

The Effects of Cooperative Learning on the Academic Achievement,
Social Interaction, Behavior, and Affect of Secondary English and Social
Studies Students

by

Wayne Stauffer

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

Wayne Stauffer

has been approved for

The Evergreen State College

by

Lester Krupp, Member of the Faculty

June 2013

ACKNOWLEDGEMENTS

I wish to thank my family for supporting me in my efforts to complete this paper and the degree of Master in Teaching. I would also like to extend my thanks to the Evergreen MiT faculty for their guidance. In addition, I would like to thank my MiT cohort for their unwavering support. Finally, I would like to thank Kelly Ann Smith for the time and energy she spent revising and editing this project.

ABSTRACT

This critical review of the literature explores the effects of cooperative learning on the academic achievement, social interaction, behavior, and affect of secondary-level English and social studies students. The majority of cooperative learning research has initially focused on elementary-level students; however, in recent years more research has been conducted at the secondary level. This review suggests that cooperative learning has positive effects on academic achievement when students are accountable only to themselves, when they are accountable to both themselves and their group, and when they are solely accountable to their group. It also suggests that students who engage in cooperative learning are more likely to work with others even when not told to. Other research suggests that students who work in cooperative settings are more likely to accept peers of different ethnicity, class, race, and ableness. These findings have major implications for teachers. Teachers can use different accountability structures of cooperative learning to reinforce different skills and values.

TABLE OF CONTENTS

TITLE PAGE	i
APPROVAL PAGE	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
CHAPTER 1: INTRODUCTION	1
Introduction	1
Rationale.....	3
Historical Background.....	3
Definitions	7
Limitations.....	14
Statement of Purpose	14
Summary	15
CHAPTER 2: CRITICAL REVIEW OF THE LITERATURE	16
Introduction	16
Effects of Cooperative Learning on Academic Achievement.....	17
Individual Accountability Structure.....	19
Mixed Individual and Group Accountability Structure	24
Group Accountability Structure.....	37
Effects of Cooperative Learning on Social Interaction, Behavior, and Affect.....	65
Effects of Previous Exposure to Cooperative Methods	66

Student Attitude, Perception, and Preference Toward Cooperative Learning.....	70
Effects of Cooperative Learning on Student Attitudes and Behavior Toward Others.....	85
Summary	100
CHAPTER 3: CONCLUSION.....	102
Introduction	102
Summary of Findings.....	104
Classroom Implications.....	119
Further Research.....	123
Conclusion	127
REFERENCES	130

CHAPTER ONE: INTRODUCTION

Introduction

In the last 40 years cooperative learning has been studied in a variety of environments ranging from kindergarten to military applications, and it is one of the most researched and implemented teaching methods to date (Hagman & Hayes, 1986). However, even though there have been over 1,200 studies conducted on the effects of cooperative, competitive, and individualistic efforts since the beginning of the twentieth century (Johnson & Johnson, 2009), there has been relatively little research on the effects of cooperative learning on attainment of mastery of concepts in humanities classes at the secondary level (Holliday, 2000) when compared to the body of research on the effects of cooperative learning on primary-level students. The purpose of this paper is to review the literature on the academic and social effects of cooperative learning on secondary English language arts and social studies students.

Research on cooperative learning in the last four decades has always focused on both the academic and cognitive effects of instruction on students, as well as the social-interactive implications that accompany these effects (Sharan, 1988). Innumerable variables and constraints of the classroom environment make relating the cognitive and social domains difficult. Slavin's 1983 review of the literature sought to reveal which aspects of various cooperative learning methods contributed to student achievement in some studies and not in others. Slavin concluded that the most prominent component in studies with measurable gains was a group reward structure, but his conclusion did not apply to the group

investigation (GI) method (Slavin, 1983). Johnson, Johnson, & Stanne (2000) completed a meta-analysis of 164 studies of cooperative learning, in which they examined the effects of cooperative goal structures compared to individualistic and competitive reward structures among eight of the most thoroughly researched models of cooperative learning. Although they found several significant effects on academic achievement, age and subject area were not variables. Roseth, Johnson, & Johnson (2008) completed yet another meta-analysis of the effects of cooperative learning on academic achievement on 12-15 year-olds, comparing the effects of competitive, individualistic, and cooperative goal structures. Of the 148 studies analyzed, the cooperative comparison yielded the most significant positive effects on academic achievement, and although grade level was taken into consideration, academic subjects of study were not (Roseth, Johnson, & Johnson, 2008).

In sum, a great deal of research has been conducted on the effects of different structures of cooperative learning on academic achievement, as well as the effects of cooperative learning on social interaction, behavior, and affect, but there has not been a comprehensive review of the literature focusing on the effects of cooperative learning on secondary-level English language arts and social studies students. This review, therefore, will examine the effects of different structures of cooperative learning on academic achievement and the effects of cooperative learning on social interaction, behavior, and affect on secondary level English language arts and social studies students. After examining the data, educators will be better able to determine which methods

align with their teaching practices and have a better understanding of which methods to employ.

Rationale

According to Roger and David Johnson, perceptions and interactions between students are often neglected aspects of instruction. A great deal of emphasis is placed on teachers' responsibilities to ensure that students have appropriate interactions with materials (such as textbooks and computers) and that students have appropriate interactions with the teacher, but relatively little time is spent structuring opportunities for students to have meaningful and appropriate interactions with each other (Johnson & Johnson, 1994).

Interpersonal interaction is a necessary component of a functioning society. It promotes cognitive and social development and also impacts the quality of peer relationships. The rationale for conducting a critical review of the literature on the effects of cooperative learning on the academic achievement, social interaction, behavior, and affect of secondary English language arts and social studies students is therefore simple: an examination of different methods and structures of the pedagogy will reveal which practices align with educators' personal and professional educational expectations so that cooperative learning can be effectively included in the classroom. This review gives insight into which methods of cooperative learning are most effective in the cognitive and social development of students.

Historical Background

The ability to coordinate and organize human effort toward a common

purpose is one of the greatest differences between humans and the rest of the animal kingdom. Likewise, our ability to do so has been responsible for many of humanity's greatest achievements. Cooperative learning is not a new idea. The Talmud, a compilation of early Jewish religious and civil laws, clearly states the need for one to have a partner in learning (Putnam, 1997). Socrates taught in small groups and engaged his students in his famous art of discourse, meaning to learn from others through dialogue and discussion rather than simply receive input (Johnson, Johnson, & Smith, 1998). Comenius, the famous Moravian educator, also believed that students could greatly benefit by playing the role of the teacher. This is very similar to the art of discourse that is found in cooperative learning; students teach each other through dialogue and discussion. In the late 18th century, Joseph Lancaster and Andrew Bell used cooperative learning groups in England and India, and later opened the Lancaster school in New York in 1806, which greatly emphasized the use of cooperative learning in instruction. Cooperative learning was also employed as a means of instruction in the common schools of the early 1800s (Johnson, Johnson, & Smith, 1998).

Late in the 19th century, Francis Wayland Parker introduced methods of cooperative learning at his experimental Francis W. Parker School of Chicago, in an attempt to move away from formalism in education and toward a classroom that acknowledged the interconnected nature of living and learning. The practices he introduced dominated American education throughout the last three decades of the 19th century, and were later labeled as progressive education as they gained acceptance by John Dewey, who also introduced the method at his own

experimental school in Chicago (Johnson, Johnson, & Smith, 1998). Dewey became the next great American advocate of cooperative learning, and his project method came to be widely used in American schools until the 1930s (DeVillar, Faltis, & Cummins, 1994). The practices informed by Parker and Dewey sparked the debate over how children in the United States should be educated, and the controversy over pedagogy is still impassioned to this day. In the 1930s, interpersonal competition became emphasized in schools. (DeVillar et al., 1994). During these years Dewey was viewed as a radical, and although the methods he advocated were mostly pushed aside in favor of a traditional model of education, his ideas were not completely forgotten.

Cooperative learning has also been influenced by a number of sociocultural theories. In the early 1900s, Gestalt psychology began to move from mechanistic to dynamic whole theories, and Kurt Koffka proposed that groups of people were similarly dynamic with varied interdependence among the members of a group. Kurt Lewin later posited that the essence of a group is the interdependence of its members, resulting in a dynamic whole that changes as a result of changes within any member or subgroup. Morton Deutsch took the idea further by conceptualizing positive and negative interdependence in groups; positive interdependence exists when individuals believe they can only achieve their goals if others in the group achieve their goals as well, and negative interdependence exists when individuals believe they can achieve their goals only if other group members fail to achieve theirs. This theory is known today as the social interdependence theory. (Johnson & Johnson, 2009). Cooperative

learning methods rely greatly on the notion of positive interdependence among group members.

Modern cooperative learning is also grounded in the theory of cognitive learning, which was first outlined by Jean Piaget and later refined in Lev Vygotsky's theory of learning as a sociocultural phenomenon. Cognitive learning theory, simply put, is the notion that humans learn through cognitive organization and adaptation in a process of seeking equilibrium with the environment. Vygotsky's theory introduced the idea of the zone of proximal development, which is an area between a child's independent level of ability, and a child's potential ability level given the help of an adult or more capable peer. A combination of these ideas later took on the name of social constructivism (Miller, 2011).

Between the 1940s and 1970s cooperative learning was relatively unknown and unused as a result of the social Darwinism that promoted individualism and competition for survival. After World War II, progressive ideas such as Dewey's were rejected in favor of instructional methods based on rote learning, drill, and memorization. In 1957, with the launch of Sputnik, education faced its greatest reform. Math and science were at the forefront of the curriculum. Instruction reverted to more traditional methods focused on student behavior. In the late 1960s, the notion of direct competition in schools began to receive criticism, and the idea of rugged individualism, that strong individuals were developed through isolation and independent learning, began to take hold. Operant conditioning and programmed learning, procedures designed to let

students control the pace and level of their own content and allow for behavior modification, were recommended (Johnson & Johnson, 2009). Opposing social scientists challenged this individualistic learning, citing the importance of peer interaction and socialization in education; these challenges also led to further research into the effectiveness of the theories of social interdependence and cooperative learning.

The practice of cooperative learning has been unusually successful in comparison to many others that have emerged and subsided in the last 60 years. Skinner's programmed learning and Madeline Hunter's Steps of Teaching were at first popular for a few years, but shortly thereafter were forgotten or abandoned in favor of other methods. Cooperative learning has followed a different path. Although it, too, was ignored after the 1940s, it experienced a resurgence in the 1960s and has continually evolved and progressed to become one of the most fundamental teaching practices throughout the world (Johnson & Johnson, 2009). By the 1990s, cooperative learning had become popular in schools in the United States, Canada, Israel, Australia, New Zealand, Europe, and Africa, and since then it has also established a foothold in Asia (*Cooperative Learning*, 1991; Nguyen, Elliott, Terlouw, & Pilot, 2009),

Definitions

Some definitions of terms in the area of cooperative learning must be understood before the pedagogical model and its various methods can be fully comprehended. First, cooperative learning is not simply the act of students working together and sharing materials. It involves these principles, but it is

imperative that it also involves “face-to-face interaction by all students, heterogeneous teams, structured goal interdependence (including a group goal and group recognition), individual accountability, and an emphasis on practicing social skills” (*Cooperative Learning*, 1991, p. 1).

The term *cooperative learning* is often incorrectly used interchangeably with the term *collaborative learning*. Collaborative learning is a broader concept that encompasses all forms of working together, such as peer tutoring. Joseph Trimbur (as cited in Stewart, 1988) wrote:

Collaborative learning is a generic term, covering a range of techniques that have become increasingly visible in the past ten years, practices such as reader response, peer critiques, small writing groups, joint writing projects, and peer tutoring in writing centers and classrooms. The term refers to a method of conducting the business at hand—whether a freshman composition course or a workshop for writing teachers. By shifting initiative and responsibility from the group leader to the members of the group, collaborative learning offers a style of leadership that actively involves the participants in their own learning. (p. 59)

In contrast, cooperative learning is a structured method that relies on real or perceived individual and group roles and responsibilities, and often a reward structure. Likewise, in cooperative learning settings, instructional goal structures are dependent on individuals and the group as a whole, whereas in a collaborative learning setting such as peer tutoring, the goals of the participants may have no relation. For example, the tutor’s goal might be recognition or

money, and the tutee's goal might be increased academic ability.

There are differing philosophical views on the goal and reward structures of different cooperative learning models. Slavin's models are typically structured to include rewards, usually in the form of public recognition for achievements, both at the individual and group level. Therefore the models rely on extrinsic motivation. The Johnsons' models, on the other hand, rely on a group goal structure. The positive interdependence and accountability of group members to each other is the impetus for success; intrinsic motivation is the primary means of student engagement. Regardless of these different philosophies underlying the purpose of the group structure, the structure itself is the focus of the first section of the critical review of the literature. Consequently, the term *accountability structure* will be used throughout the remainder of this paper to represent both group- and reward-structure methods of cooperative learning.

Another term in need of clarification for the purpose of this paper is *secondary*. In common usage, the term *secondary-level* education refers to grades 9-12 or 10-12, and the term *middle-level* is used to refer to grades 6-8 or 7-9. At other times, the term *secondary-level* refers to grades 6-12, which is how the term shall be used throughout this essay. In addition, although the bulk of the articles in the critical review section are studies of the effects of cooperative learning on secondary-level students, there are some studies whose subjects do not fall into that age range and are instead slightly younger or older.

One aspect of cooperative learning that sets it apart from many other pedagogical models is its engaging of the students in active, rather than passive,

learning. Though there are many different cooperative learning strategies, all of them can be viewed as an alternative to teacher-centered pedagogy or assigned seatwork (*Cooperative Learning*, 1991). There are dozens, if not hundreds, of different cooperative learning methods, and the most popular examples of the present day were primarily developed through the efforts of Elliot Aronson, David DeVries, David W. Johnson, Roger T. Johnson, Spencer Kagan, Nancy Madden, Shlomo Sharan, Robert Slavin, and Noreen Webb. Some of the strategies discussed in this review are: Cooperative Integrated Reading and Composition (CIRC), Group Investigation (GI), Jigsaw I, Jigsaw II, Student Team Achievement Divisions (STAD), and Teams-Games-Tournament (TGT).

Cooperative Integrated Teaching and Composition (CIRC)

In the CIRC model, students are assigned to reading groups of 8-15, as well as pairs or triads within each reading group. After a story is introduced to the class, partners read silently, then take turns reading aloud. Halfway through the story, students are given questions on story-structure and related concepts. Students are also given lists of difficult words to practice with their partners, and an additional list of words to look up in the dictionary; students are expected to paraphrase the definitions and incorporate them into their speaking vocabulary. After reading the story and discussing it in their groups, students retell the main parts with their partners. Students also pretest each other on spelling at the beginning of each week, then spend the remainder of the week helping each other master the words. The structure is fairly open to teacher interpretation, and it incorporates direct instruction, pair work, and large group work. Tests have

shown that comprehension, vocabulary, and spelling skills were improved through this method when compared to only frontal instruction (Slavin, Madden, & Stevens, 1989).

Group Investigation (GI)

In this model, developed by Yael and Shlomo Sharan, students form interest groups to plan and implement an investigation, then synthesize their findings into a presentation given in front of the class. The teacher's general role is to let students know where they can find resources. The four I's, or four main components, are investigation, interaction, interpretation, and intrinsic motivation. There are six main steps: First, the teacher presents a multi-faceted problem to the class, and students choose an interest group. Groups the plan and carry out their investigations. Groups plan their presentations and then present. Finally, students and the teacher evaluate the investigations and presentations (Tan, Sharan, & Lee, 2006)

Jigsaw I

Jigsaw was developed by Elliot Aronson at the University of Texas, then later adapted by Robert Slavin. Students are assigned to groups of five or six members, and then academic material is presented in text form. Each student is responsible for learning a part of the given material. Members from each group form an expert group and meet to study and learn more about the topic. Then the expert members return to the original group and share their findings. After this meeting, students take quizzes individually. High-scoring teams and individuals are then given recognition (Arends, 1997).

Jigsaw II

This version of the Jigsaw method was Slavin's extension to Aronson's model. As in Aronson's model, students are responsible for mastering their own portion of the material, and experts are responsible for teaching their material to the rest of the home group. The difference between this model and the original is that the experts take an expertise test before returning to their home groups (Şahin, 2010). Then, after instruction, students are tested individually, and team scores are produced based on achievement relative to previous group performances. This is known as "equal opportunity scoring," and it may or may not be the final evaluation for the students. Group recognition is then given based on total academic achievement (Mattingly & VanSickle, 1991).

Student-Teams-Reading-Writing (STRW)

Developed by Robert J. Stevens in 1992, STRW is an integrated approach to cooperative learning for early adolescents in language arts. The program comprises cooperative methods and a variety of other features: a literature anthology of high-interest material; explicit reading comprehension instruction; integrated reading, writing, and language arts instruction; and a writing process approach to language arts. Students work in heterogeneous learning teams and follow a routine of teacher presentation, team practice, independent practice, peer pre-assessment, and individual assessments that form the basis of team scores (U.S. Department of Education, 2011).

Student Team Achievement Divisions (STAD)

Developed by Slavin and his colleagues at Johns Hopkins, this is

probably the simplest and most straightforward cooperative learning approach. The teacher presents material or text, and heterogeneous student teams of four or five members work together to master the material. Students take individual quizzes regularly, and each individual is given an improvement score based on past averages. Both individual and team scores with the most improvement are then recognized (Arends, 1997).

Teams-Games-Tournament (TGT)

TGT was designed for use in upper-elementary and secondary school classrooms. Students are divided into four-member heterogeneous teams, each representing the academic range of the classroom. The students play instructional games, based on relevant academic material, which reinforce knowledge of concepts or skills. Academically comparable students from each team compete individually with students from other teams in tournaments designed with content from the lesson they have been studying. Scores are reported at the team and individual level, but grades are given to the teams (DeVries, Edwards, & Wells, 1974).

There are other mixed-method variations of cooperative learning that will be discussed in chapter two. Some are based on ability-grouping, while others are based on heterogeneous grouping. Methods not described in detail here will be further described in the next chapter. Additionally, the language of formal research and cooperative methods are often interchangeable. Often, the words *group* and *team* are used synonymously, while other times the word *group* is used to refer to the sum of the teams or the sum of the individuals participating in

the study.

Finally, the student age group primarily addressed in this paper is secondary level, meaning grades 6-12. There are a few studies included that are outside of this grade range which are included either to support or to contextualize the paper. Studies involving students not in the 6-12 age range are relevant in that they can either lend to or detract from generalizability of cooperative learning methods.

Limitations

The research reviewed in this paper is primarily from grade 6-12 classrooms in the United States; however, some of the studies are from foreign nations including Turkey, Singapore, and Israel. The subjects of the research review come from a wide variety of socioeconomic backgrounds, and although most studies took into account factors of race, gender, and socioeconomic status, there are always variables that are unaccounted for.

Another limitation of this review is the lack of qualitative research addressed in the first section of chapter two, which covers the effects of different accountability structures on academic achievement. The effects of cooperative learning on academic achievement are typically measured in quantitative values, and therefore little qualitative research has been done on the subject.

Statement of Purpose

The purpose of this review of literature is to assess the effects of cooperative learning on the academic achievement, social interaction, behavior, and affect of secondary level English language arts and social studies students.

Cooperative learning has been one of the most thoroughly researched pedagogical methods to date, but relatively little of the research conducted on it has been on secondary English language arts and social studies students, particularly in the upper secondary grades. This review shall examine the existing literature, assess its strengths and weaknesses, and provide suggestions for further avenues of related research.

Summary

This first chapter gave a historical overview of cooperative learning and outlined the definitions and limitations of this paper. Cooperative learning has been thoroughly researched in a variety of settings and results have been generally positive. The second chapter is a critical review of the literature. Studies on the effects of cooperative learning on the academic achievement, social interaction, behavior, and affect are analyzed and critiqued. The third chapter of this paper reviews the findings of chapter two, discusses classroom implications, and suggests areas for future research. After reading this review, educators of secondary social studies and English language arts students will be able to make more informed pedagogical decisions regarding which methods of cooperative learning may be most suitable in their own classrooms.

CHAPTER TWO: CRITICAL REVIEW OF THE LITERATURE

Introduction

Chapter one gave a brief overview of the history of cooperative learning and the rationale for its study; though several forms of cooperative learning have been researched over the years, there has been a relative shortage of research on cooperative learning in English language arts, especially studies concerning students in grades 9-12. Cooperative learning, in general, draws upon the strengths of individual members to achieve a goal through interdependence. However, some methods are based on individual accountability. Other models such as Jigsaw II and Student Team Achievement Divisions (STAD) rely on both individual and group accountability. Finally, other methods of cooperative learning are solely focused on the group accountability structure. Therefore, the first section of this chapter is divided into three subsections: the effects of individual accountability methods of cooperative learning, the effects of mixed individual/group accountability methods of cooperative learning, and the effects of group accountability methods of cooperative learning. All three of the subsections pertain to the achievement of secondary level English language arts and social studies students. Because of the dearth of research in the area, some studies of cooperative learning in other content areas and with students at different grade levels have been included to contextualize the findings.

