

Seton Hall University eRepository @ Seton Hall

Seton Hall University Dissertations and Theses
(ETDs)


Seton Hall University Dissertations and Theses

Fall 10-19-2016

The Influence of Grade Level Flexible Grouping on Reading Achievement

Jeanna Czerniecki Velechko
jeanna.velechko@student.shu.edu

Follow this and additional works at: <https://scholarship.shu.edu/dissertations>

 Part of the [Curriculum and Instruction Commons](#), and the [Elementary and Middle and Secondary Education Administration Commons](#)

Recommended Citation

Velechko, Jeanna Czerniecki, "The Influence of Grade Level Flexible Grouping on Reading Achievement" (2016). *Seton Hall University Dissertations and Theses (ETDs)*. 2231.
<https://scholarship.shu.edu/dissertations/2231>

The Influence of Grade Level Flexible Grouping on Reading Achievement

Jeanna Czerniecki Velechko

Dissertation Committee

Carolyn Sattin-Bajaj, Ph.D., Mentor
Christopher Tienken, Ed.D.
Richard Schwester, Ph.D.
Thomas W. Dowd, Ph.D.

In partial fulfillment of
the requirements for the degree of
Doctor of Education

Seton Hall University

2016

© Copyright by Jeanna Czerniecki Velechko

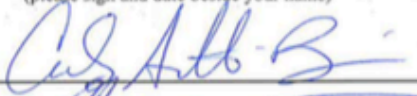
All Rights Reserved


SETON HALL UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN SERVICES
OFFICE OF GRADUATE STUDIES


APPROVAL FOR SUCCESSFUL DEFENSE


Jeanna Velechko, has successfully defended and made the required modifications to the text of the doctoral dissertation for the **Ed.D.** during this **Fall Semester 2016**.

DISSERTATION COMMITTEE
(please sign and date beside your name)

Mentor:
Dr. Carolyn Sattin-Bajaj 

Committee Member:
Dr. Christopher Tienken 

Committee Member:
Dr. Richard Schwester 

Committee Member:
Dr. Thomas Dowd 

The mentor and any other committee members who wish to review revisions will sign and date this document only when revisions have been completed. Please return this form to the Office of Graduate Studies, where it will be placed in the candidate's file and submit a copy with your final dissertation to be bound as page number two.

Abstract

This study examined the influence of grouping formation on the scores on the New Jersey Assessment of Skills and Knowledge (NJASK 3) in Grade 3 English Language Arts. Grouping within the general education classroom was compared to grouping between the same grade level. The analysis included a multiple regression model for student variables gender, race/ethnicity, prior ability and grouping status. All data explored in this study pertained to 155 third graders in one New Jersey suburban district during the 2012-2013 academic school year. The results of the study revealed prior ability in reading influenced the scores on the NJASK 3 reading section when combining gender, race/ethnicity, and grouping status in the model. Grouping status, gender, or race/ethnicity were not significant influences on the NJASK 3 reading scores in the multiple regression model.

Keywords: between grade level grouping, flexible grouping, reading, elementary

Acknowledgments

I completed this work on the shoulders of many scholars, family members, and friends. I would like to start by thanking Dr. Carolyn Sattin-Bajaj, who guided me along this learning experience. Dr. Sattin-Bajaj, your standards of excellence, feedback, and words of encouragement were appreciated. I have learned to examine data, question, and find connections in scholarly journals and books thanks to you. You also realized the role of “mentor” included being a teacher, editor, and most importantly, motivator. Your words of support always seemed to arrive at the perfect time and you understood when circumstances sometimes were pressing. I cannot thank you enough for the pursuit of knowledge you have encouraged me to seek. Dr. Tienken, even while you were on sabbatical in Italy, you found time to reflect on my work and provide suggestions to improve the reliability and validity of my study. Your contributions to the quantitative research design and analysis complemented my team of mentors. Grazi! Dr. Schwester, thank you for always finding the time after a round of golf to answer one of my numerous questions on statistics. Your contributions to my understanding of the data and research design were invaluable. You also kept me grounded during the process. I appreciate your offer to serve on my committee when we first met, and I hope you did not regret this gift of time! Dr. Dowd, I appreciate your reading my work, providing feedback, and offering the practitioner’s perspective to my dissertation committee. I am so grateful I can use your career path and goal setting as a model.

