Effects of Co-Teaching on the Biology Achievement of Typical and At-risk Students Educated in Secondary Inclusion Settings

Polly G. Haselden Ph.D.
khaselden@fmarion.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/ejie

Part of the Curriculum and Instruction Commons, Curriculum and Social Inquiry Commons, Disability and Equity in Education Commons, Special Education Administration Commons, and the Special Education and Teaching Commons

Repository Citation

Effects of Co-Teaching on the Biology Achievement of Typical and At-risk Students Educated in Secondary Inclusion Settings

Polly G. Haselden, Ph.D.

Francis Marion University

ABSTRACT

School accountability is at the forefront of education with the recent passage of the *No Child Left Behind Act* (NCLB) in January 2001. One well-known instructional strategy, co-teaching has the potential to improve the academic performance of students (i.e., typical and at-risk) educated in general education classrooms. A co-teaching intervention that included operationalized components of instructional delivery and a support class was compared to the traditional instructional delivery of students receiving science instruction from a general education teacher alone in four high school biology classrooms. Results indicated that there were no significant differences between the groups of students educated in the co teaching and typical settings overall. However, post hoc analyses showed significant differences between settings for: (a) exceptional students, (b) students with 504 plans, and (c) students receiving free or reduced lunch. Limitations, future research, and recommendations for future investigations are offered.

Statement of the Problem

The inclusion movement embodies one of the most important initiatives in the field of special education today (Deshler & Schumaker, 1993). The importance is confirmed by several pieces of disability legislation that have been written or reauthorized since 1990 (i.e., The Americans with Disabilities Act of 1990, The Individuals with Disabilities Education Act, 1990 and 1997, and The Rehabilitation Act Amendments of 1992). School accountability is at the forefront of education with the recent passage of the *No Child Left Behind Act* (NCLB) in January 2001. This legislation mandates that all children will learn, and no child will be left behind, regardless of educational setting, disability, and socioeconomic status.

Because of accountability considerations and the number of students with disabilities who are no longer pulled out of the general curriculum for support services, it is crucial that effective instructional strategies be established in inclusive settings. In 1999, it was reported that 47% of special education students spent 80% of their day in regular classes, which is a 10% increase from 1989 (NCES, 2001-034, table 53). Within the report, it was also noted that the number of special education students being educated in general education classrooms is growing faster than total school enrollment. Since school accountability must be proven through the successes of each and every child, using teaching methods that will work for all learners is vital.

However, designing effective instruction programs for diverse learners in general education settings is a formidable task. First, there is an expectation of all learners to meet curriculum standards that have been adopted by individual states and specialized organizations (Erickson, Ysseldyke, & Thurlow, 1997). Second, diverse learners are expected to acquire mass amounts of information and authentically incorporate these skills within subject areas (e.g., history and science) (Kameenui & Carnine, 1998).

The term cooperative teaching was coined to represent this collaborative idea for instruction. Later, Cook and Friend (1995) changed the name cooperating teaching to co-teaching, which is defined as instruction occurring between special educators and general educators and consisting of "two or more professionals delivering substantive instruction to a diverse, or blended group of students in a single physical space" (Cook & Friend, 1995, p.1). More specifically, co-teaching occurs when a special educator and general educator are instructing in the same setting, using the same materials, and are both participating equally in the service delivery (Vaughn, Bos, & Schumm, 2000).

Purpose

The purpose of this current investigation was to determine if there were any differences in academic achievement between classrooms implementing the co-teaching service delivery model and those classrooms that use the traditional instructional delivery of having one instructor. Of primary importance was to operationally define a pre-existing intervention (i.e., the Co-teaching/ Support Model (CSM)) so that data could be collected in biology classes to determine the academic achievement of typical students and those who were considered to be at-risk. A crucial component of this research was the collected procedural reliability, which would allow researchers the opportunity to replicate the process. To date, there is no research to validate an operationalized model of co-teaching to implement fidelity measures.

Given the emphasis of evidence-based practices as a result of the NCLB (2001), this research investigated the effectiveness of co-teaching as an instructional strategy in inclusive settings. Data was collected in academic areas using the instructional strategy to support its use to promote student achievement. Furthermore, since science will be factored into accountability considerations by 2006-2007 (NCLB, 2001), science classrooms using co-teaching as an instructional strategy were investigated.

The initial primary dependent variable was the students' scores on the End of Course examinations. These tests were given in the settings implementing the Co-Teaching/ Support Model in biology. This course was chosen for two reasons. Biology was an academic area that will be measured according to the new requirements of the NCLB (2001), which mandates that science be assessed annually by 2006-2007. A second reason for examining the biology EOC

scores of students educated in academic settings using the CSM model was to determine if any differences were found between the various at-risk groups within classes. Typically, co-teaching has been used as an instructional strategy for educating students with disabilities in inclusion settings.

The independent variable in this study was the CSM intervention. The pre-existing intervention was implemented in two biology classes, and data were collected from the same number of similar classes in biology in the control setting. The study implemented a quasi-experimental, posttest only design to investigate if academic achievement differences existed in biology among the groups and settings involved in this research.

Significance of the Study

Co-teaching. Even though co-teaching is a popular service delivery method in inclusion settings, there has been a dearth of research supporting its effectiveness on student achievement, with virtually no studies supporting the efficacy of using it as a service delivery option (Rice & Zigmond, 2000). However, school systems continue to embrace this service delivery method as a viable method of instruction in inclusion settings (Dieker & Murawski, 2003).

Another limitation to the studies included within the meta-analysis was that only two of the studies selected reported how co-teaching was implemented (i.e., Lundeen & Lundeen, 1993) and none of the studies reported treatment fidelity. Weiss and Brigham (2000) found additional problems associated with the existing literature on co-teaching: (a) no consistency in outcome variables, (b) no consistency in an operationally defined variable of co-teaching, (c) lack of

experimental and control conditions, (d) the findings of most studies are based on teacher personality, and (e) most research designs assessed change in behaviors qualitatively.

As reported in the meta-analysis conducted by Murawski and Swanson (2001), there were only three studies on co-teaching that have been conducted at the secondary level (Lundeen & Lundeen, 1993; Walsh & Snyder, 1993). Of these three studies, none were conducted in the area of science instruction. Science instruction in special education has been a research area that has been overlooked, especially at the secondary level (Scruggs & Mastropieri, 1994) there if a need to conduct initial research in this area. However, with no research to validate co-teaching science at the secondary level and little research conducted in secondary science instruction in special education in general, the need exists for research that operationally defines co-teaching with procedural reliability measures to determine if this type of instruction can be validated as an appropriate instructional delivery method for increased academic achievement of all students educated in general education settings.

Accountability. Confounding the issue of including diverse learners in general education classes are the addition of high stakes testing and the important consequences that these assessments have on students educated in inclusive settings. High stakes testing is a result of school districts attempting to meet a wide variety of standards, and the instruments that are being used vary across the nation (Langenfeld, Thurlow, & Scott, 1997). Importantly, high stakes assessments have direct and important consequences for students, educators, school systems, and parents as they are used to determine promotion or retention considerations for students educated in general education classrooms. Nowhere is that more obvious than at the secondary level, where school completion is imminent. Graduation is the summation of the United States'

education process, and our completion rates are used as indicators as to our competitiveness in the overall society (Thurlow, Ysseldyke, & Anderson, 1995). Since graduation requirements are not mandated by the U.S. Department of Education (Thurlow et al., 1995), states have been able to develop and require students to complete any individual or combination of the following: (a) Carnegie units (specific number of classes passed in selected areas), (b) minimum competency exams, (c) exit examinations, and (d) a series of benchmark examinations. As a result, educators, parents, students, and the public have a keen interest in how well schools are preparing students to meet state requirements for graduation.

Students with disabilities are being held accountable for their performances on large-scale assessments for accountability considerations; however, data on the achievement of students with disabilities on these assessments are hard to find (Bielinski & Yssledyke, 2000). The limited data that have been analyzed on the achievement of students with disabilities are discouraging (Thurlow, Ysseldyke, Nelson, & Teelucksingh (2000). For instance, in a multi-state study conducted by Ysseldyke, Thurlow, Langen- feld, Nelson, Teelucksingh, and Seyfarth (1998), it was reported that general education students consistently outperformed students with disabilities. In addition to group differences in achievement, the achievement gap between the groups continues to increase over time (Trimble, 1998), making secondary schools crucial settings for determining appropriate practices for overall student achievement. Recently, it was found students without disabilities score an average of 37% higher than students with disabilities on large-scale assessments (Thurlow et al., 2000).