The second section of this chapter examines the effects of cooperative learning on social interaction, behavior, and affect for secondary English and social studies students. This section is also divided into three subsections: the

effects of previous exposure to cooperative learning methods on students' group interactions; the effects of cooperative learning on students' perception of cooperative learning methods; and the effects of cooperative learning on students' perceptions of each other.

Effects of Cooperative Learning on Academic Achievement

As stated in the introduction to this chapter, this first section of the critical review is divided into three subsections based on accountability structures. The first subsection deals with methods of cooperative learning based on individual accountability structures, the second subsection deals with methods that have both individual and group accountability structures, and the third subsection deals with methods that rely solely on group accountability structures. The impetus for this organization was derived from a quasi-experimental study by Slavin (1978), in which he investigated the independent effects of level of reward (recognition based on the performance of a four- to five-member cooperative learning team; recognition based on individual performance) and comparison of student quiz scores (comparison with ability-homogeneous groups and comparison with entire class) on student achievement and attitudes. His research on level of reward was a model for the investigation of the different accountability structures. He had four groups that were each subjected to different treatments. The first group was *team reward and comparison with equals*, meaning that the team was rewarded as a whole and individuals' scores were given a rank value, rather than a point value, before being added up. The second group was *team reward and comparison with entire class*, meaning that students worked in

groups just the same as in group one, but their scores were simply added together instead of being ranked and added together. The other two groups were *individual reward and comparison with others* and *individual reward and comparison with equals*. Students in both of these groups worked individually on the activities but their scores were calculated in the same fashion as the first two groups. His study was conducted on 205 seventh-grade students in eight intact English classes in the principal town of a rural Maryland county. After analyzing the standardized and treatment-specific measures through an ANCOVA, the researcher found no academic achievement effects for either factor: level of reward or comparison with individuals or group, meaning that there was no significant difference in achievement when students were rewarded individually or rewarded as a group. Similar forms of two different tests were used as measures of academic achievement, as well as twice-weekly quizzes during the last three weeks of the ten-week study. All eight classes were taught exactly the same material, and a counterbalanced method in assigning teachers and classes was used to distribute teacher effects evenly across the main effects. A counterbalanced distribution is a method of distributing inputs in a different order, so as to balance their potential effects. The use of a counterbalanced design gives the results of the study an increased degree of reliability. Slavin's thorough analysis of the academic achievement results also gives the study internal validity, though the omission of correlational values on academic achievement, despite an admission that results of the treatment were insignificant, leads to somewhat questionable objectivity.

Although Slavin's study found no significantly positive effects on academic achievement for either level of reward or comparison with individuals or groups, the level of reward structures he examined are similar to the models of accountability structures to be examined, and serve as a framework for this first section of the review of the literature. The first subsection contains studies on individual accountability structures, the second contains studies on methods with both individual and group accountability structures, and the third section contains studies on group accountability methods of cooperative learning.

Individual Accountability Structure

The three studies examined in this subsection focus solely on individual accountability structures and methods of cooperative learning, meaning that even though students work together in groups, they are scored or rewarded on the basis of their own contributions, not on the results of the group effort. Lucker, Rosenfeld, Sikes, and Aronson (1976) found, in a study of the effects of Jigsaw cooperative learning on academic achievement among fifth- and sixth-grade European American and ethnic minority social studies students, that the interdependent Jigsaw setting had a significant effect on academic achievement. In another investigation of the effects of Jigsaw, Gocer (2010) conducted a comparative pre- and posttest study on Turkish high school students' ability to learn literary genres through cooperative (Jigsaw) and individualistic settings. The students in the group using the Jigsaw method showed significantly greater improvement between pre- and posttest scores than the control group. Using an unclearly defined individual accountability structure method, in the final article of

the subsection, Adeyemi (2008) conducted a pre- and posttest design study comparing the effects of a version of an individual accountability structure of cooperative learning and a traditional whole class method on academic achievement among 150 Nigerian social studies students, ages 11-15, and found that the cooperative setting had a greater effect on improved academic achievement than the traditional whole-class setting.

Focusing on a curriculum unit on Colonial America in fifth- and sixth-grade classrooms, Lucker, Rosenfeld, Sikes, and Aronson (1976) studied the effects of Jigsaw on the performance of European American and minority students working cooperatively when compared to the performance of students working in a traditional setting. The study revealed that the Anglo-Americans performed equally well in both settings, but minorities performed significantly better in the interdependent classroom environment. The sample was composed of 303 fifth- and sixth-grade students in the Austin, Texas school system. The sixth-grade sample was composed of 65 Anglo-American, 14 Mexican American, and eight African American students. The fifth-grade sample included 177 European American, 21 Mexican American, and 18 African American students. Throughout the two-week study, one half of the students worked in interdependent Jigsaw groups while the other half were in traditionally-taught classrooms. The pre- and posttest questions were true/false, multiple choice, and matching, and student responses were converted to percentages for analysis. The authors found no significant differences in reading level or pretest scores between the traditional and interdependent groups (both F s < 1.0), strengthening

the internal validity. There were, however, significant ethnic group differences between European Americans and minorities, $F(1,299) = 22.14, p < 0.001$, and $F(1,299) = 4.37, p < 0.05$, respectively. In both the traditional and interdependent groups, European American students scored higher. The difference of the pretest scores were accounted for via ANCOVA when researchers considered the results of the posttest for the different student groups. This analysis showed a significant effect for classroom type, $F(1,280) = 6.73, p < 0.02$, but not for ethnicity, $F(1,280) = 2.22, p < 0.17$. This effect for type showed that students performed significantly better in the interdependent classes. Oddly, the researchers found that the effect for type of classroom was mainly caused by differences in minority performance, $F(1,280) = 8.77, p < 0.01$, and not by differences in the performance of Anglo-Americans ($F < 1.0$). Another analysis was performed using the learning group as the nested variable instead of the teacher/classroom as that variable, and the results were nearly identical.

Another analysis was performed to determine if the interdependent format was more beneficial for the low-ability students than the high-ability students. European Americans and minorities were grouped according to their reading level scores, and the performances of the upper and lower performing 25% of students were compared against both the interdependent and traditional classroom settings. Both groups showed no interaction between classroom type and ability level ($F < 1.0$), meaning that high-ability students benefited just as much as the low-ability students in the interdependent classroom type.

What is suggestive about this study is the fact that repeated analyses

showed that classroom type had a significant effect on achievement, but ethnicity did not. However, it was determined that improvement of minority scores was the primary cause of the interdependent classrooms' achievement. Repeated analysis accounting for different nested variables yielded nearly identical results. Despite this difficulty, the methods of analysis were thorough and repeatable, therefore lending to the credibility and reliability of the study.

In Gocer's (2010) study comparing the effectiveness on academic achievement of the Jigsaw technique and conventional teaching methods, he found that the experimental group's posttest scores increased significantly. The participants were sixty eleventh-grade students in Kayseri State High School in Turkey during the 2008-2009 academic year. The students were randomly divided into control and experimental groups and given a genre questions list test to determine their prior knowledge of literary genres. The experimental group was divided into six groups of five students each, which were taught via cooperative methods, and the control group was taught via traditional whole-class methods. In the cooperative groups, after students discussed what they already knew about the topic, each group sent a representative to meet with other representatives from the other groups. After these expert groups met, they sent the representatives back and Jigsaw groups were reformed. These groups were then assigned to produce a final version of their lesson product. After finalizing what they knew about the topic, students retook the genre questions list test to see if there were any gains in knowledge between the pre- and posttest.

Results were obtained from the genre questions list, and analyses were

performed using the Statistical Package for the Social Sciences (SPSS) version 11.0. The data showed that the control group's mean scores went from 57.3 to 56.4, and the experimental group's mean scores went from 58.8 to 68.6. The statistical variances were insignificant. The *t*-test score between pre- and posttest results of the research group was -10.373, indicating significance at the $p < 0.0001$ level. The *t*-test score for the control group was insignificant (1.785, $p < 0.05$). These results suggest that study is internally valid, and the research methods suggest that it is externally valid. The high reliability of the *t*-test further strengthens the internal validity. There is no mention of attrition or mortality in the study.

Adeyemi (2008) conducted a non-randomized quasi-experimental study to investigate the effects of an individual accountability method of cooperative learning, on social studies achievement among 150 Nigerian students ages 11-15. The experimental group of students was put into groups of five and presented with instructional packages containing instructions and content to complete a task. The author does not specifically describe the academic content but states that it was related to social studies. The control group was taught via conventional lecture. The treatment lasted four weeks, and students were taught for three periods of 35 minutes each week. The investigator and three research assistants handled all treatment conditions.

The results of the study indicated that students in the experimental cooperative learning condition showed greater improvement between the pre- and posttests. Analysis of the results using the Duncan Multiple Range

Comparison showed that students in the conventional setting had a mean score of 7.26 and students in the cooperative setting had a mean score of 11.2036. All scores were significant at $p < 0.05$. The instruments of measurement administered to the students all had a split-half reliability over 78%. Split-half reliability is a measure of consistency in which a test is split into two, and the scores of each half are compared with one another. This reliability of 78% is fairly high. Pre- and posttest means were not included, costing the study both internal validity and methodological objectivity. Finally, the exact details in which the cooperative learning methods were administered are not specified, further threatening the study's reliability, objectivity, and internal validity.

The three studies on individual accountability structures and methods in this subsection found significant results for cooperative learning on academic achievement, though the study by Adeyemi (2008) was unclear as to how some the results were attained. Although two of these three studies were published somewhat recently, individualistic reward structures for cooperative learning are not as prevalent as mixed individual and group accountability methods and group-structure methods, which shall both be examined in the following subsections.

Mixed Individual and Group Accountability Structure

The six articles in this subsection examine cooperative learning methods that combine both individualistic and group accountability structures. The first five studies find significant effects of individual and group accountability structures of cooperative learning on academic achievement. The sixth study by Devries,

Edwards, and Wells (1974) finds several other positive effects of cooperative learning, but academic achievement is not among them. The first article is Mattingly and VanSickle's (1991) comparative study of the effects of Jigsaw II on academic achievement among heterogeneously grouped ninth-grade geography students at a Department of Defense high school in Germany, in which the researchers found that Jigsaw II led to higher levels of academic achievement than traditional whole-class instruction. Şahin (2010) also conducted a study on the effects of Jigsaw II on Turkish preservice language teachers' academic achievement, and found that academic achievement was significantly higher in the experimental group than in the control group, the latter of which was taught via the traditional teacher-centered method. Slavin and Oickle's (1981) study of the effects of the STAD method of cooperative learning on academic achievement and race relations among 230 sixth- through eighth-grade English language arts students found that minority students showed much higher improvement than did their European American counterparts, though both groups showed improvement. Johnson, Johnson, Johnson, and Anderson (1976) conducted a study on the effects of ability-grouped cooperative learning on fifth-grade students' academic achievement and found significant effects for the cooperative groups, which made fewer errors and had higher scores on the posttest they took as a group. Lew, Mesch, Johnson, and Johnson (1986) conducted a study comparing the effects on academic achievement of three conditions: (a) cooperative learning with group contingencies, or forms of accountability; (b) cooperative learning without group contingencies; and (c)

individualistic learning. Lew et al. found that group contingencies cooperative learning had significant positive effects on academic achievement. Finally, Devries, Edwards, and Wells' (1974) study on the effects of Teams-Groups-Teams on student academic achievement, student attitudes, cognitive beliefs, and classroom climate revealed several significant results, but academic achievement was not among them.

Mattingly and VanSickle (1991) set out to determine if Slavin's Jigsaw II method (a modification of Aronson's original Jigsaw method to include equal opportunity scoring, or EOS) would yield higher academic achievement than traditional whole-class instruction, and found that the Jigsaw II method had a significant positive effect on academic achievement. The subjects were two heterogeneous classes of ninth-grade geography students at a Department of Defense high school in Germany. Both the whole-class control ($n = 23$) and Jigsaw II ($n = 23$) groups were randomly assigned. Both groups contained students with a wide achievement level, including students from the learning disabilities program. The number of boys and girls were roughly equal, and diversity was representative of the rest of the school. Socioeconomic means varied along the military pay grade. During an experimental period of nine weeks, students covered an entire chapter on Asia. They were assigned to teams balanced in terms of low, medium, and high past achievement. All class materials for the two groups were the same.

Three different pretests were administered to the subjects because they could not be assigned randomly to the groups. The tests were a measurement of

academic ability within groups. First, the students took a 135-item multiple-choice test produced by the publisher of the class textbook. The test had a high internal consistency coefficient of 0.92. Second, the researchers gave the Hennin-Nelson Test of Mental Ability to check for any major differences in ability between members of the two groups. The third pretest was the 75-item, multiple-choice, Competency-Based Geography Test, secondary level, Form 1. This was given to estimate what students already knew about geography prior to the treatment. The posttest consisted of the cumulative total of the nine chapter tests from the textbook. The chapter tests were similar in content to the pretest but were more detailed.

Pretest scores were similar between both groups, and mean posttest scores were 86.2 for the cooperative group and 81.0 for the conventionally-taught group, with respective standard deviations of 6.1 and 6.4, $t(43) = 2.77$, $p < 0.01$. The t -score indicates that the results were statistically significant, adding to the study's objectivity and internal validity. The results of the study were well documented and highly generalizable, adding to its external validity. Finally, although the study occurred over a period of nine weeks, maturation of the subjects is not a threat to the study's validity due to the presence of a control group.

Şahin's (2010) experiment to uncover whether the Jigsaw II method would yield higher posttest and retention scores in Turkish preservice teachers found that the experimental group of Jigsaw II subjects ($n = 42$) had higher posttest scores than the control group ($n = 38$) that learned the material through

traditional teacher-centered instruction. The data for the academic achievement segment of the test was obtained by using the written expression achievement test (WEAT). The questions were selected from a graduate school entrance examination and a government employment selection examination. A pilot study was first carried out with 156 Turkish language teachers in order to measure the reliability of the test. The results showed that five of the 30 questions were unreliable, and thus were removed. The scaled-down test consisted of 25 questions and was given to control and experimental groups as WEAT pretest, WEAT posttest, and WEAT retention-test. The groups were given an attitude questionnaire as well, but it is not relevant to this review. The participants in the experimental Jigsaw treatment were divided into six groups, and each member was responsible for taking on one expert writing task, such as correct punctuation or appropriate word usage. During the first week the groups discussed how they would study their topics, and in the second week they prepared study materials. In the third week, members met with their expert groups, and in the fourth week expert groups were given tests on their level of subject expertise. Because all of the expert groups had an achievement score over 90%, they returned back to their home groups in the fifth week and taught their own topics. In the sixth week, the topics were presented to other groups by a randomly chosen member from each group.

The mean pre- and posttest scores of the control group ($n = 38$) were 14.76 (SD = 2.10) and 21.79 (SD = 1.74), $t = 14.940$. The pre- and posttest scores for the experimental group ($n = 42$) were 14.59 (SD = 2.24) and 23.33

(SD = 1.41), $t = 23.498$. The t -scores show statistical significance between pre- and posttest scores for both groups ($p < 0.05$), but were higher for the experimental group. It can therefore be said that both the teacher-centered instruction and Jigsaw II method were effective, but the scores were nonetheless more significant for the Jigsaw II group. The experimental group also had higher scores on the retention test administered five weeks later. The score for the control group ($n = 38$) was 18.71 (SD = 1.83), and the score for the experimental Jigsaw II group ($n = 42$) was 20.47, $t = 4.916$ at $p < 0.05$. These findings suggest that Jigsaw II is more effective for learning and retention than traditional teacher-centered instruction. However, because the participants in the experiment designed their own study materials, it is difficult to say how they compared to the control group materials. Little information is given for either.

Slavin and Oickle's (1981) study investigating the effects of cooperative learning (STAD) on academic achievement and race relations among 230 students in grades six through eight studying language mechanics in a desegregated rural middle school showed that the African-American students in a group setting had significantly higher scores than the African American students in the traditionally taught control group. The study involved four classes and 84 students in the STAD group, and six classes and 146 students in the traditional group. Thirty (35.7%) of the STAD students were African American, and 48 (32.9%) of the traditional group students were African American. Three of the five teachers taught one STAD and one traditional class; one teacher taught one traditional class only; and one taught one STAD class and two traditional classes.

Teachers therefore served as their own controls for nine of the ten classes – which could have affected instrumentation and internal validity if the teachers had a predisposition toward either method.

The Hoyum-Sanders test was administered as a pre- and posttest of language mechanics, such as punctuation, capitalization, and English usage. There was a significant gain between STAD and traditional groups, $F(1,191) = 5.41, p < 0.021$. European American students gained slightly more in STAD groups than traditional groups, $F(1,129) < 1$, but African Americans had significant gains, $F(1,62) = 11.01, p < 0.002$. These different outcomes between races were reflected in a significant interaction between race and teaching method, $F(1,191) = 3.92, p < 0.049$. In the end, the scores were analyzed both by overall gain (the amount the scores changed) and analysis of covariance, and the results were identical. The traditional classes' difference in pre- and posttest scores by race were nearly identical, with European Americans scoring 6.79 points higher than African Americans, $F(1,120) = 9.07, p < 0.003$. In the STAD classes the difference in scores between racial groups was less than one point; European American students scored 0.92 points higher on the posttest, which was non-significant, $F(1,75) < 1.0$

The high F -scores and improved posttest scores indicate that there is a significant effect on African American students' achievement when working in STAD groups. The level of analysis and transparency of work indicate that the test is highly objective as well as internally valid. Furthermore, concerning academic achievement, the significance of the findings on the interaction

between ethnicity and cooperative learning confirm those of Lucker et al., whose study was reviewed in the previous subsection on individual accountability structures. The confirmation of these findings strengthens the reliability of both studies.

Johnson, Johnson, Johnson, and Anderson (1976) investigated the effects of cooperative and individualistic goal structures on academic achievement among 30 fifth-grade working-class students studying language arts. The researchers found that students in the cooperative group had fewer daily errors, fewer posttest errors, more words used in predicate phrases, and a lower percentage of original subject phrases beginning with the first-person *I*, therefore concluding that cooperative learning had a significant positive effect on academic achievement. Students were divided into the control and experimental groups after taking a language arts achievement test. Scores were ranked 1 (best) – 30 (poorest), and a coin was flipped to determine whether odd or even numbered students would be assigned to the individualized control treatment. During the 17-day research period, the assigned daily section of the book was read orally by the students in the whole class of 30. This was followed by discussion. When the teacher thought students had a firm enough grasp on the material, assignment sheets were passed out. Students in the cooperative groups, which were stratified for ability in language arts and included one high achiever, two average achievers, and one low achiever, met to complete the assignment sheet together; students in the individual setting studied alone. Members of groups received the same score, while individuals received their own score. Compared with students

in the individual setting, cooperative groups made fewer errors on daily assignments ($t = 1.73, p < 0.05$) and made fewer errors on the second posttest that they took as a group ($t = 3.14, p < 0.01$). There were no significant differences between the two conditions on the first posttest, which all students took individually. Because the study was done with one teacher and one class, it is difficult to tell how reliable and externally valid the results are, particularly given the fact that there were no significant differences between the two conditions on the posttest. Also, the fact that criteria for the teacher's assessment of students are not explicitly defined in the study means that the methodology of the study is somewhat unreliable.

Lew, Mesch, Johnson, and Johnson (1986) conducted a 21-week study on the effects of cooperative learning with group contingencies (as compared with cooperative learning without group contingencies and individualistic learning) on student achievement among three eighth-grade English classes. Contingencies, in this article, are forms of accountability. Lew et al. found that the group-contingencies condition had higher achievement scores than the no-contingencies and individualistic group. The students in the no-contingencies group in this study were only accountable through individual grades, not through a group grade. Four total conditions were studied: (1) opportunity to interact with classmates; (2) positive goal interdependence; (3) positive goal interdependence with a collaborative-skills group contingency; and (4) positive goal interdependence with a collaborative-skills and academic group contingency. The positive goal interdependence groups were grouped heterogeneously by

academic and social standing. The subjects were in an upper-middle-class school district near Boston, and all three classes were taught by the same teacher who had been trained in cooperative methods for one year. No data was given for racial and ethnic makeup of the students.

The baseline condition was to assign vocabulary words on Monday, on Tuesday to give students the choice of studying together or alone to complete a non-vocabulary instructional task, on Thursday to let them meet in groups to review the vocabulary, and then to administer a quiz on Friday. These baseline conditions were used for the control group. Students in the positive goal-interdependence condition followed the same conditions as the baseline, except they worked in heterogeneous mixed-ability groups for 20 minutes on Thursdays (instead of working individually). The collaborative-skills treatment group built upon the conditions of the positive goal-interdependence group. The difference was that students in this group were explicitly taught contribution skills such as sharing ideas, keeping the group on task, praising and encouraging, and checking for understanding among group members. The group was also given the incentive of two bonus points toward the quiz grade, per member, if each member was observed engaging in three out of four of these behaviors.

The collaborative skills and academic group contingency condition added to the collaborative skills group condition in that students were also offered an additional two bonus points toward the quiz grade per member if all group members scored 70% or more on the Friday quiz.

