyiwa, S. & Akinsola, M. (2008). Improving senior secondary school students' attitude towards mathematics through self and cooperative-instructional strategies. *International Journal of Mathematical Education in Science and Technology, 39*(5), 569-585.

Johnson, D. W., & Johnson, R. T. (1974). Instructional goal structure: Cooperative, competitive, or individualistic. *Review of Educational Research*, *44*(2), 213-240.

Johnson, D.W., & Johnson, R.T. (1991). *Learning together and alone: Cooperative, competitive and individualistic learning.* (3rd ed.). Englewood Cliffs, NJ: Pearson.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365-379.

Johnson, D.W., Johnson, R.T., & Johnson – Holubec, E. (1994). *The new circles of learning: Cooperation in the classroom and school*. Alexandria, VA: Association for Supervision and Curriculum Development.

Johnson, D. & Johnson, R. & Scott, L. (1978). The effects of cooperative and individualized instruction on student attitudes and achievement. *The Journal of Social Psychology, 104*, 207-216.

Ke, F. & Grabowski, B. (2007). Gameplaying for maths learning: Cooperative or not? *British Journal of Educational Technology, 38*(2). 249-259.

Kramarski, B. & Mevarech, Z. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, *40*(1), 281-310.

Kramarski, B. & Mevarech, Z. & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*(2), 225-250.

Leikin, R. & Zaslavsky, O. (1997). Facilitating student interaction in mathematics in a cooperative learning setting. *Journal for Research in Mathematics Education.* 28(3), 331-354.

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

Linchevski, L. & Kutscher, B. (1998). Tell me with whom you're learning and I'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education, 29*(5), 533-554.

Mevarech, Z. (1991). Learning mathematics in different mastery environments. *Journal of Educational Research, 84*(4)*,* 225-231.

Mevarech, Z. & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34(*2), 365-394.

Miller, P. (2011). *Theories of developmental psychology* (5^{th} ed.). New York, NY: Worth Publishing.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211-246.

Ross, J. (1995). Effects of feedback on student behavior in cooperative learning groups in a grade 7 math class. *The Elementary School Journal, 96*(2), 125-143.

Sharan, Y. (2010). Cooperative learning for academic and social gains: Valued pedagogy, problematic practice. *European Journal of Education*, *45*(2), 300-313.

Shaul, M. S., & Ganson, H. C. (2005). The no child left behind act of 2001: The federal government's role in strengthening accountability for student performance. *Review of Research in Education*, *29*, 151-165.

Sherman, L. & Thomas, M. (1986). Mathematics achievement in cooperative versus individualistic goal-structured high school classrooms. *Journal of Educational Research. 79*(3), 169-172.

Slavin, R. & Karweit, N. (1984). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal, 21*(4), 725-736.

Slavin, R. & Karweit, N. (1985). Effects of whole class, ability grouped and individualized instruction on mathematics achievement. *American Educational Research Journal, 22(*3), 351-367.

Slavin, R., Sharan, S., Kagan, S., Hertz-Lazarowitz, R., Webb, C., & Schmuck, R. (Eds.). (1985). *Learning to cooperate, cooperate to learn.* New York, NY: Plenum Press.

Souvignier, E. & Kronenberger, J. (2007). Cooperative learning in third graders' jigsaw groups for mathematics and science with and without questioning training. *British Journal of Educational Psychology, 77,* 755-771.

Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary Mathematics classroom. *Cognition and Instruction*, *25*(2-3), 161-217.

Stewart, D. C. (1988). Collaborative learning and composition: Boon or bane? *Rhetoric Review*, *7*(1), 58-83.

Tarim, K. & Akdeniz, F. (2008). The effects of cooperative learning on Turkish elementary students' mathematics achievement and attitude towards mathematics using TAI and STAD methods. *Educational Studies in Mathematics, 67,* 77-91.

Vaughan, W. (2002). Effects of cooperative learning on achievement and attitude among students of color. *The Journal of Educational Research, 95*(6), 359-364.

Veenman, S., Denessen, E., van den Akker, A., & van der Rijt, A. (2005). Effects of a cooperative learning program on the elaborations of students during help seeking and help giving. *American Educational Research Journal, 42*(1), 115-151.

Webb, N. & Cullian L. (1983). Group interaction and achievement in small groups: Stability over time. *American Educational Research Journal, 20(*3), 411-423.

Webb, N. & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31*(2), 369-395.

Webb, N. & Mastergeorge, A. (2003). The development of students' helping behavior and learning in peer-directed small groups. *Cognition and Instruction*, 21(4), 361-428.

Whicker, K. & Bhol, L. (1997). Cooperative learning in the secondary mathematics classroom. *The Journal of Educational Research, 91*(1), 42-48.