Data analyses of students with disabilities who were proficient on general state reading assessments indicate a critical need for future research. Educators must seek out instructional

strategies, programs, and supports to increase the academic achievement and percentages of proficiency in subject areas for students with disabilities. Research must also focus on closing the gap between the achievement of students with disabilities and students without disabilities, especially at the secondary level (Trimble, 1998). There is no pre-existing research that has investigated the academic achievement of students with disabilities as measured by a high stakes assessment. Only reports from the National Center for Educational Outcomes (NCEO) have determined how students with disabilities are performing on high stakes tests and how they compare to their typical peers. Now research is needed to improve the identified problems.

It is imperative to improve educational outcomes by identifying and validating appropriate instructional practices for diverse learners in general education settings. Additionally, it is crucial to determine interventions that are appropriate to academic area and school level. Co-teaching could be a viable intervention in general education settings, although research has not validated its use (Murawski & Swanson, 2001). Combining the efforts of special education and general education teachers could serve to increase the academic achievement of all students. Combining specializations within an operationalized co-teaching intervention may be one of the first steps towards increasing student achievement and closing the gap between groups educated in general education classrooms.

Secondary school has been identified as a crucial area in closing the gap between students (Trimble, 1998). However, only three empirical studies exist on the effects of co-teaching on student achievement in secondary settings and none were conducted in the area of science (Murawski & Swanson, 2001). Since science will be a new mandated assessment area (NCLB, 2001), researching a specific co-teaching strategy in a science setting at the secondary level will

provide a setting of unexplored research. Furthermore, since there is no research to validate instructional practices on improving the achievement for students with disabilities in general education classrooms as measured by high stakes testing for accountability considerations, using state administered high stakes assessment measures in the area of science will initiate the formation of a non-existent data base. Ultimately, this research could serve to identify an instructional strategy that will improve the academic achievement of students educated in inclusive settings as measured by high stakes tests administered for state accountability considerations.

Research Questions

The research questions posited for the present study were based on a review of the existing research on co-teaching (e.g., Marawaski & Swanson, 2001) and accountability considerations (e.g., Quenemoen, Lehr, Thurlow, & Massanari (2001). The focus was to address the following questions throughout the study.

1. Were there significant differences in the academic scores in biology among group of students who were educated in the co-teaching and control settings?

2. Were there significant differences in the academic scores in biology between typical and at-risk students who were educated in the co-teaching and control settings?

3. Were there interaction effects between type of student and treatment condition?

4. Were there significant differences in the academic scores in biology between students with disabilities who were educated in the co-teaching and control settings?

5. Were there significant differences in the academic scores in biology between students who receive free and reduced lunch who were educated in the co-teaching and control settings?

6. Were there significant differences in the academic scores in biology between African Americans who were educated in the co-teaching and control settings?

Summary

Research in the area of co-teaching is clearly needed to examine if it is a viable service delivery option for students who are at-risk, especially at the secondary level. The intent of this study was to advance the existing limited knowledge in co-teaching as an instructional strategy and potentially identify the possibility of using it with specific subgroups as identified by the NCLB (2001). Acquisition of student achievement as a result of co-teaching with a support model was examined specifically. Results of this study may have implications for teachers involved with high stakes testing and those educating diverse learners in inclusive settings.

Methodology

Overview

This study employed a quasi-experimental, non-equivalent, posttest only design. The goal of quasi-experimental research is to examine cause and effect by observing how participants react to phenomena (Cook & Campbell, 1979). Quasi-experimental research is different from true experimental research, as participants are not randomly assigned (Mertler & Vannatta, 2002). In educational settings, it is oftentimes not possible to randomly assign participants because of difficulty or ethical considerations (Hadley & Mitchell, 1995).

The independent variables in this study were the method of teaching (i.e., CSM or traditional) and the type of student (i.e., at-risk or typical); the dependent variable was defined as student achievement in biology on the North Carolina End of Course (EOC/B) examination. The researcher collected data at the end of an academic semester (i.e., 90 days) to examine the effects of implementing the Co-teaching Support Model (CSM) throughout a school semester. The setting in which the CSM occurred was designated as the "experimental setting;" the general education setting without the CSM intervention was designated as the "control setting."

All of the students involved in this study were enrolled in secondary biology classrooms, and most were tenth grade students. There was the possibility that upper classmen who have previously failed the course could be enrolled in any of the groups. Students were educated at one of two high schools during the 2003 school-year, and the students took the state administered End of Course (EOC) examination at the end of the 2003 semester in December. The EOC/B scores were analyzed using a two-way analysis of variance (ANOVA) to determine differences in academic achievement.

Participants

The participants in this study consisted of students attending biology classrooms at two high schools in the south central part of North Carolina. All of the typical students in the experimental condition and all of the participants in the control condition were assigned to classes in a lottery format as a result of the master schedule component of the Student Information Management Services (SIMS) computer system.

Experimental. The experimental classrooms were at one of the high schools. These students were educated in biology classes implementing the co-teaching model. Thus, students in

classes with co-teaching served as the experimental group. All data were obtained from the

SIMS computer system. Typical students represented an estimated 50% of the experimental

participants, while at-risk students represented the other estimated 50%. Within the at-risk

category, students with disabilities represented an estimated 50% and other subgroups (i.e., 504

and FDR) represented the other 50%. A special education teacher and a general education

teacher provided the instruction for these students. Students represented typical and at-risk

students educated simultaneously for biology instruction. The experimental condition took place

in a small, rural high school with a total enrollment of approximately 800 students. The

composition of the students attending the high school was African Americans (10%), Caucasians

(87%), and others (3%). Students receiving free and reduced lunch represented about 35% of the

population.

Control. The control classrooms were at the second high school in the study. At this

setting, students were educated by general education teachers. Two classes were selected to serve

as the control based on the criteria that they have comparable numbers of students with each at-

risk group (i.e., FDR, 504, EC, and racial minorities). Students placed into control classes

represented general education and at-risk students. Typical students represented an estimated

50% of the experimental participants, while at-risk students represented the other estimated 50%.

Within the at-risk category, students with disabilities represented an estimated 50% and other

subgroups (i.e., 504, FDR, and Race) represented the other 50%.

The control condition took place in a medium-sized, rural high school with a total

enrollment of about 1300 students. The composition of the students attending the high school

was African Americans (24%), Caucasians (68%), and others (8%). Students receiving free and reduced lunch represented about 28% of the population.

Comparability. Prior to this research endeavor, potential outliers were screened by using pre-existing achievement data in reading and math for all of the participants (i.e., End of Grade (EOG) reading and math scores). A participant was considered an outlier if his or her scores were three standard deviations above or below the overall mean scores on either test and was eliminated from the study. In addition, an ANOVA was run to determine any differences in mean scores on the EOG in reading and in math between the experimental and control settings. If differences were found, another biology class was selected.

First, the number of students identified as FDR, ESL, EC, 504, and racial groups were counted in each setting and compared for equivalence considerations. Students' socioeconomic status between the experimental and control condition was based on free and reduced lunch eligibility. Socioeconomic status between the students receiving free and reduced lunch was considered comparable between the experimental and the control if the difference between the numbers was less than five.

Additionally, students with disabilities were measured by the number of students who were identified as exceptional students and received educational support through special education. The numbers of students with disabilities were considered comparable between the experimental and the control setting if the difference between the numbers of students was less than five.

Comparability was measured for at-risk students by determining the number of students who received academic and behavioral support as a result of a 504 plan created by the school

services management team (SSMT). The numbers of students who were considered to be at-risk were considered comparable between the experimental and the control setting if the difference between the numbers of students was less than five. African American students were measured by the number of students who represented that group. The NCLB (2001) requires subgroup performance to be measured if members within a specific group total 40 or more students. These populations represented the largest minority population in this school district, and adequately yearly progress (AYP) was reported for this subgroup for accountability considerations. Therefore, the number of students who were identified by SIMS to be African American was considered comparable between the experimental and the control setting if the difference between the numbers of students was less than five.

Because quasi-experimental measures were used to evaluate group comparisons, specific variables were analyzed to determine the equivalence between the experimental and control conditions. Control and experimental classrooms were analyzed for comparability considerations by calculating the number of students who were identified in the following subgroups per setting: (a) students' socioeconomic status, which was measured by the number of students who received free and reduced lunch (b) students with disabilities, which was measured by numbers representing specific disabilities, (c) students who were considered to be at-risk, as measured by the number of students who were provided additional academic support through 504 services, and (d) students representing African American groups as defined by SIMS. Data were analyzed using descriptive statistics to provide group information to determine the equivalence of the experimental and control groups.