When I began this journey, I did not realize the sacrifices my family would need to make. Scott, you proofed many papers, spent more time with our son, and even arranged weekends away to allow me time to work on this goal. I also know you

overlooked many items on our “to do” list that needed to be postponed until I finished my doctorate. I appreciate your love and understanding, and I am glad we are on this life journey together. Harrison, I know there were times you needed to wait for me to join you to play or complete an activity because I was working on my doctorate. I can only hope you will someday understand that we are perpetual students just trying to satisfy our curiosity. The gift of an education comes with so many rewards, and I hope your education enriches your life as much as mine does for me. For my parents, Eugene and Jean Czerniecki, who made many sacrifices so I could have the gift of education, thank you. I learned the importance of an education every time you read me a bedtime story, supervised projects, attended teacher conferences, looked at colleges, worked BINGO and paid for tutors, school, uniforms, etc.

Finally, thank you to my friends in Cohort XVI for all your support during our two years together. Our collaboration, reassurance, and therapy sessions helped me get through this journey.

Dedication

To all of those who nurtured and supported me throughout my childhood.

Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
Dedication.....	v
List of Tables.....	ix
List of Figures.....	x
I. INTRODUCTION.....	1
Background.....	1
Statement of the Problem.....	12
Purpose of the Study.....	12
Research Questions.....	13
Hypothesis.....	13
Design and Methodology.....	14
Conceptual Framework.....	15
Significance of the Study.....	17
Limitations.....	18
Delimitations.....	19
Variables.....	20
Dependent/Outcome Variables.....	20
Independent/ Treatment Variables.....	20
Control Variables.....	20
Definition of Terms.....	20
Organization of the Study.....	25
II. REVIEW OF THE LITERATURE.....	26
Introduction.....	26
Literature Search Procedures.....	27
Inclusion and Exclusion Criteria for Literature Review.....	28
Focus of the Review.....	28
Limitations of the Review.....	29
Review of Literature Topics.....	29
Structures of Grouping: Within-Class and Between-Class.....	29
Flexible Grouping.....	31
Historical Background.....	33
Intelligence Quotient and its Effect on Grouping.....	33
Influence of Immigration on Grouping.....	34
The Equality Movement.....	36
Accountability Era and Student Grouping.....	39
Teacher Accountability Movement.....	41
Grouping in the Contemporary Classroom.....	44
Prevalence of Grouping.....	44
Justification for Grouping: Range of Abilities.....	45
Justification for Grouping: Improves Achievement.....	46
Justification for Grouping: Management.....	48
Negative Consequences of Grouping: Labeling.....	49

Negative Consequences of Grouping: Segregation.....	50
Negative Consequences of Grouping: Hinders Achievement.....	51
Empirical Studies.....	52
First Lens: Comparing Homogeneous and Heterogeneous Grouping in Totality.....	53
Second Lens: Comparing Homogeneous and Heterogeneous Grouping by Ability Level.....	54
Third Lens: Comparing Within-Class Grouping with Whole Group Instruction.....	59
Fourth Lens: Comparing Flexible and Whole Group Instruction.....	61
Theoretical Framework.....	62
Conclusion.....	64
III. METHODOLOGY.....	66
Introduction.....	66
Research Design.....	66
Research Questions.....	67
Sample Population/Data Source.....	68
Instruments.....	73
Data Collection.....	74
Units of Analysis.....	75
Analysis Construct (Model Specifications).....	75
Analysis of Data.....	76
IV. ANALYSIS OF THE DATA.....	78
Introduction.....	78
Predictor Variables.....	79
Descriptive Statistics.....	80
Research Questions.....	83
Hypothesis.....	83
Results	83
Multiple Regression.....	92
Research Questions and Answers.....	97
Summary.....	99
V. CONCLUSIONS AND RECOMMENDATIONS.....	101
Introduction.....	101
Conclusions Policy Recommendations: Gender.....	105
Conclusions Policy Recommendations: Race/Ethnicity.....	108
Conclusions Policy Recommendations: Prior Ability.....	114
Conclusions Policy Recommendations: Grouping Formation.....	119
Recommendations for Future Research.....	121
Summary.....	124
REFERENCES.....	125