Over a course of 21 weeks, the conditions given below were implemented

for the three classes. One class stopped the progression at positive goal interdependence, the next class stopped at collaborative skills group contingency, and the final class progressed to collaborative skills and academic group contingency. Then the three groups worked their way back to the baseline. The initial baseline condition ran for four weeks, then the positive goal interdependence condition ran for the next three weeks. This was followed by the collaborative skills group contingency condition, which ran for four weeks, and then by the collaborative skills and academic group contingencies condition, which also ran for another four weeks. Then the collaborative skills group contingency 2 condition ran for three weeks, followed by the baseline 2 condition for the final three weeks. The three dependent variables of these conditions were achievement, social interaction, and interpersonal attraction.

Overall ANOVA and posthoc contrasts showed that students in the two cooperative learning conditions had higher achievement than students in the individualistic condition, $F(2,77) = 2.32, p < 0.10$, with no significant differences between the contingencies and no-contingencies conditions. The mean contingency achievement score was 82.47, the no-contingencies mean score was 80.01, and the individualistic score was 77.27 ($F = 2.32, p < 0.10$).

The results of this study indicate that both cooperative learning conditions are more effective than the baseline, and that there is little or no significant difference between them. This means that there was no significant difference between the effects on academic achievement of a dual individual-plus-group accountability structure and an individual accountability structure. The

methodology of the study and the significant findings it yielded give it internal reliability. The significant results of contingency-based cooperative learning have been found in other studies, giving external validity to this one. Although the article is somewhat difficult to interpret, the findings are nonetheless significant.

Devries, Edwards, and Wells (1974) conducted a study to determine the effects of Teams-Groups-Teams (TGT) on academic achievement – particularly on low ability students – among students studying social studies in a suburban high school. They were also looking for effects on attitude, classroom climate, and cognitive beliefs. They found that TGT had a number of significant outcomes, but that academic achievement was not one of them. The subjects were 191 public, suburban, high school students comprising 64% tenth graders, 26% eleventh graders, and 10% twelfth graders. Seven percent were of an ethnic minority group, and 46% were males. The study was conducted over a period of 12 weeks, and the three treatment levels were Individual Competition (IC), Teams-Games-Tournament-Average (TGT-A), and Teams-Games-Tournament-Weighted (TGT-W). There were two classroom arrangements: One teacher had large classes and was assisted by a junior teacher, and the other teacher had smaller classes in a more traditionally structured classroom. All classes were morning classes. Classes met daily over the twelve-week period, and all books and materials were the same. The lessons covered the topics of the Constitution, political institutions, the Civil War, and Reconstruction. Two measures of social studies achievement were administered representing differing levels of specificity of treatment. The first was a test representing a general measure of social

studies skills. The second measure was the specific weekly tests of six questions. The IC students were handed a list of questions Monday and were expected to take a test on Friday. Students were to find the answers alone in their textbooks. The Friday test then contained six of the questions. Students were encouraged to be competitive because they would be graded on the curve. Students were also told that the grades were important towards the final class grade. Grades were written at the top of the paper and returned the following Monday. The TGT-A groups followed a similar schedule and were broken into random groups of five or six students. Teams were told that their scores would count heavily, and that they would be in competition with other teams. The Friday quizzes for these students were labeled TGT-A tournaments; students from different teams randomly picked one of the six test questions and a panel of members from the other groups rated the answer according to a six-point scale. Teams' scores were announced the following Monday and were compared on a weekly and seasonal basis. TGT-W groups only differed from the TGT-A groups in the calculation of the team scores. Low-performing teammates' scores were weighted to account for more points toward the overall team score.

Academic achievement, analyzed through the multiple regression method, saw no teacher main effect ($F = 1.40$; $df = 1,184$; $R^2 < 0.01$), a "marginally significant treatment main effect" ($F = 2.40$; $df = 2,184$; $p < 0.10$; $R^2 = 0.02$), and a significant teacher X treatment effect ($F = 4.60$; $df = 2,184$; $p < 0.025$; $R^2 = 0.05$) (Devries et al., 1974, p. 16). Compared to both TGT conditions, the achievement of IC methods was significant at a low level of reliability ($F = 3.57$, $p < 0.10$).

Overall there were few or no significant achievement effects between the TGT-A and TGT-W treatments ($F < 1.0$). Because such thorough analysis was used to disentangle the variables, potential confounding factors that may be present in other populations were accounted for, and the analysis therefore increases the reliability of the study's instrumentation. The large sample size improves the internal validity of the research because there is a greater opportunity to demonstrate that the independent variable had an effect on the results.

Mixed individual and group accountability structures of cooperative learning seek to combine aspects of interdependence with individual accountability. Several studies found significant positive effects on academic achievement for methods of cooperative learning that incorporate both individual and group accountability structures (Mattingly & VanSickle, 1991; Şahin, 2010; Slavin & Oickle, 1981; Lew et al., 1986; and Johnson, Johnson et al., 1976). However, Lew et al. (1986) found that the teacher X treatment effect was more significant than the treatment main effect. Lastly, Devries et al. (1974) found several significant results for the use of TGT, but academic achievement was not among them. It is therefore implied that mixed-accountability methods of cooperative learning have a significantly positive impact on increased academic achievement.

Group Accountability Structure

The eight studies in this subsection focus on cooperative learning methods based on group accountability structures. The rationale behind the use of group accountability structure, as opposed to an individual or individual/group

accountability structure, is that interdependence between group members in a group accountability structure is the impetus for learning; when students are given individual rewards or scores, there is competition within the cooperative environment that leads to opposing goals for group members. Shachar and Sharan (1994) found that group investigation (GI), the first of several group accountability structures examined here, had a significant effect on the academic achievement of Israeli junior high students studying history and geography. Tan, Sharan, and Lee (2007) conducted a quasi-experimental study to determine if mixed GI groups of low-achieving students and high-achieving students would have significant differences in academic achievement. They found that high-achieving students had higher posttest scores, whereas low-achieving students had diminished academic achievement. Stevens, Madden, Slavin, and Farnish (1987) conducted two similar but separate quasi-experimental studies on the effects of cooperative integrated reading and composition (CIRC) on academic achievement in reading and writing and found that the method effected significant improvement in both.

Stevens, Slavin, and Farnish's (1991) study of the effects of CIRC, CIRC mixed with direct instruction, and traditional basal reader instruction on academic achievement found that the mixed CIRC and direct instruction method had the most significant effect on posttest achievement scores. Calderón, Hertz-Lazarowitz, and Slavin (1998) found that a version of CIRC adapted to the transitioning of limited English proficiency bilingual students (BCIRC) into English-based classes produced significant gains in student achievement than

the comparison group. Durukan's (2011) quasi-experimental study of the effects of CIRC on academic achievement showed that students taught via CIRC had significantly higher reading and writing achievement scores than did the control group students, who were taught via traditional whole-class instruction.

Other lesser-known methods of group accountability structures of cooperative learning are also critiqued in this section, such as the one by Stevens' (2003) student team reading and writing program (STRW). Stevens conducted an experiment to determine the effects of implementing (STRW) on academic achievement in high-poverty schools, and found that even though the experimental group had slightly lower pretest scores, they had significantly higher posttest scores than the control group. Johnson, Johnson, and Tauer (1979) conducted a study on the effects of interpersonal cooperation, competition, and individualistic efforts on students' academic achievement and found that there were significant positive effects for the interpersonal cooperation condition. Phuong-Mai, Terlouw, Pilot, and Elliott (2009) conducted a quasi-experimental study on the effects of culturally appropriate cooperative learning and found significant effects on achievement in what they termed the social sciences, but ultimately had to reject the hypothesis because there were no significant effects on academic achievement in the natural sciences; the researchers had hypothesized that there would be a significant increase in both.

Shachar and Sharans' (1994) quasi-experimental study on the effects of group investigation (GI) cooperative learning on academic achievement found that ethnically heterogeneous groups of students using GI in a middle-class

Israeli junior high school had higher achievement scores than students taught via the traditional whole-class method. The subjects were 351 eighth-grade students studying history and geography (197 in GI and 154 in whole-class treatments), and although the school was in a middle-class neighborhood, 33% of the subjects were bused in from lower-income neighborhoods. Students were categorized as either Western or Middle Eastern. Five out of nine classes were taught via GI, and four were taught via the whole-class method. During the study, the learning groups were reconstituted six times to increase the academic heterogeneity. As a result, academic homogeneity in the groups declined steadily as the ability level became more heterogeneous. The history exam consisted of eight questions (four of which were more cognitively demanding the others, according to Bloom's Taxonomy). A team of four teachers estimated that the eight questions would take 45 minutes to answer. There were 14 questions on the geography exam (divided similarly in half according to high and low cognitive-demand) and the teachers estimated that it would take about the same amount of time to answer the questions. Two teams, each comprising of three outside teachers, were brought in to determine the exact criteria for grading the exams. A large number of students in both learning settings did not answer two specific questions on the geography test and one specific question on the history test, so the researchers eliminated those questions from the final analysis. Because the results from both tests were very similar, only the results from the history test were reported in the study.

A first ANCOVA was used to determine if there were any differences

between achievement in classes in which teachers had attended a cooperative learning workshop but taught with the traditional methods, and teachers of the cooperative classes who did not attend the cooperative learning workshop. No differences were found. Classroom-level analysis revealed that the mean achievement gain of every GI class was higher than every traditionally-taught class, even though there were significant differences in method (three-way interaction effect of Method x Class x Time, $F = 136.79$, $p < 0.001$). Another 2 x 2 x 2 split-plot design ANCOVA was used to determine the results of both ethnic groups' achievement, and although the Western students had higher scores than the Middle Eastern students in both settings, the Middle Eastern students in the GI setting had significantly higher scores than the Middle Eastern students in the whole-class setting ($F = 129.26$, $p < 0.001$), on average scoring twenty points higher between pretest and posttest.

The high F -scores in all areas and repeated use of ANCOVA to disentangle outside variables give the study a high level of internal validity, and the similarity of the researchers' findings to other studies of GI give the study reliability and external validity (Tan et al., 2005). The validity of these findings increases because of the researchers' efforts to balance the potential effects of teachers' training. The methods were transparent and well-documented, and the article was clearly written so that all of the information was easily accessible. The results show GI to be highly effective in teaching all students, particularly students of ethnic minority and lower socioeconomic status.

Tan, Sharan, and Lee's (2007) research on the effects of GI compared to

whole-class teaching among 241 eighth-grade geography students in Singapore with mixed groups of high- and low-achieving students found that high-achieving students had improved test scores, but that low-achieving students' scores declined slightly. Seven classes from a total of two schools participated, and students were divided into a whole-class control treatment ($n = 103$) and an experimental GI treatment ($n = 138$). The method by which students were divided was not disclosed in the article. Three teachers from the two schools that participated were trained in six three-hour workshops conducted by Dr. Shachar, which were held four months prior to the study. The researchers had five hypotheses, of which only two are relevant to this paper. The first was that "the group investigation method will have a more positive effect on students' academic achievement than does the traditional whole-class method of instruction" (Tan et al., 2007, p. 143). The second was that "the group investigation method will have a more positive effect on the success of lower achieving students than of higher achieving students as compared with the traditional whole-class method of instruction" (p. 143). A pilot study was conducted prior to this one, and words and ideas on the test that were commonly misunderstood were changed. The experiment lasted six weeks.

The students in both the control and experimental treatments studied two units, one on pollution and one on climate change, and took an achievement test after completing each unit. A two-way ANCOVA was performed on the means, which showed a significant main effect for achievement level, $F(1,237) = 23.85$, $p < 0.001$. The high-achieving groups scored significantly higher than the low-

achieving groups in both the experimental and control groups. There was no significant method X achievement level effect and no significant effect for method, meaning that the findings are inconsistent with both of the aforementioned hypotheses, and that GI was not more effective than the whole-class method. Additionally, although high-achieving students in the GI treatment improved their scores, low-achieving students' scores decreased slightly. This is also inconsistent with the hypotheses that GI would have a positive effect on student achievement than traditional whole-class instruction and that GI would have a more positive effect on the success lower achieving students than higher achieving students. After the researchers cited that the students in the GI group did not have significantly lower scores than the control group, they analyzed three factors that potentially affected the lower GI scores.

The first factor was student learning characteristics. After citing that most of the earlier studies on group investigation (Lazarowitz & Karentsky, 1990; Shachar & Sharan, 1994; Sharan & Hertz-Lazarowitz, 1980, Sharan, S., Kussell, Bejarano, Raviv, Hertz-Lazarowitz, & Sharan, Y., 1985, cited in Tan et al., 2007) had significant positive effects on student achievement and that, as expected, the higher-achieving students' scores improved significantly in both settings, the researchers used data from student surveys to determine that students were accustomed to passive learning, in which they took notes and prepared for tests. Students wrote that the GI method took too much of their time and that they did not have time to prepare for other classes and examinations .

The second factor cited by the researchers was student motivation.

Previous research (Johnson & Johnson, 1985; Qin, Johnson, & Johnson, 1995; Sharan, 1980; Slavin, 1980, 1983, 1995, cited in Tan et al., 2007) had yielded positive effects on academic achievement anywhere from two days to two weeks after implementation. Slavin (1985, cited in Tan et al., 2007) later stated that cooperative learning needs to be in place at least four weeks to be effective. Therefore the researchers decided on six weeks as the duration of their study. Even so, after implementation and execution students expressed a preference for teacher-centered instruction.

This leads to the third factor cited by the researchers: school and classroom organizational norms. The authors suggested that because the teaching method was not part of the school norm, nor part of the teachers' existing repertoire, student behavior and attitude toward it may have been affected. Because the instructional norms of the school are interconnected, the researchers suggested, a longer-term duration of the study in only one academic content area may not have a positive effect on academic achievement. Students have similar demands and expectations in different classes. Therefore, the researchers suggested that if GI is to have a greater positive effect on academic achievement, it will need to be considered in the context of a greater educational change. And if it is to be systematically researched, classroom norms of organization and evaluation must be changed.

The methods of analysis are well-founded, and ANCOVA was used to disentangle potential outside variables, meaning that the statistics themselves are reliable. The achievement results, however, are somewhat inconsistent with

previous research. Shachar and Sharan's (1994) study (reviewed above) showed that GI had a significant positive effect on gains in achievement. There is no reason to suspect that the research of Tan, Sharan, and Lee (2007) is not internally valid. Instrumentation was reliable and analyses were transparent. Interrater reliability of the test questions was sufficient, and the six-week duration of the study was sufficient for analysis without a major risk of maturation of the subjects affecting the outcome. The duration of the study, as is mentioned above, has also been proven sufficient in other contexts. Moreover, the researchers were thorough in examining confounding factors that may have affected student achievement, such as student learning characteristics, exposure to group investigation, and school and organizational norms.

Stevens, Madden, Slavin, and Farnish (1987) performed two similar but separate quasi-experimental studies on the effects of CIRC on academic achievement in reading and writing; they found significant effects for both. The first study's subjects were 461 third- and fourth-grade students in a suburban Maryland school district. The mean percentage of minority students in the sample population was 16.1% Ten classes served as the control group and 11 as the experimental group during the twelve-week duration of the study. The experimental classes were matched on the California Achievement Test (Total Reading battery) with 10 classes from four other control schools to ensure that the students were of similar academic ability levels to check for equivalency. The teachers of control-group classes continued to use their traditional methods while the experimental teachers were trained in CIRC methods for a total of six hours.

All of the pretest and posttest materials were taken from the California Achievement Test and were standardized with z-scores to ensure that grade-level differences did not affect the outcome. Individual-level ANOVAs found that there were significant pretest differences favoring the control group students for Total Language, $F(1,459) = 9.13, p < 0.003$, and Writing Mechanics, $F(1,459) = 10.61, p < 0.002$. Class-level posttest ANOVAs found significant differences in four of the five categories in favor of the experimental group: Reading Comprehension, $F(1,19) = 4.85, p < 0.04$; Reading Vocabulary, $F(1,19) = 4.62, p < 0.05$; Language Expression, $F(1,19) = 4.45, p < 0.05$; and Spelling, $F(1,19) = 11.29, p < 0.003$.

This first study was highly objective and transparent. Thorough analyses and strong *F*-scores showed that there was a significant improvement in academic achievement among students in the experimental group. The study by Stevens et al. (1987) demonstrated that this first study's results were externally valid and reliable.

In Stevens, Madden, Slavin, and Farnish's (1987) second study, the methods were roughly the same, but student demographics and the duration of the study differed. The subjects were 450 third- and fourth-grade students in a suburban Maryland school district, with nine experimental classes and 13 control classes. The researchers attempted to control for ethnic and socioeconomic background by selecting matched classes from schools in the same or similar neighborhoods. The mean percentage of minority students in each class was 22.4%. The study lasted for six months of the 1985-1986 school year. The same

pre- and posttests from the previous study were used, and again the mixed grade-level scores were standardized to z-scores. Unlike the previous study, no significant differences were found between the groups' pretest scores.

The class-level analyses found significant differences favoring the experimental group on three segments of the posttest: Language Expression, $F(1, 20) = 4.76, p < 0.042$; Language Mechanics, $F(1, 20) = 7.57, p < 0.012$; and Reading Comprehension, $F(1, 20) = 12.86, p < 0.002$. The Reading Vocabulary segment, however, did not show significant differences, $F(1, 20) = 1.09, ns$.

As in Stevens et al.'s first study, there are positive significant effects on academic achievement for the experimental groups; experimental groups had higher levels of academic achievement between the pre- and posttests. This second study was equally thorough and transparent. Although the sample population was comprised of more diverse demographics and the study occurred over a longer duration, the results were roughly the same, meaning that they are generalizable and have some degree of external validity.

Stevens, Slavin, and Farnish (1991) conducted a similar study a few years later and found similar results. In their study on the effects of CIRC mixed with direct instruction, pure direct instruction, and traditional teaching through basal readers on the academic achievement of 486 third- and fourth-grade students in four elementary schools from an ethnically diverse medium-sized city in central Pennsylvania, the researchers found the mixed CIRC and direct instruction treatment had the most significant effect on students' posttest scores. The 30 classes of students were randomly assigned to one of three

counterbalanced treatment groups, meaning that the order of the treatments was changed for each group to account for the possible influence of external factors. The 30 teachers involved volunteered for the study. The experiment lasted four weeks, and the same amount of time was given to teaching reading in each class based on grade level: 90 minutes each day for third-grade and 60 minutes each day for fourth-grade – both for four days each week. In the four schools, African American students made up 56.3% to 85% of the sample population.

In the experimental treatments, instruction focused on recognizing the main idea or thematic elements of a passage. These two experimental treatments were direct instruction with cooperative learning (CL), and direct instruction (DI). The control group teachers used traditional whole-class readers and basal readers, and employed their traditional teaching methods, meaning that instruction between the classes varied greatly. Both the CL and DI treatment groups used CIRC materials for main idea comprehension. When students in the CL treatment were consistently answering the questions correctly, the students moved on to independent practice and exchanged papers for peer assessment. Following the CIRC model, if a student made one mistake, the independent practice and peer review cycle was repeated. If a student made more than one mistake, the teacher provided remediation before the cycle was repeated. Students who made no errors went directly to the mastery test, and students who made errors needed to correct them before taking the mastery test. Afterward, their scores were used to determine team scores.

The DI group, like the CL group, spent about half of their reading time

four days a week using CIRC reading comprehension materials related to main idea skills. The DI students did not have the group practice time and moved on to the questions after the teacher delivered the lesson. After students completed about 1/2 to 1/3 of the questions, the teacher paused to check answers, provide feedback, and reteach any content if necessary. Once the students completed the practice activities they were given the same mastery test as the CL groups. The control whole-class groups used basal readers in reading groups and completed related workbook activities as a follow-up. The skill-instruction and follow-up activities in the control groups therefore varied greatly from day to day. Most of the activities involved comprehension assessment rather than explicit instruction of particular comprehension skills, and main idea comprehension was one of the many skills covered over the four weeks.

All students were given a 30-item multiple-choice pretest. The test was a 10-paragraph passage with a detail, inference, and main idea follow-up question at the end of each paragraph. A 20-item, multiple-choice, 10-paragraph posttest with a main idea question and inference question was given to the students at the end of the four-week study. The reliability of the tests was determined by using a randomly selected subsample of 90 students, or three per class. The alpha coefficients of the pretest were 0.71 for details, 0.71 for main ideas, and 0.62 for inferences. The alpha coefficients for the posttest were 0.80 for main idea and .077 for inferences. Because the students were from two grade levels, the test scores were converted to standard z-scores for comparison. Both of the experimental groups had higher achievement gains in main idea comprehension

than the control group, $t(25) = 4.45$, $p < 0.001$. However, there were no statistically significant differences between the two experimental groups. The DI mean score was higher than the CL group, but the difference was not statistically significant. Inference questions and school-level effects were also examined and were not found to be significant. It is difficult to determine the study's degree of internal validity because the control group methods varied so greatly. Such variability in teaching methods reflects the reality of reading comprehension instruction, but such instrumentation may not serve as a reliable control in a quasi-experimental study such as this.