If the potential control settings did not meet the criteria set forth for equivalence, then

two more biology classrooms, not implementing the CSM or other team teaching methods, were recruited to participate and the comparability study was repeated.

Instrumentation

Academic achievement in biology was measured using the End of Course (EOC) examination in biology (EOC/B) of the North Carolina testing and accountability program. The EOC/B was designed as a curriculum-based achievement test to measure students' academic skills within the specific subject of biology. The instrument assessed the newly revised biology curriculum (i.e., 1999) of the *North Carolina Standard Course of Study.* The test measured students' knowledge of biology principles and concepts, laboratory simulations and activities and skills in relating these into practical experiences. The 1999 EOC/B incorporated more processing information and higher order thinking skills as compared to the old version of the EOC/B.

Test Description. The 88 multiple-choice items on the EOC/B were derived from the content objectives of the Science Standard Course of Study (See Table 2). The EOC/B was designed to be administered during a fixed amount of time within the last week of a semester for students on a block or summer school schedule and within the last two weeks of school for students educated within a traditional schedule. Three forms of the EOC/B were administered per classroom to gather information for planning and curriculum evaluations.

Test Development. Expert biology teachers developed items on the EOC/B during the 1999-2000 and 2000-2001 school years. NCDPI field-tested the items during the fall and spring semesters in 2000-2001. Participants (N=24,250) were randomly selected throughout the state of North Carolina to participate in the field-testing of the EOC/B. A revised edition of the EOC/B was implemented throughout the state in 2001-2002 for the first time.

Scores. The primary dependent variable for this research consisted of students' scale scores on the EOC/B. Initially, data obtained from the EOC/B tests were collected in the form of raw scores and were converted to scale scores. This conversion allowed for the EOC/B to be equated (NCDPI, 2002). However, results of the EOC/B were reported to school systems as scale scores and achievement levels. The scale was designed to have a range of 20 – 80 with a mean of 50 and a standard deviation of 10. The scale scores were used to provide a comparison of scores from test to test.

Reported achievement levels allowed for students' scores to be compared to a predetermined standard. The EOC/B achievement levels were determined by using the contrasting groups' method. The EOC/B achievement level matched a range of scale scores for each level. The levels are represented in Table 3.

Reliability and Validity. Reliability refers to the consistency of scores obtained by a person, when examined by the same test with different equivalent items or on different occasions; whereas, internal consistency reliability examines the extent to which an assessment measures a basic concept (NCDPI, 2003). For the EOC/B, internal consistency was determined by using the coefficient alpha (α) procedure. The coefficient alpha was used to set the upper limit of the reliability of the EOC/B constructed in terms of the domain-sampling model (NCDPI, 2003). Coefficient α estimates were found to be .88 for the EOC/B pretest and .94 for the EOC/B test.

The standard error of measurement was determined for the EOC/B. The magnitude of the standard error of θ (i.e., students' estimated achievement level) was determined by the following: (a) the number of test items, (b) the quality of the test items, and (c) the match between student

ability and item difficulty. Table 4 represents the standard error of measurement ranges for scores on the EOC/B. Measures of standard error were typically 2 to 3 points. Extreme scores outside of the two standard deviations of the mean were associated with less measurement accuracy (NCDPI, 2003).

Validity is the degree to which a test measures what it asserts to measure. Within test validity are three interrelated components: (a) content validity, (b) criterion-related validity, and (c) construct validity. Content validity for the EOC/B was built into the EOC/B during the development of the measure. All of the items included in the EOC/B were aligned with the North Carolina *Standard Course of Study* in biology and reviewed by expert biology teachers (NCDPI, 2003).

Criterion-related validity was established for the EOC/B by using achievement levels that were based on the contrasting groups' method of standard setting. This method involved teachers assigning expected achievement levels to the students. During the field-testing of the EOC/B, teachers categorize their students on the basis of "absolute" levels. These informed judgments were based on the teacher's experiences with the students throughout the school year. During the standard setting process, teachers (N>650) judged the perceived achievement performances of approximately 50,000 students across all EOC subject areas. The results were similar across subjects for percentages of students that were assigned to achievement level (see Table 5).

Construct validity determines the extent to which the test may be said to measure a theoretical construct or trait (NCDPI, 2003). North Carolina Open-Ended Tests in Biology (OET/B) were designed to evaluate higher level thinking skills by requiring students to relate or demonstrate acquired knowledge beyond the recall level. University professors, curriculum specialists, testing consultants, and teachers determined scores for the OET/B. Items (N=4) on

the OET/B were field tested in March 1994 with randomly selected students (*N*=1000) throughout the state of North Carolina to examine items performance (i.e., score distribution) and to improve the scoring OET/B scoring rubric. A second field-testing was conducted to verify the scoring rubrics. Results were analyzed by using the Samejima's graded item response theory model. In May, the same students, approximately 200 per matched field test form, were administered the OET/B and a multiple-choice field test. Results indicated that the correlations between the scores on the open-ended items and multiple-choice total scores ranged from 0.24 to 0.51, with a mean of .40.

Research Design

This study used a quasi-experimental, nonequivalent, posttest only design to examine the differences in academic achievement in biology among students who were educated in inclusive settings using the Co-teaching/Support Model (CSM) intervention to groups of students in the same grade level who were educated in biology classrooms without the CSM intervention. The data were analyzed using a two-way analysis of variance. The factors of the two-way ANOVA were the type of intervention and the type of student.

Procedure

Intervention. Classrooms in the experimental condition followed the *Standard Course of Study* for biology. One general education and one special education teacher delivered instruction for 90 minutes each day for 90 days during the fall semester. The CSM intervention operationalized how instruction was delivered within the experimental condition.

Specifically, the CSM intervention was an intervention that has been used in inclusion classes at a rural high school in south central North Carolina for ten years. It was developed to support students who were considered to be at risk within and outside of inclusion settings. The intervention consisted of specific components with an additional support class for the students who were educated in the CSM settings.

80% rules. Co-teachers implementing the CSM intervention jointly planned instruction. This was done during planning periods or before or after school and was documented. Also, instruction occurred using the 3X3X3 strategy for 80% of an instructional unit (i.e., Biomes in biology). The 3X3X3 strategy broke down the block schedule of time followed in the settings into three 30-minute blocks. The "3s" represented review, new instruction, and application, although this did not have to occur in any specific order. It was important to note that other strategies have been incorporated into the 80% instructional strategy. Due to testing or reviewing material not yet achieved, teachers could plan for instruction during the course of a unit by implementing the 2X1 (i.e., one 45 minutes of review and one 45 minutes of testing) and 1X (i.e., one complete 90 minutes of testing, usually occurring on EOC or other high stakes testing dates).

50/50 instruction. Co-teachings participating in the CSM intervention provided equal amounts of instruction. They chose from one of the following: (a) one teacher was responsible for the primary instruction for a complete unit and then that responsibility alternated to the other teacher and (b) teachers rotated daily providing the primary instruction.

50/50 evaluation. Evaluation of students educated in the CSM settings was shared between both teachers. Both teachers documented how the following instructional tasks would

break down (i.e., grading, parental contacts, updating portfolio IEP and 504 pieces (i.e., products, modifications used, etc.).

Support class. Students who were considered at risk with documented disabilities educated in the CSM model took an elective support class offered by the special education department (i.e., Curriculum Assistance). Students with 504 plans educated in the CSM model took an elective support class offered by the dropout prevention counselor. Both classes followed the same format. Students defined what they needed to do for other courses, since their work would be individual to every student with the exception of the similar co-taught academic classes.

For 90 min students worked on course assignments, completed tasks offered in class with additional modifications if needed, and were instructed according to their goals of their Individualized Education Plan (IEP) or 504 plan. Portfolios of student samples were kept for all students educated within the CSM model. Typical instructional strategies for the support classes were strategy instruction and direct instruction (DI) for specific IEP or 504 objectives, though no formal DI program had been adopted by either class. It should be noted that students within the support classes worked on assignments from all of their classes. These classes were not specific to biology instruction.

The CSM co-teaching model was implemented five days a week in 90-minute block schedule format for 90 days (i.e., one semester). Classroom instruction for the experimental groups followed the first three components of the CSM intervention and occurred within the general education classroom. The support component of the CSM was implemented five days a week for 90 days for selected at-risk students during another class period. The support

component occurred outside of the general education classroom for 90-minutes a day for 90 days.

Control Setting. Participants in the control setting were educated in general education biology classrooms. In each setting, the general education teacher assumed full responsibility of biology instruction. Students educated in these classes were comprised of typical students and those who were considered to be at-risk (i.e., FDR, EC, 504, and African American). Consistent with the experimental condition, these biology classrooms followed the *Standard Course of Study* in biology. Instruction occurred for 90 minutes for 90 days during the fall semester. Instructional delivery in typical biology classrooms was usually comprised of lecturing, labs, and cooperative group learning.

Procedural Reliability. Procedural reliability was collected by the chair of the exceptional children's department at the experimental setting. The co-teachers completed checklists documenting the time and their roles in implementing the first three components of the CSM intervention daily (See Appendix A). The department chair collected procedural reliability by observing the co-teachers implementing the intervention weekly. Reliability data were collected as the department chair completed a checklist as the CSM intervention was implemented by the co-teachers (See Appendix B). Procedural reliability for planning and evaluation was measured by analyzing the co-teacher planning sheets, and procedural reliability for instruction was collected by the department chair randomly observing the CSM intervention and comparing actual instruction to pre-planned instruction.

Data Collection and Procedures. Students in the each setting were administered the End-of-Course examination in biology upon the completion of the semester of biology. Each school's

Testing Coordinator trained teachers and proctors who administered the test. In each setting, teachers had proctors to ensure that testing procedures were followed by the guidelines issued by the North Carolina Department of Public Instruction.

Scale scores and achievement levels on the EOC/B in both the experimental and control settings were obtained from the Student Information Management System (SIMS). Scale scores were entered into an Excel spreadsheet. Following the initial collection of data, the data were transferred to the Statistical Package for the Social Sciences (SPSS) for analysis. Descriptive statistics were used to analyze the achievement levels obtained by the participants in the study. Because of the small sample sizes, student achievement by subgroups (i.e., free and reduced lunch and special education) was analyzed using nonparametric statistics. The Mann-Whitney was used to analyze scores by rank since the populations of the subgroups were normal due to expected numbers. To protect against a Type I error, a Bonferroni correction was made for the significance level. In this study, the level of significance was .017 for the three nonparametric tests.

Results

The purpose of this study was to investigate the effects of a co-teaching model on the academic achievement of diverse learners educated in inclusive classrooms. To examine this question, the performance of the experimental group, which received the co-teaching intervention, was compared to the performance of the control setting, which received the traditional instructional delivery of one educator. The type of student (i.e., typical or at-risk) and type of setting (i.e., inclusion or general education setting) were the independent variables, and the academic achievement scores on the EOC/B were the dependent variable. Additionally,

follow-up analyses examined differences between the experimental and control conditions for specific subgroups (i.e., socioeconomic status, ethnicity, 504 students, and students with disabilities).

Procedural Reliability

Procedural reliability data were collected for instructional sessions by a trained observer (i.e., expert in the exceptional children's department) using a procedural checklist developed directly from the intervention outlined in the methodology of the study. The observer checked the occurrence or nonoccurrence of each portion of the checklist (e.g., number of days of collaborative planning, type of instruction each day, primary instructor each day per type of instruction, and percentage of time performing instructional duties) to determine the fidelity of the treatment or the consistency with which the instructors followed the intervention as outlined.

Several steps were taken to ensure that the procedural reliability data were accurate and consistent. First, the researcher trained the observer, who was an expert in the area of exceptional children, in procedural data collection. The teacher in the co-teaching class also collected procedural reliability data and was trained by the researcher. To examine the interobserver agreement between the observer and the teacher, the percentage of agreement was calculated by dividing the number of agreements by the total number of items rated and multiplying by 100. Interobserver agreement mean percentages were the following: (a) collaborative planning (81%), (b) type of instruction (92%), (c) primary instructor (83%), and (d) instructional tasks (97%). The mean percentage of agreement was 88% (range = 81%-97%), which indicated reliable procedural ratings. Procedural reliability data were collected for 20% of the intervention implementation. The procedural reliability scores for nine weeks of the intervention are reported

in Table 6. The mean percentage in the intervention phase was 93% (range = 66% to 100%).

<u>*Demographic Characteristics of Participants*</u>

Participant scores (*N*=103) from the experimental (*N*=55) and control (*N*=48) settings were included in this study. Participants were composed of typical students at the experimental (*N*=25) and control settings (*N*=20) and at-risk students at the experimental (*N*=30) and control settings (*N*=28). Within the at-risk category, participants were placed into subgroups according to existing documentation. At the experimental site, students with 504 plans (*N*=9), students who received free or reduced lunch (*N*=11), students with disabilities (*N*=10), and African American students (*N*=11) made up the at-risk subgroup. At the control site, students with 504 plans (*N*=7), students who received free or reduced lunch (*N*=8), students with disabilities (*N*=12), and African American students (*N*=14) made up the at-risk subgroup. A break down of students by ethnicity by setting revealed that the experimental setting was composed of African American students (*N*=11) and Caucasian students (*N*=51). A break down of students by ethnicity by setting revealed that the control setting was composed of African American students (*N*=14) and Caucasian students (*N*=34). The number and percentage of students in each subgroup are reported in Table 2 by setting during this research.

Comparability Results between the Experimental and Control Conditions

To examine the equality of academic skills in the experimental and control groups before treatment was implemented, standardized test scores in reading and mathematics were compared. Means and standard deviations are reported in Table 3 for End of Grade examination scores in reading for students who were educated in the experimental and control settings. Scores on the End of Grade examination in reading (EOG/R) for students in the control setting were somewhat

higher (M=175.00) than the scores on the EOG/R for students educated in the experimental

setting (M=170.00). The assumption of homogeneous variances was satisfied (Levene's test,

F=.50, p=.49). The mean score for the students in the experimental group were not statistically

different from the mean scores for the students in the control setting (t=1.8, p=.07).

Means and standard deviations are reported in Table 4 for End of Grade examination

scores in math for students who were educated in the experimental and control settings. Scores

on the End of Grade examination in math (EOG/M) for students in the experimental setting were

higher (M=291.50) than scores on the EOG/M for students in the control setting (M=290.13).

The assumption of homogeneous variances was satisfied (Levene's test, F=1.50, p=.23). The

mean score for the students in the experimental group was not statistically different from the

mean scores for the students in the control setting (t=.82, p=.41).

Results by Research Question

Prior to the data analysis, the data were screened for outliers and normality of

distribution. There was one outlier (i.e., scores was greater than three standard deviations below

the mean). All analyses were conducted with the outlier included and not included, and the

results were the same; therefore, all the analyses reported included the outlier. Skewness test

indicated no serious departures for normality (i.e., all coefficients resulted in absolute values of

less than 1).

The first three research questions examined the effects of type of student and type of

setting on the EOC/B. A two-way analysis of variance was conducted to investigate academic

differences in type of student and type of instruction among biology students. The means,

standard deviations, and sample sizes for the experimental and control groups by typical and at-

risk are reported in Table 5. Levene's test for homogeneity of group variance was nonsignificant indicating the assumption of homogeneity of group variance as tenable. The two-way ANOVA results, presented in Table 6, showed a significant interaction effect ($F_{(1, 99)}$=6.34, p=. 013, partial η^2=0.06). There were no significant main effects found for type of student ($F_{(1, 99)}$=1.2, p=. 28) or type of instruction ($F_{(1, 99)}$=. 27, p=. 6).

The disordinal interaction is illustrated in Figure 1. To follow up on the statistically significant interaction, simple effect analyses were conducted to examine differences between types of students within each treatment. There was a statistically significant difference between the experimental and control groups for the at-risk students on EOC/B mean scores (F=5.32, p<.05). At-risk students in the experimental group scored on average much higher (M=57.77) than those in the control group (M=53.64), with an effect size of .72. There was not a statistically significant difference between the traditional students in the experimental (M=55.84) and control groups (M=58.55) on EOC/B mean scores (F=1.76, p>.05).

In the first follow up analysis, students who received free or reduced lunch were examined. Means and standard deviations are reported Table 7 for students who received free or reduced lunch in the experimental (N=11) and control settings (N=8). The assumption of homogeneous variances was satisfied (Levene's test, F=.40, p=.54). The mean score for the students with disabilities in the experimental group was significantly higher than the mean scores for the students with disabilities in the control setting (t=3.0, p=.01). There was a large difference between the students with disabilities in the experimental and control settings (d=1.35). Students who received free or reduced lunch did differ on their EOC/B scores in the experimental and control settings.