The final article on CIRC reviewed in this paper is Calderón, Hertz-Lazarowitz, and Slavin's (1998) study on version of CIRC that was adapted to aid in transitioning limited English proficiency (LEP) bilingual students from bilingual instruction in both Spanish and English to instruction in English. The adapted version was called bilingual cooperative integrated reading and composition (BCIRC), and the researchers found that subjects who participated for one year had significantly higher achievement scores than their comparison group counterparts. The study was conducted in a large school district in El Paso, Texas. In the district, 79% of students were Hispanic, 27% were LEP, and 33% qualified for Title I services. The seven schools that participated were divided into three experimental and four comparison schools, and all seven of the schools were among the most impoverished in the district. The subjects were 222 students in bilingual programs. The BCIRC experimental schools and the comparison schools were similar in demographic characteristics and academic

ranking within the school district. The BCIRC schools were slightly smaller than the comparison schools and had a higher percentage of students that qualified for Title I services. Classes in the experimental BCIRC schools were individually matched with classes in control schools according to mean pretest achievement scores.

The project began with a pilot cohort of second graders in 1989-1990, then followed additional second-grade cohorts in 1990-1991 and 1991-1992. Scores for all students were collected in 1992, when one cohort had finished second grade and the other had finished third grade. Teachers in the comparison group used traditional textbook reading methods, such as round-robin reading and independent workbook activities. Comparison group teachers received training in CIRC but not BCIRC, and although they sometimes used cooperative methods, they never did so on a sustained basis. Students in the comparison classes received 90 minutes of reading/language instruction a day as well a 30-minute English as a second language (ESL) class.

Students in the BCIRC classes used the same texts as the students in the comparison classes, but were taught through the BCIRC process. The BCIRC classes incorporated the 30-minute ESL class so that students were taught for an uninterrupted period of 120 minutes. Although BCIRC was based on the CIRC model, it incorporated several additional factors to be adapted to bilingual instruction: extensive constructivist-based teacher development; integration of first-language principles and theories; integration of second-language principles and theories; integration of transitional Spanish-to-English

reading and writing methods and techniques; adoption of student-centered, constructivist philosophy. The BCIRC method was comprised of a total of 15 activities that took place before, during, and after reading: (1) building background and vocabulary, (2) making predictions, (3) reading a selection, (4) partner reading and silent reading, (5) story comprehension treasure hunt, (6) story mapping, (7) story retell, (8) story-related writing, (9) words out loud and spelling, (10) partner checking, (11) meaningful sentences, (12) tests, (13) direct instruction in reading comprehension, (14) writing workshops, and (15) independent reading.

Students' academic achievement in both the comparison and experimental groups was analyzed according to three measures: (1) Bilingual Syntax Measure (BSM), (2) Texas Assessment of Academic Skills (TAAS), and (3) Norm-Referenced Assessment Program for Texas (NAPT). BSM scores are a measurement of students' oral abilities in English and Spanish. Scores range from 1 to 5. Students take the test upon entering school and for several years after to determine whether or not they are assigned to bilingual or monolingual education classes. Because many of the students who participated in the study had BSM scores from kindergarten and first-grade, the scores were used as covariates for analyses of later test data. The TAAS is a state-mandated criterion-referenced assessment of reading, writing, and mathematics that was administered to all students at the beginning of third-grade. Students whose dominant language was Spanish took the test in Spanish. The NAPT was also administered at the end of third grade. It is a norm-referenced assessment of

reading, writing, mathematics, science, and social studies in English.

An ANCOVA was used to analyze second-grade BCIRC and comparison classes' reading and writing scales from the TAAS, using kindergarten English and Spanish BSM scores as covariates. An ANCOVA was also used to compare third-grade reading and language NAPT scales, again using kindergarten BSM scores as covariates. Before performing these analyses, however, the researchers analyzed the comparability of the BSM scores from the end of kindergarten to use as baseline data for the second- and third graders. There were no significant differences between the BCIRC and comparison group second graders in either Spanish or English at the end of kindergarten. The only significant finding was that third-grade comparison group students ($N = 38$) had a slightly higher Spanish BSM score ($M = 4.39$, $SD = 0.71$) than the BCIRC group ($N = 64$, $M = 3.95$, $SD = 1.04$) at $t = 2.30$, $p < 0.02$.

The Spanish TAAS scores for second graders' reading and writing showed significantly higher scores for the BCIRC group. In reading, the BCIRC group ($N = 51$) scored $M = 1,522$ ($SD = 200$) and the comparison group ($N = 42$) scored $M = 1,434$ ($SD = 246$), $F = 3.44$, $p < 0.06$. In writing, the BCIRC group ($N = 51$) scored $M = 1,522$ ($SD = 240$) and the comparison group ($N = 42$) scored $M = 1,441$ ($SD = 179$), $F = 5.44$, $p < 0.02$.

The English norm-referenced NAPT scores for third graders' reading and language showed significantly higher scores for the BCIRC group in reading, but not in language. The BCIRC group ($N = 52$) scored $M = 32.36$ ($SD = 15.44$) while the comparison group ($N = 33$) scored $M = 23.54$ ($SD = 14.98$), $F = 7.14$, $p <$

0.01. These findings indicate that students in the BCIRC group in both second- and third-grade had significantly higher levels of achievement than students in the comparison group.

The researchers also compared the effects of the duration of time spent in the BCIRC program for third graders (one and two years) with the comparison group and found that students who spent two years in the BCIRC program ($N = 26$) scored $M = 36.83$, students who spent one year in the program ($N = 26$) scored $M = 28.83$, and students in the comparison group ($N = 33$) scored $M = 23.83$, $F = 4.27$, $p < 0.01$, indicating that BCIRC students initially have higher achievement than comparison group students in the first year, and that their achievement increases even more significantly in the second year of BCIRC.

Finally, the researchers compared the number and percentage of third graders who transitioned out of the bilingual reading and writing programs. Of the 66 students in the BCIRC group, 32% of them ($n = 21$) met the reading level criteria to exit the program, whereas in the comparison group 10% of the 52 ($n = 5$) participants met the reading level criteria ($\chi^2 = 7.11$, $p < 0.01$). Similarly, 39% of BCIRC students ($n = 26$) met the language level criteria to exit the program, while 21% of the comparison group students ($n = 11$) met the criteria ($\chi^2 = 3.53$, $p < 0.06$).

Overall, the findings strongly indicate that students who participated in BCIRC learning generally had higher reading, writing, and overall language scores on both criterion- and norm-referenced assessments than comparison group students who were taught using traditional whole-class methods.

Therefore, significantly more BCIRC students met the criteria to exit the bilingual education program and enter the monolingual one. The variety of measurements and analyses employed by the researchers gives the study a high degree of internal validity, while the two-year duration of the study allows for deeper analyses of student growth in the BCIRC program. Little research has been conducted on BCIRC, but the program's similarity to CIRC and the similarities among the findings add to the reliability of the method (Stevens et al., 1987; Stevens et al., 1991).

More recently, Durukan's (2011) pre- and posttest quasi-experimental study of the effects of CIRC on 24 seventh-grade Turkish students' academic achievement showed that they had significantly higher reading and writing achievement scores than did the 21 control-group students taught via traditional whole-class instruction. Students were randomly assigned to the control group ($n = 21$) and experimental group ($n = 24$). The study lasted for five weeks and students worked together two hours each week. During the first week, students were divided into four- to six-member mixed-ability groups. During the second week, groups prepared for their cooperative work by creating team names and slogans. During the third week, students were divided into pairs and given sentences to read. They were asked to read the sentences in pairs, correct them if necessary, and focus on pronunciation and stress patterns. Then students reconvened in their CIRC groups and were asked to answer two reading comprehension questions as a group. During the fourth week, students were asked to copy sentences written by the teacher to improve their writing skills. The

groups then sent a copyist to the board to write down group answers. Groups were encouraged to critique each other's work. During the fifth week, students were asked 10 questions about the reading and writing activities they had completed, and the group with the most correct answers was given an achievement certificate. To formalize the data, a Written Expression Achievement Test (WEAT) and Reading Comprehension Achievement Test (RCAT), both developed by the researcher, were used to collect data related to the study groups' writing skills and reading comprehension skills. The RCAT test was found to be 0.79 reliable according to the KR-20 formula, and the WEAT test was found to be 0.85 reliable according to the KR-20 formula.

Both groups were given the RCAT and WEAT tests as a pretest, posttest, and retention test. Both tests were scored on a scale of 25 points and translated into percentages. The experimental group's RCAT pretest score was 13.42 (53%), their posttest score was 23.29 (93%), and their retention test score was 19.92 (75%). The control group's RCAT pretest score was 13.52 (54%), their posttest score was 19.95 (80%), and their retention test score was 16.86 (67%). The experimental group's WEAT pretest score was 12.13 (49%), their posttest score was 22.54 (90%), and their retention test score was 19.08 (76%). The control group's WEAT pretest score was 13.00 (52%), their posttest score was 20.62 (82%), and their retention test score was 17.05 (68%). Standard deviations for all were between 1.55 and 2.54.

The RCAT findings are significant for both groups regarding the common effect of being in different student groups and different measurement periods,

$F(2,86) = 25.216, p < 0.05$. Analysis of the WEAT scores yielded similar findings, $F(2,86) = 22.204, p < 0.05$. In sum, although there were slightly higher arithmetical scores for the experimental group on both tests, that fact in itself does not determine the significance of the method and test-period effects. However, F -tests did find that both methods were significantly effective, and that the experimental CIRC method was more effective in increasing academic achievement.

.The study's demonstration of the effectiveness of CIRC, when compared to similar studies, strengthens the external validity of the method (Calderón et al., 1998; Durukan, 2011; Stevens et al., 1987; Stevens et al., 1991). However, the article gives little detail on the content that students were taught and the CIRC activities used for instruction during the duration of the five-week study. For example, the article states "Taking into consideration various student characteristics such as sex, achievement, interest, skills, age, and culture, the experimental group was divided into 46 [*sic*] member groups according to CIRC technique." The vague description of the instructional procedures, as well as the fact that the control group also had significantly increased academic achievement throughout the study, calls into question the instrumentation and, hence, the internal validity of the research methods. Because both groups had statistically significant increases in reading comprehension and written expression, it is difficult to determine which factors of CIRC, if any, led to the slightly higher rates of achievement.

In Stevens' (2003) research on the effects on academic achievement of

implementing a student team reading and writing program (STRW) in high poverty schools middle schools in the eastern United States, with a sample size of 3,916 students (80% of whom were students of ethnic minorities), the researcher found that while the comparison students in the traditionally-taught group had significantly higher pretest scores on total reading and total language, posttest results indicated that the experimental classes had significantly higher achievement on measures of reading vocabulary, comprehension, and language expression at the end of the one-year duration of the study. The two experimental schools were matched with three comparison schools on their initial achievement in reading and language arts on the California Achievement Test that had been administered by the school district. There was also an attempt to match the schools by ethnicity and socioeconomic background of the students. Students took the reading and language grammar sections of the pretest only. The pretests were given to match the initial achievement and as a covariate to increase the power in the analyses of the outcome data. The posttest administered as part of the study was parallel in form to the pretest.

During the one-year duration of the STRW program, students in the experimental group participated in a reading and writing activities, and in all of these activities they worked in heterogeneous learning teams. Within heterogeneous student teams, students were assigned a partner to work with after they had completed their activities. Individual student scores contributed to an overall team score. The reading part of the program was comprised of three main components: "literature-related activities, direct instruction in reading

comprehension activities, and selection-related writing” (Stevens, 2003, p. 142). Student groups did partner reading, a variety of comprehension activities, word mastery activities, summary activities, and selection-related writing. For the writing part of the program, students were initially taught about the writing process: planning, drafting, revising, editing, and finalizing. The process was modeled and students actively participated in each step. The students worked with peers in each of the steps to incorporate the cooperative learning process into the writing process. After students learned the writing process, they had writing concept lessons and integrated English language arts lessons, both of which were designed to allow students to develop further skills to improve their writing.

There were significant differences on the pretests of total reading ($F = 11.2, p < 0.01$) and total language ($F = 54.2, p < 0.01$). In both cases the comparison students had significantly higher initial achievement than the experimental students. No possible explanations were offered as to why the comparison students had higher pretest scores. MANCOVA results from the posttests indicated that the experimental classes had significantly higher achievement on measures of reading vocabulary ($F = 4.32, p < 0.05$), reading comprehension ($F = 3.95, p < 0.05$), and language expression ($F = 5.74, p < 0.05$).

These results support the researcher’s hypothesis that restructuring English language arts programs in high-poverty schools through the use of research-based instruction, elements of cooperative learning, and integrated

reading and writing instruction can effect a significant increase in academic achievement. Stevens cited the use of cooperative learning and use of the writing process as particularly important elements in actively engaging students, which in turn causes them to take more responsibility for their own learning and leads to gains in achievement.

The findings on the positive effects of STRW on academic achievement are highly significant. The strong *F*-scores and large sample population add to the study's internal validity. Although the study took place over the course of the entire year, maturation is not a threat to its validity because of the presence of a comparison group. Stevens' (2003) research the first study ever conducted on STRW and few, if any, similar studies have been conducted since then. Therefore, it is difficult to evaluate the reliability of the findings. Despite this, overall, it is a strong study, indicating that the implementation of STRW in high poverty schools can lead to significant gains in academic achievement.

Johnson, Johnson, and Tauer (1979), in investigating the effects of interpersonal cooperation, competition, and individualistic effort on students' attitudes and academic achievement, found that cooperative learning had significant positive effects on achievement. The subjects were 69 fourth, fifth, and sixth graders (32 males; 37 females); males and females were randomly assigned to conditions so that within each condition there were 23 students and an approximately equal number of males and females. Students were from a large Midwestern city in the United States. Students in all three conditions studied introductory geometry; none had studied geometry before.

The independent variable was the goal structure: cooperative, competitive, or individualistic. Students in the cooperative class were instructed to work together as a group, making sure that all members were involved and understood the material, and to seek help from fellow students rather than from the teacher. Students in the competitive class were put into different groups every day, reinforcing the idea that all students had a chance of winning. Competitive-treatment students were instructed to work alone and compete within their groups. Individualistic-condition students were instructed to work on their own, avoid interaction with other students, and seek help and clarification from the teacher. The teacher was instructed to praise and reward them individually. Students were instructed to work at their own pace, doing as much of the assigned material as possible in the time allowed. The two dependent variables were (a) student attitudes toward perceived acceptance by the teacher, peers, and oneself and (b) achievement. Achievement was measured by a teacher-made test consisting of 15 problems of increasing complexity, involving the material taught during five instructional periods. Three teachers participated in the study. All of them had been given 30 hours of training on how to conduct cooperative, competitive, and individualistic lessons and had spent over 20 hours teaching with each goal structure prior to this study. Teachers were then randomly assigned to conditions and then taught the same class for the five-day duration of the study.

The results of the achievement test indicated that the mean achievement score for the cooperative group was 13.68, the mean score for the competitive

group was 9.18, and the mean score for the individualistic group was 7.39 ($F = 37.03$, $p < 0.01$). However, as the authors' suggest, these findings are limited by the size and nature of the sample, as well as by the lesson content used. Still, the cooperative group had a higher level of achievement than the individualistic and competitive groups.

One weakness of the study is the lack of a pretest to determine whether the treatment groups were reasonably similar in academic ability. One possible reason for this is that the study was also concerned with student attitudes and perceptions of the teacher and fellow students depending on different goal structures, and that the effects of cooperative learning on academic achievement was but one of a few dependent variables. The high reliability of the F -scores, despite the lack of a pretest, may still be valuable in determining the effectiveness of the methods. The small sample and relatively short duration of the experiment are a weakness of its instrumentation, and therefore threaten the internal validity of the research.

In a quasi-experimental, posttest-only study, Phuong-Mai, Terlouw, Pilot, and Elliott (2009) hypothesized that a cooperative learning strategy emphasizing culturally appropriate leadership, reward allocation, and group composition while studying a variety of subjects would yield higher learning outcomes among Vietnamese high school students than a traditional cooperative learning scenario that did not take cultural characteristics into consideration. Analysis showed, however, that although the experimental group scored higher than the control group, the only statistically significant difference was the score in what the

researchers termed social science, and thus the hypothesis was rejected.

Four classes comprising a total of 181 students aged 16-18 were involved in the study, which covered several academic subjects. Students were volunteers and were paid a small sum for participating. Each class was randomly divided into two experimental settings. One group of students was given CL lessons designed to be more culturally appropriate to Vietnamese learners in that they had a formal leader and that group composition was based on friendship. The second experimental group was given the same content and lessons, but the groups had no formal leader and the group composition was based on mixed-ability grouping. In both treatment groups, students were given group grades, not individual grades. The segment of the experiment applicable to academic achievement consisted of 10-12 lessons that each lasted 45 minutes. Students were assigned to complete tasks related to, as the researchers described, different social science (ESL, literature, history) and natural science (geometry, applied physics) disciplines. These were referred to as task cluster activities. Students were also assigned to complete a group project, but this project was not relevant to the researchers' researchers on academic achievement. The researchers' hypothesis of academic achievement was, "When CL is organized in Viet Nam in a fashion that is culturally appropriate in terms of leadership, reward allocation and group composition, Vietnamese students will have learning outcomes superior to those in conventionally operated groups" (Nguyen et al., 2009, p. 862) In the task cluster activities, the first group of students scored higher than the second group, but the only statistically significant difference

found was for the social science cluster, so the hypothesis was rejected. For the social science activities, group one scored 8.2 ($SD = 1.3$) and group two scored 7.3 ($SD = 1.8$), $t = 2.2$, effect size (d) = 0.58, $p = 0.02$.

Although the results of the study indicate that the researchers' hypothesis was rejected, there are still implications relevant to this review of the literature. Despite the lack of significant difference in achievement in the natural sciences, there was a significant effect in social science achievement gains. Despite these significant results in the findings, there are several weaknesses in the study. Firstly, it is a posttest only study, so the researchers never empirically demonstrated that the subjects in both groups were of similar ability levels before the treatments were applied. Therefore it is impossible to evaluate the degree of internal validity because it is impossible to determine if the treatment had an effect. Secondly, the differences between the two experimental groups are too numerous to determine which factors led to a difference in achievement. Because leadership structure and ability-grouping were different in both groups, it is unclear which variables led to the significant outcomes. This also makes it impossible to determine whether the independent variable had an effect, again making it impossible to evaluate the internal validity of the study.

The first section of this literature review examined the effects on academic achievement of individual accountability structures, mixed individual-plus-group accountability structures, and group accountability structures of cooperative learning. The majority of the studies examined found significant effects on gains in academic achievement. Slavin's (1978) study on the effects of

type of recognition (recognition based on the performance of a four- to five-member cooperative learning team and recognition based on individual performance) on student achievement yielded no significant results. However, all three of the individualistic accountability structure studies critiqued in this review found significant effects on increased academic achievement. In addition, five of the six articles included in the second subsection on individual-plus-group accountability structures found significant effects on gains in academic achievement. Finally, of the eight studies on group accountability structures of cooperative learning, all found significant gains in academic achievement. However some of them had discrepant results, such as Tan et al.'s (2007) study on the effects of GI on high- and low-achieving students' academic achievement. In this case, high-achieving students had significantly higher posttest scores, but low-achieving students had slightly lower posttest scores. Likewise, Phuong-Mai et al. (2009) found statistically significant effects for a culturally responsive form of cooperative learning on gains in academic achievement; however, the results did not completely support their hypotheses. Chapter three considers potential classroom implications based on these studies, and provides suggestions for future research.

Effects of Cooperative Learning on Social Interaction, Behavior, and Affect

This second section of the paper on the effects of cooperative learning on social interaction, behavior, and affect is divided into three subsections. The first subsection is dedicated to the effects of previous exposure to cooperative learning on behavior. The second subsection contains research on attitude,

perception, and preference towards cooperative learning. The third subsection is a review of research on attitudes and behavior towards others in cooperative settings.

Effects of Previous Exposure to Cooperative Methods on Behavior

The two studies examined in this subsection investigated the effects of previous exposure to cooperative learning methods on individual and group behaviors and on future academic achievement. Gillies (2002) conducted a two-year follow-up study on fifth-grade students who either had or had not been taught to work in cooperative settings in the past. The researcher found that students who had participated in cooperative activities in the past were more likely to stay on task and less likely to engage in non-cooperative behavior, such as behavior that demonstrates competition or opposition to criticism. The second study by Ryan and Wheeler (1977) found that students who had been taught to work in cooperative groups were more likely to voluntarily work together than students who had not.

Gillies (2002) conducted a quasi-experimental study to determine if 88 fifth-grade Brisbane students who had been trained in cooperative learning two years earlier and who had practiced cooperative learning in the prior year would perform better in a cooperative learning situation than students who had not been trained. Gillies found that students with prior training were more task-oriented, shared more resources, and provided more solicited explanations than their untrained peers. Gillies defines solicited explanations as “requests for help that were detailed and elaborate” (Gillies, 2002, p. 18). Fifty-two children who had

been trained in cooperative learning two years earlier and had participated in cooperative learning groups (with no additional training) the previous year were assigned to the trained condition. Thirty-six other children who had also participated the previous year in the untrained groups were assigned to the untrained condition.

The activities required comparison, classification, imagination, inference, hypothesizing, and generalization in primary social studies topics. Researchers observed the student groups and recorded behavior states and frequency of verbal interactions. Only the behavior states observed are examined here. The four behavior states examined were cooperative, noncooperative, individual task-oriented, and individual non-task-oriented. For the researcher to determine whether there were differences between the frequency of behavior states a MANOVA was conducted. Gillies (2002) found a significant effect on student behavior states in the cooperative learning groups, with specific observable effects in the areas of cooperation, $F(1, 86) = 31.28, p < 0.001$; noncooperation, $F(1,86) = 62.11, p < 0.001$; and individual nontask-oriented behaviors, $F(1, 86) = 17.9, p < 0.001$. Simply put, students in the trained cooperative groups exhibited higher levels of cooperation, lower levels of noncooperation, and lower levels of nontask-oriented behavior than their untrained counterparts.

These figures illustrate significant positive results for the trained cooperative condition. The strong F -scores and low p -values give the study a high degree of internal validity. The process and results are transparent and add to the study's objectivity. The article was clearly written and there are no

questionable or ambiguous aspects of it.

In Ryan and Wheeler's (1977) quasi-experimental quantitative study on the effects of cooperative and competitive background experiences on the play of a simulation game, the researchers found that when students with previous cooperative training were given the choice to compete or cooperate, they were far more likely to cooperate than the students from the competitive background. The authors compared subjects who had recently participated in cooperative learning experiences with subjects who had not done so; Ryan and Wheeler hypothesized that students who had participated would be more likely to: (a) seek help from one another; (b) positively respond to help sought; (c) volunteer help for others; and (d) establish group strategies. The subjects were 48 randomly selected fifth- and sixth-grade students from one suburban school in a large metropolitan area. Their mean reading, language, and mathematics scores on the Iowa Test of Basic Skills were above national norms. The socioeconomic status of the surrounding school neighborhood was judged by the investigators as middle to upper-middle class.

Prior to playing a simulation game, all subjects were involved in a series of 18 lessons focused on the Iban of Borneo and the Eskimo of northern Alaska. The subjects were randomly assigned to a cooperative or competitive group. After every five lessons, groups and individuals submitted a workbook for evaluation. In the competitive group, the six individuals with the best workbooks were rewarded with a poster, and in the cooperative group, the team with the best workbooks was rewarded, with a poster given to each member. After the 18

lessons, students played a board game called Seal Hunt, in which they were tasked with trying to find seals under the ice. Pre-existing cooperative groups of six remained intact, while members of the competitive group were assigned to new groups of six. In the game there were two ways to win: as a member of the winning hunting group or as the individual hunter with the most seal meat. There were, therefore, both group and individual goal structures.

The four hypotheses stated earlier were used to develop observational categories. Two independent raters tallied their observations as they watched the students on video tape. After six hours of training, the interrater reliability was 0.98. Mann-Whitney U tests for small group data were used to analyze the observational data for each of the four categories of the observational guide. Results indicated support for each of the four previously stated hypotheses ($U = 0, p < 0.05$ in each instance). The median score of all observational categories for the cooperative condition was 73.5 with an interquartile range of 65, while the median of all observations for the competitive condition was 8.5 with an interquartile range of 20.5. What this means for the cooperative condition is that the central 50% of the cooperative behaviors observed were within 65 points of each other and highly variable, but that the median score for the observed behaviors was still quite high. For the competitive condition there was far less variability in the scores, and the central 50% of the observations were within 20.5 points of each other, with the median score being 8.5. Clearly, there is a significant positive effect for the cooperative condition; the students engaged in more cooperative behaviors and, although there was far greater variability in their

cooperative behavior, there was a much higher median score.

There are some discrepancies in the study, however. For instance, the researchers stated that students with the best workbooks were rewarded. They did not specify what the expectations were for having the best workbooks. Because of this, there is a lack of objectivity and transparency. The study suggests that students act on acquired attitudes, even when they are not explicitly told to. The study does not examine whether factors such as socioeconomic status might have influenced the outcome, or whether similar results have been found in other studies, both of which weaken the study's instrumentation. Finally, although students in the cooperative condition were not instructed to work cooperatively when playing Seal Hunt, they were left in the same cooperative groups they had been working in previously. Perhaps this may have influenced their cooperative behavior. The instrumentation could have had a higher degree of validity if students had been assigned to different groups within the cooperative setting.

Both studies reviewed in this subsection yielded significant findings demonstrating that students who had previously been exposed to cooperative learning were more likely to work together even when not instructed to do so. The next subsection investigates students' attitude, perception, and preference toward cooperative learning.

Student Attitude, Perception, and Preference Toward Cooperative Learning

The five articles in this subsection focus on students' attitudes, perception, and preference toward cooperative learning methods. Dunn, Giannitti, Murray,

Rossi, Geisert, and Quinn (1990) found that students who learned in their preferred setting (individual or cooperative) had significantly higher scores than students who learned in a non-preferred setting, and that students with no preference had significantly higher scores when learning alone than when learning as part of a group. Owens and Barnes (1982) investigated the relationships between cooperative, competitive, and individualized learning preferences among 279 seventh- and eleventh-grade math and English students; they found that eleventh-grade students preferred learning through cooperation more than competition, and that seventh-grade students expressed a greater preference for individualized learning, but both groups expressed a preference for more cooperative learning in English than in math. The third article, by Tan, Sharan, and Lee (2005), is a qualitative study that found that both high- and low-achieving students had more positive than negative perceptions of the group-investigation method of cooperative learning. The fourth article, by Ghaith and Bouzeineddine (2003), revealed several significant findings about students' perceptions of cooperative learning. Among these findings was the discovery that lower-achieving students had more positive perceptions of cooperative learning than their higher-achieving peers. The final study, by Moskowitz, Malvin, Schaeffer, and Schaps (1983), is a quantitative study on the effects of Jigsaw on student attitude and behavior toward themselves, peers, and school. The study found that sixth-grade students in Jigsaw treatment classrooms viewed the classroom as significantly more cooperative than their control-group counterparts.

Dunn et al. (1990) conducted a quasi-experimental study on the effects of

matching and mismatching students with a preference for learning alone or learning with peers with selected instructional treatments in order to determine the impact upon their attitudes and achievement in social studies. The researchers identified the 108 sixth-, seventh-, and eighth-grade social studies students' learning style preferences through a learning style inventory assessment, then gave them an assignment to complete, a posttest, and a semantic differential questionnaire. The subjects were from two American middle schools – a small private Catholic academy and a large public school. The ethnic makeup of the Catholic school was 71% European American, 25% African American, and 4% Asian American; the families' socioeconomic status ranged from middle to upper class, and the academic performance of the subjects was average or above average. The ethnic breakdown of the public school was 80% European American, 15% African American, and 5% Asian American. Both schools had about the same socioeconomic demographics, and the students' achievement levels as measured by the Stanford Achievement Test (1987) were about the same.

Students' learning style preferences were identified through a 23-element Learning Style Inventory (LSI) assessment. The LSI identified six learning preference scenarios: learning (a) alone, (b) in a pair, (c) with peers, (d) as part of a team, (e) with a teacher, and (f) in a variety of ways. Reliability of the learning alone versus with peers element of the LSI instrument was 0.83. Using the findings from the LSI, students were classified into three preference groups: those with a preference for learning alone, those with a preference for learning

with peers, and those with no preference. Each group participated in both treatments: learning with others and learning alone. There were a total of four individual social studies lessons; students worked on two lessons in each treatment. The learning-alone treatment students worked on topic-driven packages that contained the objectives and all of the necessary information students would need to complete the assignments. They included objectives that were written directly to each student, alternative resources that allowed students to learn through the media they preferred, activity alternatives that challenged students to apply their knowledge by creating a project, reporting alternatives for students to share their projects, and the pre- and posttest directly related to the objectives..The learning-with-peers treatment was applied to small groups, and all members of each group were required to participate to complete the task. Members were randomly assigned in groups of five or six and were instructed to complete a “clearly stated assignment,” however details of this assignment are not present in the study (Dunn et al., 1990, p. 489). After completing their tasks, all students took a posttest and answered the Semantic Differential Attitude Questionnaire.

Repeated measures of ANOVA on attitudes showed an overall main effect for preference ($p < 0.05$), indicating that significant differences in attitudes were also reflected in the academic scores. Therefore, students who learned in their preferred groups had higher scores and more positive attitudes. Single-factor ANOVAs showed that in the learning-alone treatment, students who preferred to learn alone obtained significantly higher scores than learning-with-

peers and no-preference students, $F(2, 101) = 4.64, p < 0.01$. Also, in the learning-with-peers treatment, students who preferred the method had significantly higher scores than the learning-alone and no-preference students, $F(2, 101) = 3.02, p < 0.05$.

The findings of this study demonstrate that students who are allowed to learn in their preferred setting have significantly higher achievement scores, and that those who learn in a non-preferred setting have significantly lower achievement scores. The study also demonstrates that students with no preference learn significantly better when alone, however detailed achievement data is not given for this statistic.

The F -values of this study are highly significant and multiple analyses to disentangle variables were employed to ensure a high degree of reliability. The findings are transparent and highly internally valid, and corroborate previous findings on the relationship between student achievement and learning preference, also giving them an increased degree of external validity. However, as the researchers suggest, the study should be replicated to determine if, in fact, students with no learning-style preference do have higher levels of achievement when learning alone.

Owens and Barnes (1982) conducted an exploratory study utilizing a structured-response survey to investigate the relationships between cooperative, competitive, and individualized learning preferences and students' perceptions of classroom learning atmosphere. The subjects were 279 seventh- and eleventh-grade secondary math and English students from two high schools in New South

Wales that ranked fairly low on the socioeconomic scale. Owens and Barnes found that the eleventh-grade students preferred learning through cooperation more than competition, and the seventh-grade students expressed a greater preference for individualized learning, but both groups expressed a preference for more cooperative learning in English language arts classes than in math.

Owens and Barnes (1982) used the Learning Preference Scale – Students (LPSS) to obtain cooperative, competitive, and individualized learning mode preferences. A 36-item test with a four-point true-false scale was administered. The response categories were *completely true*, *more true than false*, *more false than true*, and *completely false*. Three of the 12 items were negatively phrased to compensate for the acquiescence effect, a term used to describe the way in which respondents are often led to agree with one-sided statements. Three main preference subscale scores were calculated, indicating preference for cooperative, competitive, and individualized learning styles. Although there were several other more in-depth analyses performed on the data sets, they were exploratory in nature and do not relate directly to the purpose of this paper, and shall therefore be excluded.

Results of the LPSS indicated that eleventh-grade students preferred learning through cooperation more than competition, and that the seventh-grade students expressed a greater preference for individualized learning. For school A ($N = 126$), the mean score of seventh-grade students' ($N = 75$) preference toward cooperative learning in English was 36.7 ($SD = 6.51$), while the mean scores for the competitive and individualized preferences were 35.2 ($SD = 5.43$) and 34.0

(SD = 6.17), respectively. The mean score of eleventh-grade students' ($N = 51$) preference toward cooperative learning in English was 39.6 (SD = 4.94), while the mean scores for the competitive and individualized preferences were 35.9 (SD = 5.16) and 31.4 (SD = 5.00), respectively. For school B ($N = 153$), the mean score of seventh-grade students' ($N = 87$) preference toward cooperative learning in English was 35.7 (SD = 5.70), while the mean scores for the competitive and individualized preferences were 35.3 (SD = 5.95) and 33.6 (SD = 6.58), respectively. The mean score of eleventh-grade students' ($N = 66$) preference toward cooperative learning in English was 39.0 (SD = 4.63), while the mean scores for the competitive and individualized preferences were 36.7 (SD = 5.47) and 32.4 (SD = 6.15), respectively. Overall, both grades at both schools had a preference toward learning English cooperatively, rather than individually or competitively, and both grades at both schools expressed a stronger preference for more cooperative learning in English language arts than math, even more so in school A than school B.

There are no obvious inconsistencies or ambiguities that bring the results into question. However, the findings offered here are limited because they only reflect the opinions of two grades of students in two academic subjects at two schools. Although the schools were matched as closely as possible, there were still variations among the results between them. For instance, students in both grades at school A expressed a stronger preference for cooperative learning in English than students in school B. Any number of confounding factors, such as school atmosphere or climate, could have affected this outcome. Furthermore,

the findings of the study reviewed in this paper are limited by their exploratory nature; students' preferences toward particular learning styles are measured and analyzed, but unlike a purely experimental or quasi-experimental study, there is not a pedagogical model as an independent variable and results cannot be directly attributed to it. Simply put, students were asked to rate their preferences for certain learning settings, but the researchers did not implement any of these settings directly. As stated earlier, the data included in this paper is preliminary, and the researchers used it in conjunction with several other sets of data not included here. So these are not limitations of the study as a whole, just limitations of how the study relates to the purpose of this paper.

The researchers discussed the traditional norms of higher achievement among boys in math and higher achievement among girls in English language arts, and cited theories of potential genetic developmental causes as well as other potential environmental causes that were prevalent in the literature of the day. Despite addressing these differences, the researchers did not take academic success or failure into account when analyzing student learning preferences. This could be seen as a weakness in the methodology – boys might have a preference for competitive learning in math because math classes are typically more competitive and individualistic in structure than English language arts, and boys typically are higher-achieving in math than in English language arts.

Tan, Sharan, and Lee (2005) performed a qualitative study on 142 eighth-grade geography students in Singapore by administering an open-ended

questionnaire on the students' perceptions of group investigation (GI). An analysis of the results showed that both high- and low-achievers had more positive than negative perceptions of GI. Four classes of eighth-grade geography students ($n = 142$) in two schools were taught two units of geography using GI. Two classes were from the high-achieving track ($n = 80$) and two classes were from the low-achieving track ($n = 62$). After the students participated in a GI setting for six weeks, the researchers administered an open-ended questionnaire to the students that simply asked what their thoughts and feelings on GI were. Tan and an assistant worked independently to code students' statements and achieved a 91.3% interrater reliability.

Tan et al. (2005) found that of 955 statements, 651 statements (68.2%) were positive and 304 (31.8%) were negative. High-achievers wrote an average 8.1 statements per student (a total of 651 statements), of which 450 (69.1%) were positive and 201 (30.9%) were negative statements. Low-achievers wrote an average of 4.9 statements per student (total of 304 statements) of which 202 (66.5%) were positive and 102 (33.5%) were negative. In the end, both groups had more positive than negative perceptions of GI. There were four significant categories of positive statements: (a) direct positive evaluation of GI, (b) positive perceptions of achievement regarding understanding and learning, (c), positive perceptions of social relationships, and (d) positive perceptives of learning skills and ability to pursue knowledge through GI. The high interrater reliability of 91.3% adds to the credibility of the study. The article was clearly laid out and easy to navigate, adding to its confirmability.

Ghaith and Bouzeineddine (2003) investigated the relationship between reading attitudes, achievement, and perceptions of cooperative learning on 111 English as a foreign language (EFL) students in Lebanon. Students completed two questionnaires and a semantic differential scale that measured their attitudes toward reading and perception of cooperative learning. They also took a pre- and posttest designed specifically to align with both their learning objectives and the study.

A week before the study, all eighth-grade students took the pretest for a duration of 30 minutes. The test had a variety of true/false, multiple choice, and short answer questions to assess comprehension of a passage, as well as questions to assess higher cognitive demands, such as making inferences and drawing conclusions. The posttest that students took at the end of the twelve-week study was also given for a period of 30 minutes and was of the same format. Additionally, students' attitudes toward reading were assessed prior to and at the conclusion of the study through a ten-item survey ($\alpha = 0.63$). Students' attitudes toward self and school were also assessed before and after the study through the use of a seven-item scale (internal consistency = 0.70). Finally, after participating in the study, students completed a semantic differential scale to assess their perception of cooperative learning (internal consistency = 0.81). Both teachers that participated were trained in the use of Jigsaw II and their average of full-time teaching experience was 5.6 years. The teachers attended a workshop facilitated by the researchers and developed all the learning materials themselves.

Analysis of the test, attitude, and perception data revealed several significant findings. First, the students' reading attitude was positively correlated with their attitude toward school and self ($r = 0.47$) and with achievement ($r = 0.44$). Second, the students' attitude toward self and school was found to have a positive correlation with achievement ($r = 0.44$). Third, students' experience of Jigsaw II cooperative learning did not have a positive correlation with any of the other variables measured. Of relevance to this paper are the findings that were revealed through the univariate analysis of covariance. Namely, that male students found their Jigsaw II experience to be more valuable than the female students $F(2, 81) = 16.99, p < 0.01$. The mean score for the males was 31.27 (SD = 4.91) and the mean score for the females was 29.25 (SD = 4.66). Also of relevance to this paper are the MANCOVA scores by achievement level indicating that there was a significant difference between high- and low-achievers, $F(5, 78) = 18.04, p < 0.01$. High-achieving students had a higher mean score (22.22, SD = 3.01) than the low achieving students (20.00, SD = 2.69), and high-achieving students were found to have more positive attitudes toward reading than low-achieving students, $F(2, 82) = 13.82, p < 0.01$. Likewise, low-achieving students were less positive toward themselves and school, $F(2, 82) = 9.81, p < 0.01$. The mean scores of the high-achieving students was 29.88 (SD = 3.01) and the mean score of the low-achieving students was 26.38 (SD = 4.60). Additionally, high-achieving students had higher reading comprehension scores than the low-achieving students, $F(2, 82) = 77.05, p < 0.01$. The mean score of the high-achieving students was 61.86 (SD = 7.26) and the mean score of the low-

achieving students was 42.70 (SD = 6.01). However, the analysis of the data on student attitudes found that the low-achieving students were more positive toward their Jigsaw II cooperative learning experience than the high-achieving students, $F(2, 82) = 17.24, p < 0.05$. The mean score of the high-achieving students was 29.57 (SD = 3.77) and the mean score of the low-achieving students was 31.07 (SD = 5.73).

This data indicates that reading attitudes, self and school attitudes, and reading achievement are positively internally related, meaning that students who have more positive attitudes toward reading are more likely to have more positive attitudes about self and school and are more likely to attain higher levels of academic achievement. The data also reveals that students' perception of their cooperative learning experience was unrelated to reading attitudes, attitudes toward self and school, and reading achievement, suggesting that other individual learning characteristics such as gender and aptitude may have affected students' perception of cooperative learning.

The study contained no questionable or ambiguous language or methodology, meaning the instrumentation gives it a high degree of internal validity. The findings of the study corroborate previous work by Ghaith (2001), who found that low-achieving learners are more comfortable in a small cooperative learning group than in a more competitive whole-class situation, adding to the study's reliability and overall external validity.

Moskowitz, Malvin, Schaeffer, and Schaps (1983) conducted a quasi-experimental study on the effects of Jigsaw on student attitude and behavior

toward themselves, peers, and school. Moskowitz et al. studied fifth- and sixth-grade students throughout the course of a year, and although the study revealed few significant effects on affect, the participants in the Jigsaw classes rated their classes as less competitive and the sixth graders rated their classes as more cooperative. The treatment group consisted of 147 students and the control group consisted of 114 students. They attended eight elementary schools in a middle-class suburban school district in northern California. The ethnic composition of the subjects was 94% European American, with Mexican Americans making up the largest minority group at 3%. The teachers who volunteered for the Jigsaw training attended two-hour classes for nine weeks, were paid \$200, and were offered postgraduate credit.

Although measures of academic achievement and teacher process evaluation were thoroughly analyzed, the measure pertaining to this review of the literature was student self-reports. As a pretest, students were given a self-observation scale and student questionnaire, which formed scales of affective teaching climate, attitudes toward school, academic and social self-esteem, attitudes toward peers, and locus of control for success and failure. The alpha reliabilities ranged from 0.56 to 0.91. A year later, posttest data was obtained with a revised version of the questionnaire that provided researchers with the same types of data as pretest scales. Measures of cooperation and competition in the classroom were also included. The reliability measures of the scales were 0.73 and 0.77, respectively.

After performing an ANOVA on the data, Moskowitz et al. (1983) found

significant effects for three variables. The authors suggested that more significant effects may have not been revealed because some of the classes were not seen as engaging in exemplary Jigsaw practices, and one class in the treatment did not perform Jigsaw activities at all, though the students did work collaboratively in groups. The reason for this was because, after attempting to implement Jigsaw, the teacher reported that students were engaged in too much off-task behavior; the teacher chose not to continue the treatment. The three significant effect variables were cooperative climate, competitive climate, and academic and social self-esteem. Significant pretest main effects for students in the exemplary Jigsaw classes were that they at first preferred the teaching climate, $F(1,165) = 10.82, p < 0.001$, and enjoyed school better, $F(1,166) = 14.89, p < 0.001$. An ANCOVA of posttest measures (except for cooperative and competitive climate) of exemplary design classes that used the corresponding pretests as covariates confirmed the findings, and also revealed that student attendance was significantly affected, $F(1, 156-172) = 7.17, p < 0.05$. Perhaps of most importance are the findings on cooperative climate. Compared to control treatment students, sixth-grade Jigsaw treatment students viewed their classroom as more cooperative, $t(247) = 4.31, p < 0.01$. However, fifth-grade students had a lower measure of academic self-esteem, $t(239) = 3.24, p < 0.01$.