The second follow-up analysis examined the academic achievement of students with disabilities educated in the experimental and control settings. Means and standard deviations are reported in Table 8 for students with disabilities as indicated by their I.E.P. plans in the experimental (N=10) and control settings (N=12). The assumption of homogeneous variances was satisfied (Levene's test, F=1.0, p=.34). The mean score for the students with disabilities in the experimental group was significantly higher than the mean scores for the students with disabilities in the control setting (t=2.80, p=.01). There was a large difference between the students with disabilities in the experimental and control settings (g=1.17). Students who were labeled as special education did differ on their EOC/B scores in the experimental and control settings.

The third follow-up analysis investigated the academic achievement of African American students educated in the experimental and control settings. Means and standard deviations are reported for African American students educated in the experimental (N=11) and control settings (N=14). The assumption of homogeneous variances was satisfied (Levene's test, F=.16, p=.70). There was no difference in the means score for the African American students in the experimental and control setting (t=.10, p=.93).

The final follow up analysis examined the academic achievement of students with 504 plans educated in the experimental (N=9) and control settings (N=7). Means and standard deviations are reported in Table 10 for students with 504 plans in the experimental and control settings. The assumption of homogeneous variances was satisfied (Levene's test, F=4.32, p=.06). The mean score for the students with 504 plans in the experimental group was significantly higher than the mean scores for the students with 504 plans in the control setting (t=2.40, p=.03).

There was a large difference between the students with 504 plans in the experimental and control settings (*g*=1.21).

Follow-up analyses examined differences by subgroups of at-risk students educated in the experimental and control settings. Differences found for each subgroup analysis are reported in Table 11.

Percentage of Students at Each Achievement Level on the EOC/B

Achievement level descriptions by NCDPI (2003) indicated that students who scored a Level 1 did not have "sufficient knowledge of the skills of the subject to master a more advanced level within the same subject areas." Students who scored a Level 2 had an "inconsistent knowledge of the skills of the course and are minimally prepared to be successful at an advanced level in the same subject area." Students who scored a Level 3 had a "consistent knowledge of the skills of the course and are adequately prepared to be successful at an advanced level in the same subject area", and students who scored at a Level 4 had a "superior knowledge of the skills of the course and are very prepared to be successful at an advanced level in the same subject area."

Achievement levels for typical and at-risk students are reported in Table 12 on the EOC/B. Academic achievement levels corresponded with each student's score. At the experimental school, 5.5% of the students (*N*=3) scored a Level 1, and 31% of the students (*N*=17) scored a Level 2 which indicated that 36.5% of the students in the experimental condition did not meet the standard for passing. However, 58% of the students (*N*=32) at the experimental school scored a Level 3, and 5.5% of the students (*N*=3) at the experimental school scored a

Level 4. These scores indicated that 63.5% of the students educated at the experimental site met academic standards for passing the EOC/B.

At the control school 8.3 % of the students (*N*=4) scored a Level 1, and 29.2% of the students (*N*=14) scored a Level 2 which indicated that 37.5% of the students in the control condition did not meet the standard for passing the EOC/B. However, 60.4% of the students (*N*=29) at the experimental school scored a Level 3, and 2.1% of the students (*N*=1) at the experimental school scored a Level 4. These scores indicated that 62.5% of the students educated at the experimental site met the academic standards for passing the EOC/B. A chi-square analysis indicated that there was no statistically significant difference between the experimental and control groups on the achievement levels (Chi-square = 1.1, *df*=3, *p*>.05).

Additionally, the percentage pass rate (i.e., achievement level 3 and 4) for the EOC/B for the experimental and control groups by type of students are reported in Table 13. The results of Mann Whitney-U indicated that there were no differences between the percentage pass rate of the at-risk students in the control or experimental group (*Z*=-1.28, *p*>.05) and the typical students in the control or experimental group *Z*=-1.42, *p*>.05.

Social Validity Measures

Teachers completed a social validity checklist upon the completion of the semester in which this research was conducted. Scores on the checklist ranged from 1 (disagree) to 5 (strongly agree). Results indicated that teachers in the experimental condition felt adequately prepared to meet the needs of diverse learners in their classrooms. The first question "I am aware of the diverse learners that are educated in my classroom" resulted in positive responses from the teachers at the experimental site" and mixed responses from teachers at the control site. The

attitudes between the teachers at the experimental and control sites "I am aware of special modifications and accommodations for the diverse learners in my classroom" were comparable.

The next question "I am aware of other outside supports that the diverse learners in my classroom use on a daily or weekly basis" yielded favorable responses from teachers in the experimental condition and unfavorable responses in the control condition. "I enjoy including diverse learners in my general education classroom" indicated positive feelings for teachers in the experimental setting and indicated mixed feelings in the control setting. The next item "I feel that I am supported as I try to meet the educational needs of the diverse learners in my classroom" indicated more favorable attitudes in the experimental setting than in the control setting. The final item "I feel adequately prepared to meet the needs of diverse learners educated in my classroom" resulted in positive feelings for the teachers at the experimental condition and mixed reactions in the control condition.

Summary

Results indicated that there were no significant differences in the academic achievement on the EOC/B for groups of students educated in the experimental and control settings or typical students educated in those settings, but at-risk students had higher mean EOC/B in the experimental group than the at-risk students in the control group. In addition, significant differences were found for at-risk students in the following subgroups of students: (a) exceptional students, (b) students with 504 plans, and (c) students who received free or reduced lunch between settings. No significant differences were found for the African American subgroup between settings. Additionally, no significant differences were found in achievement levels for any of the groups or subgroups of students.

Discussion

Summary of Major Findings

Results indicated that at-risk students and specific subgroups of at-risk students educated in general education classes using the co-teaching model had higher average biology scores than a comparison group of at-risk students educated in traditional general education settings. Although there was no statistically significant difference between the experimental and control groups as a whole on the EOC/B, statistically significantly higher academic gains were noted for the following subgroups: (a) students with disabilities, (b) students with 504 plans, and (c) students receiving free or reduced lunch in the experimental setting. No significant differences in EOC/B means were for found between the experimental and control African American subgroup as a result of setting. Given the higher average academic scores across the at risk subgroups, (i.e., students with disabilities; students with 504 plans; and students receiving free or reduced lunch) these current findings provide teachers of diverse students an alternative instructional approach to enhance academic achievement.

High Stakes Testing and Accountability

One of the purposes of this research was to identify a strategy that could improve outcomes on large-scale assessments for students who are educated in inclusion settings. Prior research has shown a gap between performance of students with and without disabilities (Bielinski and Ysseldyke, 2000; Bielinski, et. Al, 2001; Thurlow, et. al, 1998; Trimble, 1998) on high stakes assessments.

A major contribution of the current study is demonstration that the instructional intervention appears to increase the achievement for diverse learners on high-stakes testing. Specifically, this is the first study to examine the effects of a co-teaching intervention as measured by high-stakes assessments.

Academic Achievement for Typical Students Educated in the Experimental Condition

Although this study demonstrated increased achievement for at-risk students, it did not appear to make a difference for typical students. Because this is the first study to compare the achievement of the typical students in the experimental and control settings for a co-teaching intervention, there is no way to determine if findings were unusual for typical students. Most co-teaching research has not focused on the impact of co-teaching on the rest of the class (Lundeen & Lundeen, 1993; Shulte et. al, 1990; Walsh & Snyder, 1993) and has only been interested in the gains of only students with disabilities.

This study also contributed to the research on co-teaching by examining the academic achievement of subgroups of at-risk students as identified according to NCLB (2001). Results of this investigation demonstrated that the co-teaching intervention was effective for the subgroups of exceptional students, students with 504 plans, and students who received free or reduced lunch. However, results did not show any differences as determined by co-teaching effectiveness for the African American subgroup.

The lack of impact on the African American student achievement in both settings could have occurred because the intervention was not planned specifically to be culturally responsive to this ethnic group. Since this was the first study to investigate the effectiveness of this

intervention on specific racial and ethnic groups, it is now evident that future research is needed to evaluate the method of this intervention with these specific subgroups.

An important consideration should be attributed to the fact that subgroups did overlap in this research. For example, exceptional students, students with 504 plans, and students who received free or reduced lunch could have also been dually represented as African American during the statistical analysis. Therefore, the findings could have been an artifact of this condition. In the future, it will be necessary to perform individual analyses in regards to the potential overlapping of characteristics that would place students in more than one group.

Student Achievement in Science Settings

Another unique contribution of this study to co-teaching was the focus on science. Science is an academic area that must be measured and factored into accountability considerations at each public school setting (i.e., elementary, middle, high) by the 2007-2008 school year (NCLB, 2001). To date, no other co-teaching intervention has investigated science settings at the secondary level using high stakes assessments to investigate the effects of co-teaching.