The results of the study are somewhat perplexing in areas, while in others they are inconclusive (such as the measure of fifth-grade treatment students' self-esteem). The authors stated that some of the classes that employed the Jigsaw method either did not do so at an exemplary level, or did not do so at all.

Further analysis focused on the exemplary classes. The results of this study are not easily generalizable to the other studies in this subsection and are therefore of a low degree of reliability. Finally, as stated by the researchers, the results of the study are disappointing. Even though Jigsaw influenced students' impressions of their classroom environment, they were unable to find improvements in school attitudes and self-concept, even in the exemplary classes. The researchers suggested the most likely cause for this was that the teachers adapted Jigsaw in their own classrooms, and in doing so, possibly omitted crucial aspects of it.

The five articles reviewed in this subsection focused on student attitudes, perception, and preference toward cooperative learning. Dunn et al. (1990), Owens and Barnes (1982), and Tan et al. (2005) all found that students had positive attitudes toward the use of cooperative learning. Owens and Barnes (1982) found a correlation between gender, age, and class in student preference for cooperative learning; the prevailing finding was that eleventh-grade students preferred cooperation more than competition in both math and English language arts. Tan et al. (2005) found that the majority of both high- and low-achieving students that participated in their study of group investigation had far more positive than negative things to say about the GI instructional method. Ghaith and Bouzeineddine (2003) found that lower-achieving EFL students had more positive attitudes toward cooperative learning than their higher-achieving peers. And finally, Moskowitz et al. (1983) found that sixth-grade students viewed the Jigsaw classroom as significantly more cooperative than the control students, but

many of the findings were inconclusive. In conclusion, all of the studies report that part or all of the populations in each study expressed a preference for and positive attitude toward cooperative learning. The next subsection in this essay examines students' attitudes and behavior towards others in cooperative learning settings.

Effects of Cooperative Learning on Student Attitudes and Behavior Toward Others

The next five articles in this subsection address student behavior and attitudes towards others in cooperative settings. In addition to academic achievement, much research has been done on the effects of cooperative learning on student behavior and the ways in which students interact with one another. The first article in this subsection is Gillies' (2004) quasi-experimental study on the effects of structured and unstructured cooperative learning on students' behaviors, interactions, and learning, in which the researcher found that structured groups had a significantly higher effect on cooperative behaviors than unstructured groups. The second article in this subsection is Putnam, Markovchick, Johnson, and Johnson's (1996) quasi-experimental study on the effects of cooperative learning on general education students' acceptance of their special education classmates. The authors found significant improvements in the attitudes of regular education students in the cooperative group, but did not find significant improvements in students' attitudes in the competitive group. Weigel, Wisler, and Cook's (1975) quasi-experimental study on the impact of cooperative learning experiences on cross-ethnic relations and attitudes found that students

in the cooperative setting were three times as likely to help their classmates than their peers in the control setting, and that cross-ethnic helping was five times greater in the cooperative setting. The fourth article in this subsection is Hansell and Slavin's (1981) quasi-experimental study on the effects of cooperative learning on the structure of interracial friendship. The researchers found that students in the cooperative treatment increased cross-race friendships equally for students of different sexes, races, and achievement levels. The final article is on Bratt's (2008) quasi-experimental studies on the effects of Jigsaw on improving intergroup relations within multiethnic classes. Bratt found few significant effects and ultimately concluded that no favorable result of Jigsaw was evident.

Gillies' (2004) quasi-experimental study comparing the effects of structured and unstructured cooperative learning on behaviors, interactions, and learning among 223 ninth-grade students studying geometry in six Brisbane high schools found that structured cooperative learning groups had a more significant positive effect on cooperative behaviors than unstructured cooperative learning. Three of the schools had pre-established cooperative learning activities and routines available to the students at least a few times each week, across all subjects of the curriculum. These were the structured schools. The other three schools (the unstructured condition) used groups on an ad hoc basis – the groups were actually collaborative groups (rather than cooperative ones) because there had not been the necessary supports and training to develop the communication skills and positive interdependence of a structured cooperative

learning experience.

Initially, teachers rated their students' math achievement, and Gillies (2004) found no significant differences, $F(1, 205) = 0.32, p < 0.05$ (High commitment $M = 3.44, SD = 0.96$; Low commitment $M = 3.52, SD = 0.95$). The scores were obtained via a 5-point rating scale (1 = *unsatisfactory achievement* to 5 = *very high achievement*). The number of teachers and number of classes involved are not given. Groups were formed using stratified random direction to ensure that groups included one high-achieving student, two medium-achieving students, and one low-achieving student, but teachers were told to redirect students to different groups if they noticed any personal or behavioral difficulties. The four behavior-state categories that the researchers looked for were (a) cooperative behavior (task-orientated group behavior, listening); (b) non-cooperative behavior (competitive behaviors designed to exclude others and not cooperate); (c) individual task-oriented behavior (works alone on task); and (d) individual non-task behavior and confusion (non-participation in group activities and not working individually). Over a period of four to six weeks, students were videotaped in their classes. Afterward, the researchers coded their behavior. Interrater reliability ranged from 96-100%.

Gillies (2004) found that students in unstructured groups displayed significantly more non-cooperative behavior, $F(1, 170) = 10.98, p < 0.01$, and independent non-task behavior, $F(1, 170) = 8.2, p < 0.01$, than those in structured groups. The study was remarkably transparent and a great amount of attention was paid to detail and accuracy. The analyses are thorough and all

incidents that may have affected the outcome of the study, such as attrition, are reported and taken into account. Additionally, the researchers paid a great deal of attention to the grouping of students. The first weak aspect of the study was the teacher's rating of student commitment and how that translates to achievement ability. It would have been more reliable if teachers had been able to simply submit the students' math scores, but because the students were at different schools and in different classes, it may have been too difficult to standardize all of the data and account for discrepancies in teachers' assessments. The second weakness of the study is the omission of data for the number of classes and teachers involved.

Putnam, Markovchick, Johnson, and Johnson (1996) researched the effects of cooperative learning on regular education students' acceptance of their special education classmates in a quasi-experimental study of 41 special education students and 417 general education students in two Maine schools. The students were in grades five through eight. The researchers found that general education students' perceptions of their special education classmates did not improve in the competitive conditions but improved significantly in the cooperative conditions. The study occurred over eight months of a school year, beginning with preassessments administered in October and ending with postassessments in May. All of the subjects in the experiment studied math, social studies, science, and English language arts. There were two competitive conditions and one cooperative condition. The two competitive conditions were both used as a control for the cooperative condition. The first competitive

condition was taught by the same teachers who taught the cooperative condition classes. The second competitive condition was taught by nine teachers who did not want to be trained in cooperative learning methods. Both competitive conditions were lecture-based, and assessment consisted of norm-referenced examinations. Putnam et al. (1996) observed all classes twice each week and completed an observation form to verify that the cooperative methods were properly implemented by the teachers in the cooperative classes and that cooperative methods were not implemented in the competitive classes. Structured interviews of students and teachers were conducted at the beginning and end of the year; interview subjects were asked about general education students' behavior toward the special education students.

The totaled peer ratings were subjected to an ANCOVA. Two findings are of significance to this paper. First, students' perceptions of special-education classmates became significantly more positive between October and May in the cooperative condition than in the first and second competitive conditions, $F(1, 39) = 17.72, p < 0.001; p's < 0.06$ and 0.01 , respectively. Thirty-eight percent of the special education students were rated significantly more positively in May than they previously were in October. Second, special education students in the cooperative condition reported significant changes in their experiences at school: 96% of the special education students thought that their classmates had treated them well or pretty well; 87% believed that the general education students were more friendly than in the past; 74% reported that they had made new friends who were general education students; and 96% believed that behavior improved over

the course of the year. Overall, positive changes from both groups of students occurred significantly more in the cooperative condition.

Although the study is relatively complex, the necessary data is present and there are no clear threats to its validity. The fact that the researchers monitored the implementation of the cooperative and competitive settings speaks well to the study's instrumentation and internal validity. The implementation of the study had a few limitations in that the teachers were new to the cooperative methods they were using, they were trained during the same school year, and they used cooperative methods 40% of the time. The study also had a number of strengths: it was conducted over the course of an entire year, it was methodologically well-controlled, and it focused on a broad range of age groups.

Weigel, Wisler, and Cook (1975) conducted a quasi-experimental study on the impact of cooperative learning experiences on cross-ethnic relations and attitudes of 231 European American students, 54 African American students, and 39 Mexican American students in English classes in two newly-desegregated junior and senior high schools in Colorado. They found that, among other things, students in classes composed of small, cooperative groups helped their fellow students three times as often as did those in the traditionally-taught control group, and cross-ethnic helping was five times greater in the small-group classes. The study was conducted over a four and-a-half month period at the senior high school and over a seven-month period at the junior high school. Ten English teachers participated in the study (five at the junior high school and five at the high school) and each of them taught one English class via traditional whole-

class methods and one English class using cooperative learning. All classes were observed once each week to ensure that the control and experimental methods were being implemented properly.

By using teacher observations to gather data on student behavior, Weigel et al. (1975) found that, although interpersonal conflict was relatively low in both settings, cross-ethnic hostilities made up 90% of the conflict in the traditional classes and 45% in the cooperative classes ($p < 0.001$). Several other analyses were carried out regarding ethnic attitudes, friendship choices, and group identification. Data for these analyses was gathered through student surveys. The most significant finding from this data was the main effect of teaching method was respect and liking of Mexican American students by European American students, $F = 17.75$, $p < 0.01$. Cross-racial friendship naming was either marginally significant or insignificant for the other groups. Finally, approximately three months after the in-school research was terminated, data on students' attitudes was gathered through interviews in the subjects' homes. Weigel et al. could detect no difference in student attitudes after these interviews.. Altogether, the cooperative classrooms were more harmonious and the researchers recommended the use of cooperative methods in newly desegregated schools. The researchers also conducted further analysis after the study and found that groups that had perceived their work as successful in the intergroup competitions had higher levels of attraction toward their own group members than the other members of the class ($r = 0.40$, $p < 0.01$). The authors suggested that resulting failures in several groups may have inhibited attraction

between ethnic groups. The research methods were transparent, and the article was written with great attention to detail. There are no glaring inconsistencies to suggest that the findings might be flawed. However, different instrumentation might have yielded more significant results for cross-ethnic naming of friends, liking, and respect. The cooperative learning method they used was one of intragroup cooperation, intergroup competition, meaning that it is a version of a group accountability structure. What the researchers do not mention, however, is how many of the groups were recognized as winners in each class. If each activity had only one winning group, the majority of the class may have recognized their own efforts as a failure. By using a method that relies less on competition, or at least recognizes more winners, the researchers may have had more significant method main effects for cooperative learning, friendship naming, liking, and respect.

Hansell and Slavin (1981) conducted a quasi-experimental study on the effects of cooperative learning on the structure of interracial friendship among 402 seventh- and eighth-grade students in twelve inner-city English language arts classrooms, and found that over the course of the ten-week study, the cooperative treatment increased cross-race friendships more for than the control treatment. By examining the reciprocity of close friendship choices, Hansell and Slavin studied the strength of the friendships made. The twelve classes were randomly assigned to treatments taught by five teachers, with 173 students in the control group and 229 in the experimental group. Individual classroom percentages of African American students ranged from 17% to 70%, but

demographics between control and experimental classrooms were very similar. The students studied a ten-week unit on punctuation, grammar, and English usage, and both treatments used the same materials and took the same quizzes. The experimental treatment used the STAD method of cooperative learning with mixed-ability grouping, while the control group was taught via traditional methods.

Student friendships were measured by the sociometric question: "Who are your best friends in this class? Name as many as you wish." This question was administered on the pre- and posttest. The first six friends named were said to be the closest, and Hansell and Slavin (1981) looked for reciprocity. The effects of the treatments were assessed in multiple regressions. The researchers found that there was no treatment effect on total number of choices made. However, on the pretest, the experimental group averaged 3.54 cross-race friendship choices (SD = 3.69, $n = 167$), while the control group averaged 2.93 (SD = 3.27, $n = 154$). On the posttest, the experimental group averaged 3.87 cross-race choices (SD = 3.60, $n = 129$), and the control group averaged 2.37 (SD = 2.76, $n = 133$) (for reasons not mentioned in the article, far fewer students took the survey than were actually in each group). Therefore, the experimental group, on average, made 0.33 more cross-race choices after the treatment than the control group, which made 0.56 fewer cross-race choices on the posttest than on the pretest. Controlling for other variables, the betas for the multivariate analysis of the intersection of the experimental treatment and sheer number of cross-race friendships listed in the posttest are 0.16 and 0.18 (for choices made and received, respectively), which are statistically significant ($p < 0.05$). More

interesting to this particular study is the fact that reciprocated cross-race friendships also occurred at a significant level ($\beta = 0.25$, $p < 0.05$), and cross-race friendships that were listed in the first six places on the sociometric survey were also significant (betas of 0.17 and 0.15 for friendship made and friendship received, respectively; $p < 0.05$).

Steps were taken to minimize expectations of the cooperative group among both teachers and students, adding to the reliability of the instrumentation. Students were not told that race relations were being measured and teachers were told that interracial friendship was only one of several dependent variables measured. The results were significant and were subjected to thorough analysis. There are no glaring inconsistencies or ambiguous language present in the study. A weakness of the study admitted by the researchers is that students' behavior in the groups was not actually observed. Therefore, the exact factors in the cooperative setting that led to an increase in the naming and reciprocating of friendship choices cannot be pinpointed. Another potential weakness is the omission of a reason for why so many students in both the control and experimental groups did not complete the survey.

Bratt (2008) conducted two quasi-experimental studies to determine the effectiveness of the Jigsaw model on improving the classroom intergroup relations of multiethnic classes of students aged 11-15 years in Norway and found that the both studies did not support the hypothesis that Jigsaw would have positive effects on attitudes towards minority students, common in-group identity, empathy, or attitudes towards school. The first study involved younger

sixth-grade students and the second study involved students in grades eight through ten. The first study was conducted at two multiethnic schools in Oslo, Norway over a period of seven weeks. The analysis was focused on the attitudes of ethnic Norwegian students. Each school had a Jigsaw class and a control class. One of the Jigsaw classes had two teachers while the other three classes had only one teacher. The Jigsaw teachers volunteered and were trained during a two-day course lead by the author and a teacher experienced in teaching multiethnic classes. The experimental classes ($n = 34$) used Jigsaw for seven weeks while the control classes ($n = 34$) continued with their regular course of study. In the Jigsaw groups, 24% of the participants, or one or two per group, belonged to a non-Western ethnic group. The researcher did not state what percentage of students in the control group were of non-Western ethnic background. Both schools used Jigsaw about once a day. School one had tests after every Jigsaw session while school two had them somewhat less frequently. The author stated that students in both schools talked about their Jigsaw experiences as a class rather than in their Jigsaw groups, and that it was not an ideal implementation of the method compared to the guidelines set forth by Aronson and Patnoe (Aronson and Patnoe, 1997, cited in Bratt, 2008, p. 407). The researcher used a questionnaire that was administered as both a pretest and a posttest. The questionnaire measured (a) attitudes toward boys and girls of different racial origin, (b) attitudes toward classmates, (c) empathy, (d) intergroup friendship, and (e) attitudes toward school.

An ANOVA of pre-test data did not show any differences between

experimental and control classes for any of the aforementioned variables. Comparison of pre- and posttest data for each class found that the Jigsaw class with two teachers had significantly more positive attitudes towards all students of different ethnicity than the control class at the same school. This same Jigsaw class also had more favorable development in attitudes toward Nordic looking girls. The second Jigsaw class had no favorable development of attitudes towards students of different ethnic background and a substantially more negative development of empathy (according to one of two measures of empathy used). A one-way ANOVA of change scores in the two Jigsaw classes compared to the two control classes revealed significant differences in attitudes toward class members, attitudes toward Nordic looking girls, and friendship with ethnic Norwegian students, but there were no significant findings for attitudes toward members of non-Nordic ethnic groups. Further 2 X 2 X 2 ANOVA (Jigsaw participation X school affiliation X repeated measures) revealed a significant interaction of repeated measures and Jigsaw participation $F(1,64) = 4.493-5.410$, $ps < 0.05$ for positive development of attitudes to the class and attitudes toward boys and girls that look Asian. There was also a significant interaction effect of repeated measures and Jigsaw on the development of Nordic-looking girls, though there was no significant interaction effect for attitudes toward boys or girls who appeared to be of African descent, friendship with other Norwegians, empathy, or school attitude. Additionally, two interaction effects were revealed through a 2 X 2 X 2 (repeated measures X Jigsaw X school) ANOVA, $F(1,64) = 7.468-7.834$, $ps < 0.007-0.008$. These two effects were attitudes toward Asian

boys and the negative development of empathy that one of the two measurements of empathy yielded.

Although there were significant positive outcomes for the experimental group when compared to the control group, the positive outcomes can be attributed to the Jigsaw class with 22 students and two teachers. The researcher suggested that positive findings in this class may have been attributed to the smaller class size and presence of two teachers. The researcher also suggested that the small sample size may have greatly affected the outcome and therefore modified the second study to account for these two potential weaknesses.

In the second study, like the first study, teachers participated in a two-day workshop on the implementation of Jigsaw. The method was implemented for a period of eight weeks, for about one to two hours a week.. The study was conducted in Drammen, a city of about 55,000 people, located just outside Oslo. There were 11 Jigsaw classes with 129 ethnic Norwegians taking the pre- and posttests, and 35 control classes learning via traditional methods; pretest and posttest questionnaires were administered to all ethnic Norwegians. Eleven of the 35 control classes that most closely matched the experimental classes were used in the analysis. No exact number of non-ethnic Norwegians was given due to a lack of individualized data. The assessments on views toward ethnic minorities were focused on ethnic minorities that had a substantial presence in the town (Pakistani, Vietnamese, and Turkish). The subjects were in grades eight through ten and were well balanced between control and experimental groups in terms of gender and socioeconomic status. Jigsaw groups were composed of

five to six members, of which, one to two were members of ethnic minority groups.

The analysis was focused on matched control classes (regarding school grade and sample size). Intergroup attitudes were assessed by semantic differential scales ($\alpha < 0.86$). Common in-group identity was assessed with three items using a 7-point Likert-response format ($\alpha = 0.59-0.73$). Intergroup friendship was assessed with one item for each gender in ethnic minority groups, asking for the number of close friends of different ethnicities that students like to be with outside of school. Answers were coded into a scale of 1-11. Empathy was assessed with a four-item scale ($\alpha = 0.68-0.82$).

An ANOVA did not uncover significant results in pretest data in attitudes toward different ethnic groups. Several thorough methods of analysis to disentangle potential effects showed that there were no significant effects on any of the four attitudes. Bratt (2008) suggested that a limitation to the study was that it used assessments of attitudes toward particular ethnic groups, while none of these ethnic groups were large enough to be represented in most of the Jigsaw groups. The study was complex, and the lack of consistent results can be attributed to the instrumentation and reliability of the research. The study has a low level of internal validity because of the potentially flawed methodology. It could be argued that, because an ANOVA did not uncover any significant results in the pretest data, there was no valid reason to continue the study. Students did not have significantly different views of minority groups to begin with; therefore it is unclear why the researcher chose to continue the study. Also, the low

representation of ethnic groups in the Jigsaw setting is a weakness in methodology, further diminishing the study's internal validity.

Four of the five articles on student behavior and attitudes towards others that were examined in this subsection found significant positive effects for cooperative learning. Gillies (2004) found that structured groups had significantly higher effects on cooperative behaviors than unstructured groups. Putnam et al. (1996) found significant improvement in regular education students' attitudes toward special education peers. Weigel et al. (1975) found that students in cooperative setting were three times as likely to help their peers as students in the control group, and that cross-ethnic helping was five times greater in the cooperative setting. And Hansell and Slavin (1981) found that students in cooperative groups had significantly more cross-race friendships than students in the control group. The only studies that yielded questionable results were Bratt's (2008) studies on the use of Jigsaw to improve intergroup relations of multiethnic classes; however, the studies were potentially flawed.

This second section of chapter two examined the social and affective effects of cooperative learning in three distinct areas: the effects of previous exposure to cooperative learning methods; the effects on student attitudes, perception, and preference towards cooperative learning methods; and effects on student behavior and attitudes towards others in cooperative settings. Only one of the eleven studies did not produce significant findings, meaning that overall the majority of the social and affective effects of cooperative learning in the studies reviewed are positively significant.

Summary

Chapter two was divided into two sections. The first section dealt with the effects of cooperative learning on the academic achievement of secondary-level English language arts and social studies students. The section was divided into three subsections addressing the effects of different accountability structures of cooperative learning: individual, mixed individual/group, and group. The majority of the findings for all three accountability structures were positive, suggesting that all the three forms have the potential to increase student achievement.

The second section addressed the effects of cooperative learning on the social interaction, behavior, and affect of secondary-level English language arts and social studies students. The section was divided into three subsections: the effects of previous exposure to cooperative methods; student attitude, perception, and preference toward cooperative methods; and the effects of cooperative learning on student attitudes and behavior toward others. The findings in the first subsection suggested that students who have been exposed to cooperative learning in the past are more likely to engage in cooperative behaviors with others – even when not instructed to. The findings from the second subsection suggest that students who prefer to learn in a cooperative setting have higher levels of achievement when in that setting. In general, the majority of participants in cooperative learning viewed the method favorably; one sample population whose academic achievement was somewhat diminished after participating in the cooperative setting still had positive views of group investigation. In the third subsection, the majority of the studies reviewed revealed that students in

cooperative settings were more likely to engage in cooperative behaviors and peer-helping than students in a traditional setting. The research also suggests that students have improved attitudes toward one another when learning cooperatively.

CHAPTER THREE: CONCLUSION

Introduction

Chapter one gave a brief history of the rationale for cooperative learning. Humans have been working together to achieve common goals since time eternal, but cooperative learning did not become common as an instructional strategy until Dewey introduced what he called the Project Method of learning, which was widely used in American schools until the 1930s. The Great Depression led to major changes in business and education, and cooperative learning was abandoned as schools reverted back to traditional whole-class methods.

Cooperative methods were relatively unknown between the 1940s and 1970s, as a result of World War II and the Cold War that succeeded it. During these years, American ideologies of social Darwinism and rugged individualism emphasized competition and success through isolated efforts. Individualistic and competitive methods of schooling were heavily influenced by Skinner's theory of operant conditioning, a guiding principle of what came to be known as behaviorism.

Cooperative learning methods resurfaced in the 1960s and began to take hold in the 1970s. Since this resurgence, cooperative learning has been thoroughly developed and researched in elementary schools in particular, but not to the same extent in secondary-level English language arts and social studies. Although all methods of cooperative learning address both the cognitive and affective demands of students, the relative shortage of research on the academic,

behavioral, and affective effects of cooperative learning on secondary English language arts and social studies students was the impetus for this review.

Chapter two reviewed the effects of cooperative learning on academic achievement as well as on social interaction, behavior, and affect. The first section on academic achievement was divided into three subsections addressing different accountability structures of cooperative learning. Some methods of cooperative learning are based on an individual accountability structure, meaning that even though students work together to complete a task and attain a concept, they are graded or rewarded only for their individual efforts. Other methods of cooperative learning are based on a mixture of individual and group accountability structures, meaning that students are accountable for their own contributions and are also accountable to the group. Finally, other methods of cooperative learning are based on a group accountability structure only; individuals are not accountable for their own contributions, and instead the group is held accountable as a whole. The majority of the articles reviewed contained significant findings for gains in academic achievement in all three of the structures.

The second section of the chapter reviewed the literature on the effects of cooperative learning on social interaction, behavior, and affect. The section was divided into three subsections: the effects of previous exposure to cooperative methods; attitude, perception, and preference towards cooperative learning; and attitudes and behavior towards others in cooperative settings. All but a few of the studies demonstrated that cooperative learning has positive effects on behavior,

social interaction, and affect.

Chapter three is the concluding chapter of this paper, which reviews the guiding question: What are the effects of cooperative learning on academic achievement, social interaction, behavior, and affect of secondary-level English and social studies students? The findings from chapter two that address these effects will be used to answer this question, provide classroom implications, and give suggestions for further research.

Summary of Findings

Academic achievement and social interaction in schools are often strong indicators of future success. School teaches students how to navigate the obstacles of life – obstacles that may take the form of either workplace politics or college graduation, for example. School is where students become acquainted with social norms and expectations, many of which are directly related to situations in which they are expected to work with others. Cooperative learning is an appropriate medium by which to acculturate students to these norms and expectations.

There is a fundamental question that must be asked, however, when developing or implementing a cooperative learning method: How should students be held accountable? Should they be accountable on the basis of their own contributions to the group, on both their own contributions and the final outcome of the group's effort, or on the outcome of the group's effort alone? These questions hold major ethical implications. In American society in particular, success is often viewed in terms of individual achievement; this is clearly

reflected in most educational curricula and assessments. American society is in many regards an individualistic one. Yet Vygotsky's sociocultural learning theory purports that learning with others often has greater cognitive, academic, and affective effects.

In 1978, Robert Slavin conducted a study to investigate the effects of accountability structures on academic achievement. He compared individual-level reward with group-level reward structures and found no significant results for either. Despite this, the articles reviewed in the first subsection demonstrate that individual accountability methods of cooperative learning have significant positive effects on academic achievement. The first two studies on the effects of Jigsaw on academic achievement by Lucker et al. (1976) and Gocer (2010) both showed that Jigsaw had a significant effect on gains in academic achievement. Lucker et al. showed that minority students made significant advances when working in cooperative Jigsaw groups, and that European American students performed equally well in both settings. Gocer (2010) found that students in a Jigsaw setting had significantly higher posttest scores than students taught via traditional methods. Adeyemi's (2008) individual accountability method of cooperative learning also showed gains in academic achievement when compared to traditional whole-class methods, but omission of mean scores and aspects of methodology detract from the study's internal validity. Despite this, the collected body of research in this first subsection suggests that individual accountability structures are effective for producing gains in academic achievement.

Although individual reward structures have been shown to have an effect on academic achievement in the three cases mentioned above, these studies do not answer the question of whether or not group effort should be individually assessed. Several other methods of cooperative learning that incorporate both individual-level and group-level accountability have been studied in an attempt to determine their effect on academic achievement. In 1990, Slavin modified Aronson's original Jigsaw method to include a group scoring component referred to as equal opportunity scoring, or EOS (Mattingly et al., 1991). This method, known as Jigsaw II, therefore has an individual-level reward as well as a group-level reward. This is therefore a mixed individual/group accountability structure method of cooperative learning. Mattingly et al. (1991) studied whether Slavin's Jigsaw II method would yield higher academic achievement results than traditional whole-class instruction, and found that the Jigsaw II method had a significant positive effect on academic achievement for ninth-grade students. Pretest results were similar between the traditionally-taught and Jigsaw groups, but the posttest scores for the Jigsaw group were significantly higher. Şahin (2010) performed a similar study on the effects of Jigsaw II, which yielded similar findings; however, the subjects were preservice language teachers in a graduate studies program in Turkey. Although both the control and experimental groups had significantly higher scores on the posttest than the pretest, the Jigsaw II subjects' mean scores were higher and of greater significance. Overall, Jigsaw II has been shown to be effective in increasing gains in achievement – even among different populations – adding greatly to the generalizability or transferability of

the method.

Slavin et al. (1981) found that Student Team Achievement Divisions (STAD), another method of individual-plus-group accountability structure cooperative learning, also yielded significant positive effects on academic achievement. In particular, it yielded higher gains for African American students. In fact, the scoring improvement of European American students was not statistically significant. Although the Jigsaw II and STAD studies yielded gains in academic achievement, the assignment of the subjects to the cooperative treatments was either random or designed to balance achievement scores with the control group, meaning that academic and social heterogeneity were not directly accounted for.

The studies by Johnson et al. (1976) and Lew et al. (1986) focused on ability grouping, meaning that students in the cooperative treatments were deliberately grouped heterogeneously according to academic performance. Groups were typically composed of one academic high-achiever, two average-achievers, and one low-achiever. Johnson et al. (1976) found that students working in mixed-ability groups made fewer mistakes on in-class assignments and had a higher score on the posttest that they took as a group than students in the individual condition. However, all students in the class took an individual posttest and there were no significant differences in achievement between students who worked individually and students who worked in groups. This suggests that the method is effective when accounting for students' scores as a group, but not as individuals. Lew et al. (1986) also grouped the experimental

subjects according to academic as well as social heterogeneity, and both experimental cooperative conditions produced significantly higher academic achievement scores than the baseline individual condition. In short, the cooperative treatment subjects in Johnson et al. (1976) had higher achievement scores when they took the test as a group, and the subjects in the Lew et al. (1986) cooperative conditions had significantly higher posttest scores than the students in the individual condition, implying that heterogeneous ability-grouping has a positive impact on gains in academic achievement in mixed individual/group accountability structures of cooperative learning. However, in the Johnson et al. (1976) study, students in the cooperative condition only scored higher when they took the test together.

The final mixed individual/group accountability method of cooperative learning examined in this subsection was Devries, Edwards, and Wells' (1974) study on the effects of Teams-Groups-Teams (TGT) on academic achievement – particularly on the academic achievement of low-ability students. There were three treatment conditions: an individual control (IC) treatment, a TGT-Average (TGT-A) treatment, and a TGT-Weighted (TGT-W) treatment. The TGT-A condition averaged the scores of the group members, and the TGT-W condition weighted the scores of lower-achieving members to contribute a greater number of points toward the team score. Few, if any, treatment effects on academic achievement were found between the two TGT conditions, but there were marginally significant effects on improved academic achievement when compared to the IC method. Overall, the study found that either way of

organizing TGT groups yielded higher academic achievement results than the IC method.

Several different studies examined a variety of mixed individual/group accountability structures, and the general conclusion is that the mixed accountability structures have a positive impact on academic achievement when students are grouped both randomly and heterogeneously. The sample populations in all of the studies were from a variety of different backgrounds, suggesting that the mixed accountability methods are highly generalizable with a high degree of external validity.

Although the mixed accountability methods generally showed positive results on academic achievement, the fact that students are accountable for their own contributions in addition to the overall group contribution still leaves room for potential competition among group members. Because each group member receives an individual score, students may compare scores with each other and potentially find fault in others' contributions if the group does not succeed in its task. This potential for in-group competition is one of the major reasons for implementing group-only accountability methods of cooperative learning. By holding the group accountable for successes and failures, not the group *and* its members, positive interdependence to complete a task or meet an objective is the primary driving force, not the average or weighted average of individuals' achievements.

The majority of the studies in the third subsection on group accountability methods of cooperative learning found that the experimental treatment conditions,

such as group investigation (GI) and cooperative integrated reading comprehension (CIRC), had significant positive effects on increased academic achievement when compared to traditional instruction (Durukan, 2011; Johnson et al., 1979; Shachar & Sharan, 1994; Stevens et al., 1987; Stevens et al., 1991; Stevens, 2003). In examining the effects of group investigation on both high- and low-achieving students, Tan et al. (2007) conducted a study very similar to that of Shachar and Sharan (1994), and had very different findings. Shachar and Sharan found that in an ethnically heterogeneous middle school in Israel, ethnic minority students from a lower socioeconomic class who had lower pretest achievement scores had greater gains on the posttest than the majority students from the predominantly middle-class background. In a methodologically similar study on the achievement of high- and low-achieving middle school students in Singapore (Tan et al., 2007), the researchers hypothesized that both groups of students would have higher achievement than the control treatment students taught via traditional methods. The high-achieving students' achievement increased significantly when in the cooperative treatment condition, but the low-achieving students' academic performance was actually diminished between pre- and posttests. Both studies sought to reveal the effects of GI on a heterogeneous sample population. The Israeli middle school students were heterogeneous both academically and socioeconomically, while the students from Singapore were only academically heterogeneous. This may imply that, to some degree, a greater number of factors of heterogeneity and diversity may positively impact student achievement.

The third subsection on group accountability structures included four studies on cooperative integrated reading and writing (CIRC). The two studies conducted by Stevens et al. (1987) were methodologically very similar. The first study compared the effects of CIRC on the academic achievement of third- and fourth-grade students in suburban Maryland, in which 461 students studying English language arts participated. The mean percentage of minority students was 16.1%, and the study lasted 12 weeks. The experimental CIRC group had significantly higher achievement than the control group in reading comprehension, reading vocabulary, language expression, and spelling. The researchers' second study, which was nearly identical in method, spanned a longer duration and included more racially diverse students in the sample population. The mean percentage of minority students was 22.4%, and the study lasted six months. The students in the experimental treatment condition in this second study had significantly higher achievement than the control students on language expression, language mechanics, and reading comprehension. These two studies therefore imply that CIRC is a more effective method of instruction than traditional teacher-centered instruction for middle-grade elementary students of different demographics and for differing durations.

Four years later, in 1991, Stevens et al. compared the effects of three different teaching methods on student academic achievement. The first method was CIRC, the second was CIRC integrated with direct instruction, and the third was direct instruction. The students were ethnically diverse, and African American students made up between 56.3% and 85% of the student population

in each of the schools involved. The study lasted four weeks, and the results showed that the effects of both CIRC and CIRC mixed with direct instruction were significant, and that students in those experimental groups had higher gains in academic achievement than students in the direct instruction group.

These three studies demonstrate that CIRC is an effective instructional method for increasing student reading achievement in grades three and four, and across broad ranges of ethnic diversity and socioeconomic status. The studies also demonstrate that CIRC can improve academic achievement over the course of four weeks or over the course of six months, suggesting that its effects are quickly apparent and lasting. The final study on CIRC (Durukan, 2011) found that seventh-grade language students in Turkey had higher achievement scores than their peers taught via conventional methods, which further demonstrates the flexibility, generalizability, and external validity of CIRC.

Stevens' (2003) study on student team reading and writing (STRW) was a large-scale implementation of a group accountability method of cooperative learning that involved five urban middle schools and 3,916 students. Approximately 80% of the students were of a racial or ethnic minority, and the surrounding neighborhoods were high in poverty. The yearlong study revealed that students in the experimental STRW treatment conditions had far higher scores than their control group counterparts taught via traditional methods. This study demonstrates that STRW is an effective way of improving academic achievement on a large-scale, long-term basis and with diverse and impoverished youth. It also demonstrates that this particular group accountability

form of cooperative learning is effective for improving the achievement of middle school students.

The subsection on group accountability structures of cooperative learning also contained a study by Johnson et al. (1979) on the effects of cooperative, individualistic, and competitive goal structures on student attitudes and achievement, in which the subjects were fourth-, fifth-, and sixth-grade students studying geometry. The results of the study showed that the students in the cooperative condition had the highest posttest scores, reinforcing the claim that group accountability cooperative learning is effective for upper-level elementary students. The results also suggest that group accountability cooperative learning is effective for teaching subjects other than English language arts and social studies.

All these aforementioned studies on group accountability methods of cooperative learning involved elementary and middle school students. As was mentioned previously in the paper, there has been a relative shortage of research conducted on the effects of group accountability cooperative learning on academic achievement in upper secondary grades nine through twelve. The studies involving elementary students add to the external validity of the cooperative methods they address because they were well-executed and revealed significant positive findings, but they do not directly address the effects on a secondary sample population. The final study included in this subsection was the only study involving high school students. Phuong-Mai et al. (2009) compared the effects of a simple, unnamed group accountability form of

cooperative learning with another culturally responsive form of cooperative learning. The researchers wanted to determine if a cooperative structure that was more in line with Vietnamese leadership values would yield higher student achievement results. The study was somewhat unreliable because the students were paid volunteers and there was no pretest to determine equivalency between the groups. In light of this, the students in the culturally responsive cooperative learning group did have higher achievement in the social sciences than the students in the traditional group. It may therefore be plausible to assert that cooperative learning strategies that are adapted to certain cultural values may have a greater impact on achievement than cooperative methods that are not.

In sum, studies on all three accountability structures of cooperative learning demonstrated that they have a positive impact on academic achievement. Throughout a number of age groups, academic subjects, cultures, ethnicities, and varying degrees of racial and academic heterogeneity, more often than not, academic achievement of students taught through cooperative methods is higher than the achievement of students taught via traditional methods. Even if academic achievement is not increased, there are other potential benefits of using cooperative methods.

The second section of the critical review on the effects of cooperative learning on social interaction, behavior, and affect was divided into three subsections: the effects of previous exposure to cooperative methods on present behavior; student attitude, perception, and preference toward cooperative learning; and effects of cooperative learning on attitudes and behavior toward

others. The first subsection on the effects of previous exposure to cooperative learning on behavior contained two articles that examined the ways in which students interacted with one another when they had or had not been previously exposed to cooperative methods. Both Gillies (2002) and Ryan et al. (1974) sought to uncover the effects of students' previous exposure to and training in cooperative learning, and both found that students who had been taught to work in cooperative groups either weeks or years earlier were more likely to work cooperatively even when not explicitly told to. Ryan et al. (1974) found that students were more likely to work together – even when they initially did not want to; Gillies (2002) found that students who had been trained in cooperative learning two years earlier and had a chance to practice a year earlier were more likely to engage in cooperative behaviors such as staying on task and being socially oriented to others in the group. Students who had not been trained in cooperative learning were more likely to engage in non-cooperative behaviors, such as behavior demonstrative of competition or behavior in opposition to criticism or group goals. All of the students involved in the two studies were in fifth and sixth grade. Generally speaking, students in the fifth grade are not secondary-level students and students in the sixth grade are, but they are developmentally similar, making these studies of elementary students relevant to this review. In sum, the studies demonstrate that students who are trained to work in cooperative groups at a younger age will be more likely to work cooperatively with other students throughout their secondary education and, theoretically, their lives.

The second subsection contained three studies on student attitude, perception, and preference toward cooperative learning. Dunn et al. (1990) found that American middle school social studies students who were able to learn in their preferred setting (individually or in a group) had higher scores and more positive attitudes than they had when they learned in a non-preferred setting. The second study, by Owens and Barnes (1982), found that Australian eleventh-grade students preferred learning in a cooperative environment and seventh-grade students preferred a competitive environment; however, both the seventh- and eleventh-grade students preferred to learn more cooperatively in English language arts than in math. According to Dunn et al. (1990), allowing students to choose which method of learning they engage in positively affects attitude and achievement, and according to Owens and Barnes, those preferences change as students progress through secondary education. Perhaps older students choose to work together more than younger students because of a more demanding course load, or possibly because they are at a different level of cognitive development. Regardless, according to these two studies, it benefits the students to allow some degree of choice.

While the two previous studies were conducted on subjects of Western cultures, Tan et al. (2005) found that the majority of 142 eighth-grade students in Singapore had a preference for using cooperative methods in social studies. Both high-achieving and low-achieving students preferred to learn in a cooperative setting. This means there is evidence that students from both Western and Eastern cultures have a preference for learning English language

arts or social studies cooperatively, and that these similar results are generalizable with some degree of external validity. With this in mind, the evidence from Dunn et al. (1990), showing that students who are allowed to work in their preferred setting will have more positive attitudes and higher levels of achievement, further supports the notion of giving students the opportunity to learn in cooperative groups.

Although many students may prefer to work in cooperative groups, and those students who do may have increased academic achievement, there are other positive effects of working in cooperative groups as well. The third subsection on the effects of cooperative learning on student attitudes and behavior toward others contained a number of significant positive findings. Gillies (2004) found that students working cooperatively in structured groups were significantly more likely to engage in cooperative behavior than students working together in unstructured, collaborative groups. This is an important distinction between collaborative and cooperative learning. Collaborative learning is simply the act of working together to accomplish a learning task, while cooperative learning is built on the theory of social interdependence, meaning that the task is structured so that it can only be accomplished if all of the group members have real or perceived individual and group roles and responsibilities and work to fulfill them. In sum, this study demonstrated that students in a cooperative environment are more likely to stay on task and work productively with group members than students who are simply placed in collaborative work settings. The study highlights the importance of structure in cooperative learning.

The interdependent structure of cooperative learning has a strong impact on the ways in which students perceive each other and behave toward each other. Students who have participated in cooperative learning may be more inclined to accept and befriend others, as compared to students in competitive, individualistic, or even collaborative learning settings. Putnam et al. (1996) found that when mixing general education students and special education students in a traditional classroom setting, the general education students' views of their special education peers did not improve over the course of a school year. But when general education students were mixed with special education students in structured cooperative classrooms, students' perceptions of each other were improved for both groups. Weigel et al. (1975) also found that students in a recently desegregated school were three times as likely to help their classmates when in a cooperative classroom than when in a traditional classroom. They also found that students helped students of other ethnicities five times as often in the cooperative classroom than they did in the traditional classroom. Hansell and Slavin (1981) found that students in cooperative groups named far more students of other races as friends than students in traditionally-taught classrooms. The only studies in the group that did not find significant positive results pertaining to the ways in which ethnic majority students perceive ethnic minority classmates was Bratt's (2008) experiment on the effectiveness of the Jigsaw model on improving the intergroup relations of multiethnic classes of students. However, it is highly possible that the experiment did not reveal more significant findings due to flawed methodology.

Students in the previously listed studies were a variety of ages and were studying mostly social studies and English language arts when the experiments were conducted. The fact that nearly every study found that intergroup relations were improved among students from a variety of educational settings ranging from special education to a highly diverse, recently integrated school shows that cooperative learning has strong positive effects on the ways in which students view and act toward one another.

Classroom Implications

Given that all three accountability structures have been shown to have significant positive effects on students' academic achievement, a crucial question is what a teacher wants students to get out of their cooperative learning experience. If a teacher wants students to be held accountable for their individual contributions, then individual or mixed reward structures should be used. If a teacher were trying to teach students the value of interdependence through cooperative learning, then a group accountability structure should be used.