Findings of this current investigation support co-teaching as an intervention to increase the science achievement of students who are considered to be at-risk as measured on high stakes tests. One possible explanation why the co-teaching model may have positively influenced science scores for these students included the features that the co-teaching model had in common with the dynamic science instruction components described by Scruggs and Mastropieri (1994c). For instance, components of dynamic science instruction have been identified by the authors to include: (a) administrative support, (b) support from special education personnel, (c) an

accepting, positive classroom atmosphere, (d) appropriate curriculum, (e) effective general teaching skills, (f) peer assistance, and (g) disability-specific teaching skills.

Comparisons can be made between certain components of dynamic science instruction (Scruggs & Mastropier, 1994c) and the intervention in this current investigation. First, support from education personnel was established through the teaming of the general and special education teachers and the shared responsibilities of instruction, evaluation and support. Effective teaching skills were embedded within the pre-existing intervention by using a 3X instructional sequence (i.e., 30 minutes review, 30 minutes instruction, and 30 minutes application) for 80% of the duration of the intervention.

Additionally, the educational backgrounds of the professionals in the experimental setting promoted disability specific instruction, an appropriate curriculum (i.e., modifications and accommodations), and daily support from the special education department. Given that the intervention in this research was aligned with some of the dynamic components found to be effective in science settings (Scruggs & Mastropieri, 1994c), it is feasible that academic improvements for at-risk students are a result of effectively implementing the co-teaching intervention.

Synthesis of Contributions

As a result of the co-teaching intervention, several important insights were gained. First, these results provide data indicating increases in the achievement of at-risk students on high stakes testing. Second, the focus on implementing the intervention in a secondary science classroom may provide an important instructional strategy that can be used in those content areas

to improve students' achievement. Third, results of this investigation indicate a need to further evaluate how co-teaching impacts the academic achievement of all students.

Limitations

Although this study makes important contributions to the literature on co-teaching, it also has several limitations. One limitation to this research is that a non-equivalent, quasi-experimental design was employed in this study. Also, the participants were not randomly assigned to treatment. As discussed earlier, there is no way to control for selection bias in a quasi-experimental design. This study illustrates why this can be problematic and interviews with teachers indicate that selection bias may have been present. One way that selection bias seemed evident is that the typical students had numerous documented disciplinary referrals and time out notices. As a result, the grades for these students were lower than the atypical students throughout the semester.

Another limitation to this research is related to the framework of the co-teaching support model. Most teaching interventions are typically comprised of multiple components. Therefore, it is necessary to evaluate these interventions in the same capacity that this current investigation sought to do. However, it is impossible to determine if a combination of components, one distinct component, or the entire intervention package was responsible for the increased academic achievement of subgroups and at-risk students in the experimental setting.

Additionally, since there was no procedural measure of what was done in the control setting, it is impossible to understand if some of the components of the treatment package used in this investigation were also evident in the control setting, thus increasing the achievement scores

of typical students in that setting. Questions still remain as to what occurred in the control setting.

Future Research

There are several recommendations for future research. First, collecting additional data such as unit test scores, classroom assignments, and project scores within each setting throughout the intervention would have allowed the researcher to determine consistency of student performance and allowed for a predictive component concerning the EOC/B final examination.

Measures should be taken in future research to determine comparability between and within groups participating in research. For instance, this study used pre-existing EOG scores in reading and math to establish a criterion for subgroup membership to determine if the selected control group classrooms were comparable to the existing classrooms in the experimental condition. Although comparability was established between groups in this investigation, future research should seek to determine differences and similarities of groups of students within settings as well.

Additionally, there is a need to isolate and evaluate critical variables of the co-teaching package. The current investigations' major focus was on the effectiveness of a co-teaching support model on the biology achievement of students educated in inclusion settings. Typically, instructional interventions are composed of treatment packages. In the future, researchers should identify co-teaching interventions and conduct comparative research on the packages with varying components.

Future research should replicate the co-teaching procedure with more teachers so that the effects of teacher differences can be examined. To date, there are very little data available on the effects of co-teaching and no data on the effects of differences between instructors using the co-teaching intervention.

Finally, there is a need to compare co-teaching interventions across more ethnic groups and between ethnic groups. Since the co-teaching intervention is a popular instructional strategy in inclusive settings, it is crucial to determine the effectiveness of such intervention with different populations that are typically represented in inclusive settings.

Conclusion

In conclusion, the current investigation describes a co-teaching support model that increased academic achievement in biology for at-risk students in biology as indicated by scores on a high stakes test. As a result of the NCLB (2001) evidence-based practices must be defined and incorporated in public school settings in this era of accountability. Conducting research using an instructional strategy as measured by a high stakes test in this investigation has provided initial data to take the first step in analyzing whether co-teaching should be considered a viable option in inclusion settings. Since the numbers of students with disabilities are increasingly included in general education classrooms for instruction, there is a continuing need to address what works not only for students with disabilities, but also what works for all students educated in these environments.

This research has begun to address this concern and has yielded primary data supporting the academic growth of diverse students educated in inclusion settings using the co-teaching

model for instructional delivery. Although the findings were positive for at-risk subgroups, with the exception of the African American subgroup, there is a continuing need to conduct research in the avenue of co-teaching to further understand contributions and limitations of this approach as it pertains to students and their achievement on mandated high stakes tests.

REFERENCES

Affleck, J. Q., Madge, S., Adams, A., & Lowenbraun, S. (1988). Integrated classroom versus resource model: Academic viability and effectiveness. *Exceptional Children, 54, (4),* 339-348.

Banerji, M., & Dailey, R. A. (1995). A student of the effects of an inclusion model on students with specific learning disabilities. *Journal of Learning Disabilities, 28*, 511-522.

Bauwens, J., Hourcade, J., & Friend, M. (1989). Cooperative teaching: A model for general and special education integration. *Remedial and Special Education, 10,* 17-22.

Bay, M., Staver, J. R., Bryan, T., & Hale, J. B. (1992). Science instruction for the mildly handicapped: Direct instruction versus discovery teaching. *Journal of Research in Science Teaching, 29(6),* 555-570.

Bear, G., & Proctor, W. (1990). Impact of a full time integrated program on the achievement of non-handicapped and mildly disabled children. *Exceptionality, 1*, 227-238.

Bielinski, J., Thurlow, M., Callendar, S., & Bolt, S. (2001). *On the road to accountability: Reporting outcomes for students with disabilities* (Technical Report 32). Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.   Beilinski, J., & Ysseldyke, J. E. (2000). *Interpreting trends in the performance of special education students* (Technical Report 27). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [June 13, 2003] from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Technical27.htm.

Bolt, S., Krentz, J., & Thurlow, M. (2002). *Are we there yet? Accountability for the performance of students with disabilities* (Technical Report 33). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [June 13, 2003], from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs?Technical33.htm.

Cook, D., & Campbell, T. (1979). *Quasiexperimentation: Design & Analysis Issues in Field Settings.* Boston, MA: Houghton, Mifflin Company.

Cook, L., & Friend, M. (1995). Co-teaching: Guidelines for creating effective practices. *Focus on Exceptional Children, 28(3),* 1-16.

Dalton, B.,& Morocco, C. C. (1997). Supported inquiry science: Teaching for conceptual change in urban and suburban science classrooms. *Journal of Learning Disabilities, 30(6),* 670-685.

Deshler, D. & Schumaker, J. (1993). Strategy mastery by at-risk students: Not a simple matter. *The Elementary School Journal, 94 (2*), 153-167.

Dieker, L. A., & Murawski, W. W. (2003). Co-teaching and the secondary level: unique issues, current trends, and suggestions for improvement. *High School Journal, 86(4),* 1-14.

Erickson, R., Ysseldyke, J., & Thurlow, M. (Winter, 1997). Neglected numerators, drifting denominators, and fractured fractions: Determining participation rates for students with disabilities. *Diagnostique, 23* (2), pp. 105-115.

Hadley, R., & Mitchell, L. (1995). *Counseling research and program evaluation.* Pacific Grove, CA: Brooks/Cole.

Hassel, B. C., & Wolf, P. J. (2001). *Effectiveness and accountability (Part 2): Alternatives to the compliance model*. In: C. E. Finn, A. J. Rotherham, and C. R. Hokanson, Jr., (Eds.).Rethinking Special Education for a New Century. Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Hitchcock, C., Meyer, A., Rose, D., & Jackson, R. (2002). Providing new access to the general curriculum: Universal design for learning. Teaching Exceptional Children, 35 (2), 8-17.