In the broader scheme of things, teachers can use different versions of cooperative learning to prepare students for different life situations. Although adults are required to work with others in nearly every professional setting, they are often assessed on their individual performance, and they are almost always remunerated on an individual basis. It is difficult to imagine an employer handing a lump sum payment to a group of employees who were on a development team together, but it is not so difficult to imagine all of the employees receiving their usual salary plus a bonus. On the other hand, although employees are typically

paid individually, the success of a team is often dependent on the actions of the individuals it comprises; the members must complete their jobs so that the group can succeed and everyone can remain employed and receive payment.

Individual accountability and mixed accountability methods of cooperative learning reflect this reality.

Group accountability methods of cooperative learning reflect a different reality. These methods reward and assess the entire group on the job they have done as a team. No one is left out, no one gets more or less than others, and the focus is less on the reward than on the accomplishment of completing a task or attaining a concept together. There are, however, some difficulties that can arise when implementing a group accountability method of cooperative learning in the classroom. In the era of No Child Left Behind (NCLB), both students and teachers are increasingly accountable for performance. Teachers often need to assess and evaluate students on an individual basis. This does not rule out the possibility of using group accountability methods of cooperative learning, but it means that there are some potential limitations. Teachers need to be able to assess where individual students are in their learning to ensure that they pass standardized tests. Teachers also need to have individual grades for students. On the other hand, any accountability structure of cooperative learning is an opportunity for teachers to assess participation and communication skills, which are commonly included in state or core learning standards.

The teacher's beliefs about teaching and learning are important considerations when determining which type of cooperative learning to use,

especially when a teacher determines that assessment will be involved. Some teachers argue that students should never be assessed as a group because it is important to evaluate whether or not each individual student has met a learning objective. Still other teachers might believe that group completion of a task is evidence of meeting the objective, regardless of what individuals may or may not have contributed. Other teachers may opt for a mixed-structure method of cooperative learning to address both of these concerns. The majority of the studies of cooperative learning reviewed had significant positive effects on academic achievement, so it is up to the teacher to decide what else the students should gain from the learning task.

In addition to giving students the skills to achieve interdependently as part of a group, cooperative learning carries with it a host of other social and affective classroom implications. Cooperative learning has long-term implications on students' cooperative behavior toward others. Gillies (2002) and Ryan et al. (1977) both found that students who participated in cooperative class activities were more likely to work cooperatively even when not asked explicitly to do so. This implies that cooperative learning in the classroom prepares students to function in a democratic society by teaching them how to complete tasks and activities as a member of a group with both individual and group roles and responsibilities.

Tan et al. (2005) found that both high- and low-achieving geography students had more positive than negative perceptions of learning cooperatively, and Owens et al. (1982) found that the majority of their seventh- and eleventh-

grade subjects preferred to learn English cooperatively. Contrastingly, Dunn et al. (1990) found that students who learn in their preferred setting had significantly higher achievement than students who learned in a non-preferred setting. The research of Tan et al. (2005) and Owens et al. (1982) implies that a majority of students prefer to learn cooperatively, and the research of Dunn et al. (1990) that those who work in cooperative learning groups will have higher levels of achievement. However there will still be students who prefer to learn alone, which creates an excellent opportunity for differentiation in instruction. According to Dunn et al. (1990), if students are given the choice to work alone or with a group, they will have higher achievement than if they were forced to learn in a non-preferred setting. Therefore, teachers could offer individualized learning to some and cooperative learning to others.

A further classroom implication of cooperative learning is the importance of structure in the cooperative groups – not necessarily the importance of a specific accountability structure, but the presence of a structure of interdependence. In order for students to be driven by interdependence, the task needs to be group-worthy and open-ended. Students need to be aware of what their individual and group roles and responsibilities are, and students must know why those roles and responsibilities are necessary for the group to complete its task. Gillies (2004) demonstrated the importance of this structure when her study revealed that students in an unstructured collaborative setting were far more likely to engage in competitive and oppositional behaviors. It is therefore the responsibility of the teacher to make sure that the tasks and activities assigned to

students are group-worthy, open-ended, and structured to promote interdependence.

Teachers are sometimes reluctant to use cooperative methods because they perceive that time spent teaching students to work in structured groups should instead be time spent learning academically, and that therefore teaching students to work in cooperative groups will lead to a drop in academic achievement. As the majority of the studies in this review illustrate, this is not typically the case. Students who are taught to work cooperatively have higher levels of achievement and experience a host of other positive social and affective effects. Therefore, making the time to teach students to work in cooperative groups will likely have positive academic, social, and affective results.

In regard to building a classroom community, cooperative learning has been shown to increase acceptance and improve intergroup relations (Weigel et al., 1981), particularly in the context of highly diverse schools. It was also found that long-term use of cooperative learning methods increased acceptance of special education students among their regular education peers (Putnam et al., 1996), making it a viable option for a mainstreaming model of special education. Cooperative learning is an effective way to improve students' ability to work with and accept others who are different from themselves.

Further Research

Cooperative learning has been one of the most thoroughly researched educational practices in the last four decades, with a great deal of that research focused on academic achievement and socialization; however, the bulk of it has

been conducted at the elementary level. It was not until more recently that cooperative learning at the secondary level has been more thoroughly researched. Still, research on cooperative learning in secondary English and social studies classrooms has not been conducted to the extent that it has been in elementary classes. Therefore, simply put, more research needs to be done on the effects of cooperative learning on secondary level students in these subject areas, particularly high school students. Furthermore, little or no research has been done on

In *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, Slavin (1996) proposed that three types of cooperative learning tasks or activities might not rely on individual or group accountability, as theoretical models of cooperative learning propose. The three tasks he proposed were controversial tasks lacking single right answers, voluntary study groups, and structured dyadic tasks. At the time, no research had been done on voluntary study groups, but other research on controversial tasks and structured dyadic tasks had found significant positive effects on academic achievement. Yet he suggested that additional evidence had shown that group goals and individual accountability further enhanced achievement. In light of this, he proposed that further research be done on cooperative conditions in which group goals and individual accountability may not be necessary, although he did not give any examples of what those might look like.

As a follow-up to Slavin's (1996) article, Abrami and Chambers (1996) suggested that in addition to investigating forms of cooperative learning that do

not rely on extrinsic motivation, research could also be done on the effects of what happens when groups fail. Cooperative learning is not an assurance of success, and failure of the group to complete a task or meet an objective can have negative consequences. It would be valuable to know how students are affected over time by both success and failure in cooperative learning settings. Such research should also investigate the three accountability structures outlined in this review, and take into consideration both their internal and external effects. For example, what are the effects of each of the three outlined accountability structures on *both* academic achievement and attitudes towards others? Do some of the accountability structures have different effects on the ways students view one another when participating in cooperative learning?

Furthermore, more research could be done to explore how cooperative learning affects the ways in which students perceive and respond to cultural differences. For example, although Hansell and Slavin (1981) found that cooperative learning had significant positive effects on students' naming of friends of different races when compared to traditional competitive learning, the study did not take into account students' perceptions of English language usage. Future research might focus on the ways in which cooperative learning affects students' perceptions of English usage different from their own. It would be of value to determine whether speakers of nonstandard dialects had an improved perception of standard English dialect if they were exposed to it in more cooperative settings. Likewise, it would be useful to see if students who speak standard dialect English have more positive views of non-standard dialects of

English after working in cooperative groups with speakers of nonstandard dialects.

Additionally, Hansell and Slavin (1981) found that learning in a cooperative setting increased the listing and reciprocity of close friendship choices among students of different races, but the researchers stated that a weakness of their study was their inability to actually observe the students in the cooperative setting. Because of this, the researchers were unable to determine which factors in the cooperative setting led to the increase in the naming of new friends of a different race. An avenue for further study would be to determine which factors in a cooperative English language arts settings affect the forming of close friendships.

Finally, further research on the integration of technology and cooperative learning could have major implications on the ways in which curricula are designed. Technology in the classroom is continuing to gain popularity among teachers and students, and as it is firmly grounded in the concepts of communication and sharing, cooperative learning methods might enhance the outcomes of computer-based learning experiences by adding a degree of structure. In a report for JISC Technology and Standards Watch, Anderson (2007) suggested that there is a conflict between established education and the rise of computer-based instruction because education has traditionally focused on the construction of formal knowledge, while computer-based learning typically promotes the construction of informal knowledge. Perhaps the addition of pre-established cooperative learning structures could allow students to perform

traditional tasks—such as a group investigation into the cause of polar melting—with students from classes in different schools or even different countries. A presentation of the findings could also be developed through online collaboration. As schools continue to evolve to meet the demands of an increasingly diverse and technological populace, cooperative learning may have a place in computer-based education.

Conclusion

Cooperative learning has been thoroughly researched and implemented in the last four decades, but most of the research on its effects have been at the elementary level. The comparative shortage on the effects of cooperative learning at the secondary level, specifically in English and social studies classes, therefore generated this paper's research question: What are the effects of cooperative learning on the academic achievement, social interaction, behavior, and affect of secondary-level English and social studies students? Answers to this question will afford educators access to research that allows them to make more informed decisions about their instructional strategies.

The collective body of research in this review suggests that there are a number of significant positive effects of cooperative learning. There are three accountability structures of cooperative learning: individual, mixed individual/group, and group. The majority of the studies reviewed found that students participating in all three types of cooperative learning structures had significant gains in academic achievement. The research on social interaction, behavior, and affect revealed that working in cooperative groups had several

significant effects. Students who have been exposed to cooperative learning in the past are more likely to work cooperatively of their own volition (Gillies, 2002; Ryan & Wheeler, 1977). Students in cooperative groups are also far more likely to name friends and help students of different race and ableness than students in traditional educational settings (Gillies, 2004; Hansell & Slavin, 1981; Putnam et al., 1996; Weigel et al., 1975). The research also suggests that students who are allowed to learn in their preferred setting, whether with others or alone, have higher levels of achievement than when they learn in a non-preferred setting (Dunn et al., 1990; Owens & Barnes, 1982; Tan et al., 2005). Therefore, teachers must decide how to structure their curricula to include learning opportunities that address different student learning preferences.

Because the majority of students in all three accountability structures had higher gains in academic achievement than students taught traditionally, there are other implications that a teacher must consider when choosing which method to employ. Teachers need to decide what else it is that students should gain from their learning experience. Should they gain a lesson in the value of independent reward through group effort, or should they instead gain a lesson in the value of teamwork and interdependence? These are the types of questions teachers must address when employing cooperative methods. Teachers must also ask similar questions about their social and affective goals for their instructional methods. The research suggests that students who work in cooperative groups are far more likely to help one another and far less likely to engage in competitive behaviors. Therefore, teachers must decide if and when they want competition to

have a place in the classroom, or if they should instead use a cooperative strategy. Finally, cooperative methods may have a place in building classroom community. Students who learn and grow through cooperative interdependence may develop a stronger sense of community than students in a competitive or individualistic classrooms.

As mentioned throughout this paper, cooperative learning has been thoroughly researched in recent decades. Despite this, there are still avenues for further research. Firstly, more research on its effects on secondary-level students needs to be conducted. Although there has been an increase in research on secondary-level populations in recent years, the body of research is relatively small when compared to the research on elementary-level students. Another area for further research is on the effects of failing in cooperative learning settings. Cooperative learning is designed to maximize the number of successful groups and therefore successful students. But there are still groups that fail in cooperative settings, and little or no research has been done on the affective consequences of that failure. Finally, an area for future research may be the ways in which students perceive one another's standard or non-standard dialect in a cooperative setting. Research suggests that students are more accepting of students of other races and ethnicities after working cooperatively. It would therefore be useful to understand what the effects are on dialect perception, particularly in the present standards-based educational climate in which students are challenged to obtain proficiency in standard English dialect.

References

- Abrami, P. C., & Chambers, B. (1996). Research on cooperative learning and achievement: Comments on Slavin. *Contemporary Educational Psychology, 21*(1), 70–79.
- Adeyemi, B. A. (2008). Effects of cooperative learning and problem solving strategies on junior secondary school students' achievement in social studies. *Electronic Journal of Research in Educational Psychology, 6*(3), 691–708.
- Anderson, P. (2007). *What is web 2.0? Ideas, technologies and implications for education*. Retrieved from Joint Information Systems Committee website: www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf
- Arends, R. (1997). *Classroom Instruction and Management*. Boston: McGraw-Hill.
- Bratt, C. (2008). The Jigsaw classroom under test: No effect on intergroup relations evident. *Journal of Community & Applied Social Psychology, 18*(5), 403-419.
- Calderón, M., Hertz-Lazarowitz, R., & Slavin, R. (1998). Effects of bilingual cooperative integrated reading and composition on students making the transition from Spanish to English reading. *The Elementary School Journal, 153*–165.
- DeVillar, R. A., Faltis, C., & Cummins, J. (1994). *Cultural diversity in schools: From rhetoric to practice*. Albany, NY: SUNY Press.

- DeVries, D.L., Edwards, K J., & Wells, E.H. (1974). *Teams-games tournament in the social studies classroom: Effects on academic achievement, student attitudes, cognitive beliefs, and classroom climate* (Report Number 173, Center for the Study of Social Organization of Schools). Baltimore, MD: Johns Hopkins University.
- Dewey, J. (1997). *Experience and Education*. Kappa Delta Pi lecture series. New York: Simon & Schuster.
- Dunn, R., Giannitti, M. C., Murray, J. B., Rossi, I., Geisert, G., & Quinn, P. (1990). Grouping students for instruction: Effects of learning style on achievement and attitudes. *The Journal of Social Psychology, 130*(4), 485–494.
- Durukan, E. (2011). Effects of cooperative integrated reading and composition (CIRC) technique on reading-writing skills. *Educational Research and Reviews, 6*(1), 102–109.
- Ghaith, G. (2001). Learners' perceptions of their STAD cooperative experience. *System, 29*(2), 289-301.
- Ghaith, M. G., Bouzeineddine, A. R. (2003). Relationship between reading attitudes, achievement, and learners' perceptions of their jigsaw II cooperative learning experience. *Reading Psychology, 24*(2).
- Gillies, R. M. (2002). The residual effects of cooperative-learning experiences: A two-year follow-up. *The Journal of Educational Research, 96*(1), 15-20.

- Gillies, R. M. (2004). The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction, 14*(2), 197–213.
- Gocer, A. (2010). A comparative research on the effectivity of cooperative learning method and jigsaw technique on teaching literary genres. *Educational Research and Reviews, 5*(8), 439-445.
- Hagman, J. D., & Hayes, J. F. (1986). *Cooperative learning: Effects of task, reward, and group size on individual achievement* (Technical Report No. 174). Boise, ID: U. S. Army Research Institute for the Behavioral and Social Sciences.
- Hansell, S., & Slavin, R. E. (1981). Cooperative learning and the structure of interracial friendships. *Sociology of Education, 54*(2), 98-106.
- Holliday, D. C. (2000, April). *The development of jigsaw IV in a secondary social studies classroom*. Paper presented at the 2000 meeting of the Midwest Educational Research Association (MWERA), Chicago, IL.
- Johnson, R. T., & Johnson, D. W. (1994). An overview of cooperative learning. *Creativity and collaborative learning*. Retrieved from http://teachers.henrico.k12.va.us/staffdev/mcdonald_j/downloads/21st/comm/BenefitsOfCL/OverviewOfCoopLrng_Benefits.html
- Johnson, D. W., & Johnson, R. T. (1996). Cooperative learning and traditional american values: An appreciation. *NASSP Bulletin, 80*(579), 63-65.

- 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, J., & Anderson, D. (1976). Effects of cooperative versus individualized instruction on student prosocial behavior, attitudes toward learning, and achievement. *Journal of Educational Psychology*, 68(4), 446-452.
- Johnson, David W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change: The Magazine of Higher Learning*, 30(4), 26-35.
- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Cooperative learning methods: A meta-analysis. *Minneapolis, MN: University of Minnesota*.
- Johnson, R. T., Johnson, D. W., & Tauer, M. (1979). The effects of cooperative, competitive, and individualistic goal structures on students' attitudes and achievement. *The Journal of Psychology*, 102(2), 191–198.
- Lew, M., Mesch, D., Johnson, D. W., & Johnson, R. (1986). Components of cooperative learning: Effects of collaborative skills and academic group contingencies on achievement and mainstreaming. *Contemporary Educational Psychology*, 11(3), 229-239.
- Lucker, G. W., Rosenfield, D., Sikes, J., & Aronson, E. (1976). Performance in the interdependent classroom: A field study. *American Educational Research Journal*, 13(2), 115–123.

- Mattingly, R. M., & VanSickle, R. L. (1991). Cooperative learning and achievement in social studies: Jigsaw II. *Social Education, 55*(6), 392-395.
- Miller, P. H. (2009). *Theories of Developmental Psychology* (Fifth ed.). Worth Publishers.
- Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1983). Evaluation of a cooperative learning strategy. *American Educational Research Journal, 20*(4), 687–696.
- Nguyen, P.M., Elliott, J. G., Terlouw, C., & Pilot, A. (2009). Neocolonialism in education: Cooperative Learning in an Asian context. *Comparative Education, 45*(1), 109-130.
- Owens, L., & Barnes, J. (1982). The relationships between cooperative, competitive, and individualized learning preferences and students' perceptions of classroom learning atmosphere. *American Educational Research Journal, 19*(2),
- Phuong-Mai, N., Terlouw, C., Pilot, A., & Elliott, J. (2009). Cooperative learning that features a culturally appropriate pedagogy. *British Educational Research Journal, 35*(6), 857-875.
- Putnam, J., Markovchick, K., Johnson, D. W. & Johnson, R.T. (1996). Cooperative learning and peer acceptance of students with learning disabilities. *Journal of Social Psychology, 136*(6), 741-752.
- Putnam, J. W. (1997). *Cooperative Learning in Diverse Classrooms*. Upper Saddle River, N.J: Merrill.

- Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*(2), 223.
- Ryan, F. L., & Wheeler, R. (1977). The effects of cooperative and competitive background experiences of students on the play of a simulation game. *Journal of Educational Research, 70*(6) 295-299.
- Şahin, A. (2010). Effects of jigsaw II technique on academic achievement and attitudes to written expression course. *Educational Research and Reviews, 5*(12), 777–787.
- Shachar, H., & Sharan, S. (1994). Talking, relating, and achieving: Effects of cooperative learning and whole-class instruction. *Cognition and Instruction, 12*(4), 313-353.
- Sharan, S. (1988). *Language and learning in the cooperative classroom*. New York: Springer-Verlag.
- Sharan, Y. (2010). Cooperative learning for academic and social gains: Valued pedagogy, problematic practice. *European Journal of Education, 45*(2), 300-313.
- Slavin, R. E. (1978). Student teams and comparison among equals: Effects on academic performance and student attitudes. *Journal of Educational Psychology, 70*(4), 532-538.
- Slavin, R. E. (1979). Effects of biracial learning teams on cross-racial friendships. *Journal of Educational Psychology, 71*(3), 381-387.

- Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin*, *94*(3), 429–445.
- Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology*, *21*(1), 43–69.
- Slavin, R.E., Madden, N. A., & Stevens, R. J. (1989). Cooperative learning models for the 3 R's. *Educational Leadership*, *47*(4), 22–28.
- Slavin, R. E., & Oickle, E. (1981). Effects of cooperative learning teams on student achievement and race relations: Treatment by race interactions. *Sociology of Education*, *54*(3), 174-180.
- Slavin, Robert E., Sharan, S., Kagan, S., Lazarowitz, R. H., Webb, C., & Schmuck, R. (Eds.). (1985). *Learning to cooperate, cooperating to learn*. New York: Plenum.
- Stevens, R. J. (2003). Student team reading and writing: A cooperative learning approach to middle school literacy instruction. *Educational Research and Evaluation*, *9*(2), 137–160.
- Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly*, *22*(4), 433–454.
- Stevens, R. J., Slavin, R. E., & Farnish, A. M. (1991). The effects of cooperative learning and direct instruction in reading comprehension strategies on main idea identification. *Journal of Educational Psychology*, *83*(1), 8-16.

- Stewart, D. C. (1988). Collaborative learning and composition: Boon or bane? *Rhetoric Review*, 7(1), 58-83.
- Tan, I. G.-C., Sharan, S., & Lee, C. K.-E. (2005). Students' perceptions of learning geography through group investigation in Singapore. *International Research in Geographical & Environmental Education*, 14(4), 261–276.
- Tan, I.G.-C., Sharan, S., & Lee, C.K.-E. (2006). *Group investigation and student learning: An experiment in Singapore schools*. London: Marshall Cavendish Academic.
- Tan, I. G.-C., Sharan, S., & Lee, C. K.-E. (2007). Group investigation effects on achievement, motivation, and perceptions of students in Singapore. *Journal of Educational Research*, 100(3), 142-154.
- Totten, S., Sills, T., Digby, A., & Russ, P. (1991). *Cooperative learning: A guide to research*. New York: Garland.
- U.S. Department of Education, Institute of Education Sciences. (2011). *Student team reading and writing*. Retrieved from What Works Clearinghouse <http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid+591>
- Weigel, R. H., Wiser, P. L., & Cook, S. W. (1975) The impact of cooperative learning experiences on cross-ethnic relations and attitudes. *Journal of Social Issues*, 31(1), 219-244.