Jenkins, J. R., Jewell, M., Leicester, N., O'Connor, R .E., Jenkins, L. M., & Troutner, N. M. (1994). Accommodations for individual differences without classroom ability groups: An experiment in school restructuring. *Exceptional Children, 60(4),* 344-358.

Kameenui, E. J. & Carnine, D. W. (1998). *Effective teaching strategies that accommodate diverse learners.* Upper Saddle River, NJ: Merrill.

Klingner, J. K., Vaughn, S., Hughes, M. T., Schumm, J.S., & Elbaum, B. (1998). Outcomes for students with and without learning disabilities in inclusive classrooms. *Learning Disabilities Research & Practice, 13(3),* 153-161.

Langenfeld, K. L., Thurlow, M. L., Scott, D. L. (1997). *High-stakes testing for students: Unanswered questions and implications for students with disabilities* (Synthesis Report 26). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Lipsky, D.K., & Gartner, A. (1997). *Inclusion and school reform. Transforming America's classrooms.* Baltimore: Prentice Hall.

Lundeen, C., & Lundeen, D. J. (1993). *Effectiveness of mainstreaming with collaborative teaching.* Paper presented at the annual convention of the American Speech-Language-Hearing Association, Anaheim, CA (ERIC Document Reproduction Service No. ED 368 127).

Lyon, G. R., Fletcher, J. M., Shaywitz, S.E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Schulte, A., & Olson, R. (2001). *Rethinking learning disabilities.* In: C. E. Finn, A. J. Upper Saddle River, NJ: Merrill.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1993). Success for all: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30(1),* 123-148.

Marshall, K. J. & Herrmann, B. A. (1990). A collaborative metacognitive training model for special education and regular education teachers. *Teacher Education and Special Education, 13(2),* 96-104.

Mastropieri, M. A., Scruggs, T. E., Mantzicopoulos, P., Sturgeon, A., Goodwin, L., & Chung, Hsiang (1998). "A place where living things affect and depend on each other:" Qualitative and quantitative outcomes associated with inclusive science teaching. *Science Education, 82,* 163-179.

Mertler, C. A., & Vannatta, R. A. (2002). *Advanced and multivariate statistical methods (2nd ed.).* Los Angeles, CA: Pyrczak Publishing.

Meyers, J., Gelzheiser, L. M., & Yelich, G. (1991). Do pull-in programs forster teacher collaboration? *Remedial and Special Education, 12(2),* 7-15.

Murawski, W. & Swanson, L. (2001). A meta-analysis of co-teaching: Where are the data? *Remedial and Special Education*, *22*(5), 258-267.

National Center for Educational Statistics (2001). Retrieved January 1, 2004 from http://nces.ed.gov/.

New Horizons for Learning (1995). Retrieved December, 27, 2003 from http://www.newhorizons.org.

NGA Center for Best Practices (2003). Retrieved November 24, 2003 from http://www.nga.org.

Norman, K., Caseau, D., & Stefanich, G. P. (1998). Teaching students with disabilities in inclusive science classrooms: Survey results. *Science Education, 82*, 127-146.

Partriarca, L. A., & Lamb, M. A. (1994). Collaboration, curriculum development and reflection as frameworks for exploring the integration of general and special education. *B.C. Journal of Special Education, 18(1),* 95-100.

Pomplun, M. (1996). Cooperative groups: Alternative assessment for students with disabilities? *Journal of Special Education, 30(1),* 1-18.

Pomplun, M. (1997). When student with disabilities participate in cooperative groups. *Exceptional Children, 64(1),* 49-58.

Quenemoen, R. F., Lehr, C. A., Thurlow, M. L., & Massanari, C. B. (2001). *Students with disabilities in standards-based assessment and accountability systems: Emerging issues, strategies, and recommendations* (Synthesis Report 37). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [June 13, 2003], from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis37.html.

Rice, D., & Zigmond, N. (2000). Co-teaching in secondary schools: Teacher reports of developments in Australian and American classrooms. *Learning Disabilities Research and Practice, 15*, 190-197.

Sanford, E. E. (1996). *North Carolina End-of-Course Tests: Algebra 1, Biology, Economic, Legal, & Political Systems, English 1, and U.S. History* (Technical Report #1). Raleigh, NC: Department of Public Instruction, Division of Accountability/Testing.

Sailor, W. (2002). *Devolution, school/ community/ family partnerships, and inclusive education*. In W. Sailor (Ed.), Inclusive Education and School/ Community Partnerships. New York: Teachers College Press.

Sailor, W., Anderson, J., Halvorsen, A., Goering, K. F., Filler, J., & Goetz, L. (1989). *The comprehensive local school: Regular education for all students with disabilities*. Baltimore, MD: Paul H. Brooks.

Sailor, W., Gee, K., & Karasoff, P. (1993). *Full inclusion and school restructuring. Instruction of students with severe disabilities (4th ed.).* M. E. Snell. New York, Charles Merrill Publishing Company.

Scruggs, T. E., & Mastropieri, M. A. (1994). Successful mainstreaming in elementary science classes: A qualitative study of three reputational cases. *American Educational Research Journal, 31(4),* 785-811.

Shulte, A. C., Osborne, S. S., & McKinney, J. D. (1990). Academic outcomes for student with learning disabilities in consultation and resource programs. *Exceptional Children, 57(2),* 162-172.

Shulte, A. C., Villwock, D. N., Whichard, S. M., & Stallings, C. F. (2001). High stakes testing and

    expected progress standards for students with learning disabilities: A five-year study of one

    district. *School Psychology Review, 30(4),* 487-507.

Slavin, R. E., Madden, N. A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for all: First-

    year outcomes of a comprehensive plan for reforming urban education. *American Educational*

    Research Journal, 27(2), 255-278.

The Renaissance Group (1999). Retrieved November 6, 2003 from http://www.uni.edu/coe/inclusion/.

Thurlow, M. L., Langenfeld, K. L. H., Nelson, J. R., Shin, H., & Coleman, J. E. (1998). *State*

    accountability reports: What are states saying about students with disabilities? (Tech. Rep. No.

    20). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Nelson, J. R., Teelucksingh, E., & Ysseldyke, J. E. (2000). *Where's Waldo? A third*

    search for students with disabilities in state accountability reports (Technical Report No. 25).

    Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

    Retrieved [January 23, 2004] from the World Wide Web:

    http://education.umn.edu/NCEO/OnlinePubs/TechReport25.html.

Thurlow, M., Quenemoen, R., Thompson, S., & Lehr, C. (2001). *Principles and characteristics of*

    inclusive assessment and accountability systems (Synthesis Report 40). Minneapolis, MN:

    University of Minnesota, National Center on Educational Outcomes. Retrieved [June 13, 2003],

    from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis40.html.

Thurlow, M., Wiley, H. I., & Bielinski, J. (2002). *Biennial performance reports: 2000-2001 State assessment data.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes, Retrieved [June 13, 2003], from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/.

Thurlow, M., Wiley, H. I., & Beilinski, J. (2003). *Going public: What 2000-2001 reports tell us about the performance of students with disabilities* (Technical Report 35). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [June13, 2003] from the World Wide Web: http://educaion.umn.edu/NCEO/OnlinePubs/Technical35.htm.

Thurlow, M. L., Ysseldyke, J. E., & Anderson C. L. (1995). *High school graduation requirements: What's happening for students with disabilities*? (Synthesis Report 20) Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Turnbull, R., Turnbull, A., Shank, M., Smith, S., & Leal, D. (1998). *Exceptional lives: Special education in today's schools.* Columbus, OH: Merrill.

Trimble, S. (1998). *Performance trends and use of accommodations on a statewide assessment: Students with disabilities in the KIRIS on-demand assessment from 1992-93 through 1995-96* (Maryland-Kentucky Report 3). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [December 26, 2003] from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/MDKY_3.html.

U.S. Department of Education, *Goals 2000: Educate America Act*. Retrieved June 15, 2003 from http://www.ed.gov/legislation/GOALS2000/TheAct/.

U.S. Department of Education, National Center for Educational Statistics, *The Condition of Education 2002*, NCES 2002-025, Washington, DC: U.S. Government Printing Office, 2002.

U.S. Department of Education, *No Child Left Behind 2001*. Retrieved June 15, 2003, from http://www.nochildleftbehind.gov/.

Vaughn, S., Bos, C. S., & Schumm, J. S. (2000). *Teaching exceptional, diverse, and at-risk students in the general education classroom (2nd ed.).* Boston: Allyn & Bacon.

Vaughn, S., Elbaum, B. E., Schumm, J. S., & Hughes, M. T. (1998). Social outcomes for students with and without learning disabilities in inclusive classrooms. *Journal of Learning Disabilities, 31,* 428-436.

Waldron, N., & McLeskey, J. (1998). The effects of an inclusive school program on students with mild and severe learning disabilities. *Exceptional Children, 64*, 395-406.

Walsh, J. M. & Snyder, D. (1993). *Cooperative teaching: An effective model for all students.* Paper presented at the annual convention of the Council for Exceptional Children, San Antonio, TX. (ERIC Document Reproduction Service No. ED 361 930)

Wang, M. C. & Birch, J. W. (1984). Comparison of a full-time mainstreaming program and a resource room approach. *Exceptional Children, 51(1),* 33-40.

Wang, M. C., Peverly, S., & Randolph, R. (1984) An investigation of the implementation and effects of a full-time mainstreaming program, *Remedial and Special Education, 5(6),* 21-32.

Weiss, M. P., & Brigham, F. J. (2000). Co-teaching and the model of shared

responsibility : What does the research support? In T.E. Scruggs, & M.A. Mastropieri (Eds.),

   Advances in learning and behavior disorders, Volume 14. (pp. 217-245). Stamford, CT:

   JAI Press.

Welch, M. (2000). Descriptive analysis of team teaching in two elementary classrooms: A formative

   experimental approach. *Remedial and Special Education, 21(6),* 366-376.

Ysseldyke, J. E., Thurlow, M. L., Langenfeld, K. L., Nelson, J. R., Teelucksingh, E., & Seyfarth, A.

   (1998). *Educational results for students with disabilities: What do the data tell us?* (Technical

   Report 23) Minneapolis, MN: University of Minnesota, National Center on Educational

   Outcomes.

Will, M. (1986). *Educating students with learning problems: A shared responsibility.* Washington, D.C.:

   U.S. Department of Education.

Zigmond, N. & Baker, J. (1990). Mainstream experiences for learning disabled students

   (Project MELD): Preliminary report. *Exceptional Children, 57(2),* 176-185.

Figure 1.
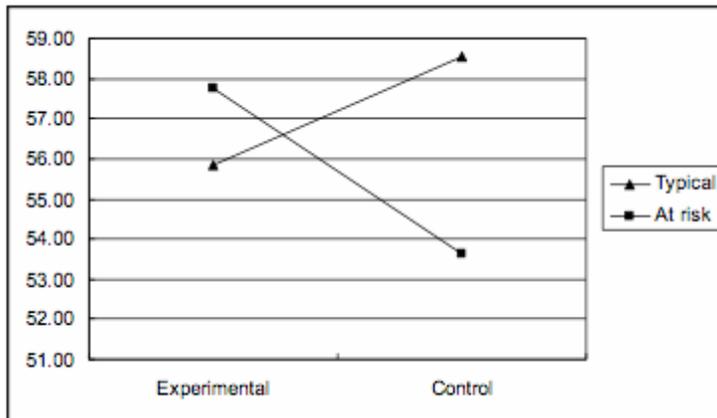
Figure 1. The interaction of setting by type of student on the EOC/B scores.

Table 2
Descriptive Information for the North Carolina Test of Biology

Goal	Description of Goal	Percentage of Items on Test
1	The student will increase his or her knowledge of cellular, physical, and chemical basis of life.	19.5%
2	The student will increase his or her knowledge of the continuity of live and evolution or organisms over	31%

time.

3	The student will increase his or her knowledge of the diversity and unity of life.	18.5%
4	The student will increase his or her knowledge of the ecological relationships among organisms.	19.5%
5	The student will increase his or her knowledge of the behavior patterns of organisms that stem from a combination of heredity and environment.	11.5%

Table 3

Achievement Level Description of the Biology End-of-Course Examination

Level	Description
Level I	Students do not have sufficient knowledge of the skills of the subject to master a more advanced level within the same subject areas.
Level II	Students at this level have an inconsistent knowledge of the skills of the course and are minimally prepared to be successful at an advanced level in the same subject area.
Level III	Students at this level have a consistent knowledge of the skills of the course and are adequately prepared to be successful at an advanced level in the same subject area.,
Level IV	Students at this level have a superior knowledge of the skills of the course and are very prepared to be successful at an advanced level in the same subject area.

(NCDPI, 2008)

Table 4
Standard Error of Measurement for Ranges of Scores on the EOC/B

Score Range	Biology
80-89	3
70-79	2
60-69	2
50-59	2
40-49	4
30-39	6
20-29	5

(NCDPI, 2008)

Table 5
Percent of Students Assigned to Each Achievement Level during Field Testing

Subject Area	Level 1	Level 2	Level 3	Level 4
Algebra 1	14.5%	32.5%	40.4%	12.6
Biology	17.3%	30.9%	36.3%	15.4%
ELPS	13.7%	27.1%	36.0%	23.2%
English 1	13.4%	32.3%	35.4%	18.9%
U.S. History	17.3%	33.6%	33.6%	15.4%

(NCDPI, 2008)

Table 6

Procedural Reliability Scores

Week	Score
1	100
2	66
3	83
4	100
5	100
6	100
7	100
8	83
9	100

Table 7

Number and Percentage of Students by Subgroup Educated in the Experimental and Control Settings

Subgroup	Experimental		Control	
	$\underline{N}$	$\underline{\%}$	$\underline{N}$	$\underline{\%}$
Students with 504 Plans	9	9.3	7	7.2
Students with Disabilities	10	10.3	12	12.4
Students who Receive Free or Reduced Lunch	11	11.3	8	8.2
African American Students	11	11.3	14	14.4
Caucasian Students	51	52.5	34	35.0
Total At Risk Students	30	30.9	28	28.8
Total Typical Students	25	25.8	20	20.6

Table 8

Means and Standard Deviations of the End of Grade Scores in Reading

Setting	N	M	SD
Experimental	55	170.00	13.40
Control	48	175.00	14.00

Table 9

Means and Standard Deviations of the End of Grade Scores in Math

Setting	*N*	*M*	*SD*
Experimental	55	291.50	7.00
Control	48	290.13	10.00

Table 10

Means and Standard Deviations for Typical and At-Risk Students

Group	Student	*N*	*M*	SD
Experimental	Typical	25	55.84	6.87
	At risk	30	57.77	8.28
	Total	55	56.89	7.66
Control	Typical	20	58.55	5.51
	At risk	28	53.64	5.74
	Total	48	55.69	6.10
All Students	Typical	45	57.04	6.38
	At risk	58	55.78	7.40
	Total	103	56.33	6.97

Table 11

Two-way ANOVA Summary Table

Source	SS	df	MS	F	η²
Student	55.90	1	55.90	1.20	.00
Instruction	12.60	1	12.60	.30	.01
Student X Instruction	293.63	1	293.62	6.34 *	.06
Error	4586.10	99	46.32		
Total	331782.00	103			

* p<.05

Table 12

Means and Standard Deviations of the Academic Scores of Students Who Received Free or Reduced Lunch

Setting	N	M	SD
Experimental	11	63.00	7.00
Control	8	53.00	8.00

Table 13

Means and Standard Deviations of the Academic Scores of Students with Disabilities

Setting	N	M	SD
Experimental	10	61.00	8.00
Control	12	52.30	7.01

Table 13

Percent Pass Rate for Students Educated in the Experimental and Control Settings

	Experimental	Control
Type Student	Percent Pass Rate	Percent Pass Rate
At-Risk	67%	50%
Typical	60%	80%

Table 14

Means and Standard Deviations of the Academic Scores of African American Students

Setting	N	M	SD
Experimental	11	55.00	6.04
Control	14	55.00	6.00

Table 15

Means and Standard Deviations of the Academic Scores of Students with 504 Plans

Setting	N	M	SD
Experimental	9	62.60	9.13
Control	7	54.00	4.12

Table 16

Differences between Subgroups in Follow Up Analyses

Subgroup	Significant Differences
Free or Reduced Lunch	Yes
Students with 504 Plans	Yes
Exceptional Students	Yes
African American Students	No

Table 17

Percentage of Students Scoring at Each Level of Achievement

Achievement Level	*Experimental*	*Control*
Level 1	*5.5*	*8.3*
Level 2	*31.0*	*29.2*
Totals for Levels 1 and 2	*36.5*	*37.5*
Level 3	*58.0*	*60.4*
Level 4	*5.5*	*2.1*
Totals for Levels 3 and 4	*63.5*	*62.1*