LIST OF TABLES

Table 1. District Third Grade Demographics Groups.....	68
Table 2. Demographics for Control and Treatment Groups.....	72
Table 3. Description of Variables Used in the Study.....	76
Table 4. Variable Names and Descriptors.....	79
Table 5. Descriptive Statistics for Group, Gender and Race/Ethnicity.....	80
Table 6. NJPASS and NJASK 3 Descriptive Statistics.....	81
Table 7. Independent Sample T-test.....	82
Table 8. Grouping Status and NJPASS/NJASK 3 Crosstabulation.....	82
Table 9. Correlation Tables.....	89
Table 10. Gender and Combined Predictor Variables Coefficients.....	91
Table 11. Race/Ethnicity and Combined Predictor Variables Coefficients.....	91
Table 12. NJPASS and Combined Predictor Variables Coefficients.....	91
Table 13. Grouping Status and Combined Predictor Variables Coefficients.....	92
Table 14. Regression Variables Entered/Removed.....	93
Table 15. Model Summary	93
Table 16. ANOVA Table	94
Table 17. Coefficients Table.....	95

LIST OF FIGURES

Figure 1. NJPASS and NJASK 3 Gender Scatter Plot.....	85
Figure 2. NJPASS and NJASK 3 Race/Ethnicity Scatter Plot.....	86
Figure 3.NJPASS and NJASK 3 Grouping Status Scatter Plot	87
Figure 4. NJPASS and NJASK 3 Prior Ability Scatter Plot.....	88

CHAPTER I

INTRODUCTION

Background

Teacher evaluations have been a topic of concern in recent years. As of September 2015, forty-five states and the District of Columbia require teacher evaluations be linked to student achievement (Doherty & Jacobs, 2015). While each of these states' lawmakers may have some variation on how they define "effectiveness," all 41 of the states and District of Columbia determine this effectiveness based on student performance on a state approved assessment tool. There are nine states and the District of Columbia who are using the Partnership of Assessment for Readiness of College and Careers (PARCC) to measure their students' performance (PARCC States, 2016) and 17 states using Smarter Balanced, a consortium of states who use the Smarter Balanced assessment of the Common Core State Standards (CCSS) (Smarter Balanced Members, 2016). In September of 2010, U.S. Secretary of Education Arne Duncan announced PARCC and Smarter Balanced consortiums would receive a total of \$330 million dollars to develop assessments to evaluate student achievement, and many states would use these results to rate teacher effectiveness (U.S. Department of Education, 2014). With so much money and many states involved in two major testing consortiums, the stakes are high for lawmakers, testing companies, teachers, and most importantly, students.

An examination of how teachers are held accountable for student achievement in their final evaluation gives insight into the high stakes placed on student performance on such tests as PARCC and Smarter Balanced. Doherty and Jacobs (2015) reviewed all the states' teacher evaluation systems to report on student learning objectives. They

concluded that student achievement on state assessments had impact on six major areas for teachers: tenure and licensure, professional development planning, termination of employment or job performance improvement plan creation, teacher compensation, layoffs, and teacher preparation program effectiveness. This move to link teacher evaluations with student achievement on state mandated standardized tests and other measures may cause numerous educators to examine the best instructional methods to improve student achievement on state mandated assessments, thus helping a teacher's overall evaluation.

Ability grouping students for academic instruction is one approach that educators have begun to adopt, or, more accurately, return to with increasing frequency. Loveless (2013) speculates that the No Child Left Behind Act (NCLB) of 2002 is a major catalyst for the increase in grouping practices. Even though President Obama signed Every Student Succeeds Act (ESSA) on December 10, 2015, to replace NCLB, ESSA still requires students in Grade 3 through Grade 8 be tested in language arts and math once a year and once during the high school years (Every Child Succeeds, 2016). However, under ESSA the accountability for these tests resides at the state level. ESSA also requires groups of student performance be monitored carefully as is the case with NCLB (Every Child Succeeds, 2016). The true ramifications of ESSA will not be learned until the federal government specifies the regulations concerning ESSA, but standardized state testing will still occur under this act. Educators will continue to explore and examine teaching practices that lead to better performance on these assessments.

Grouping, the practice of dividing students for instruction by some characteristic such as skill development or ability (Gentry, 2014; Slavin, 1987; Tomlinson, 2003) is

becoming increasingly common at the elementary level, especially in reading (Chorzempa & Graham, 2006; Ford & Opitz, 2011; LeTendre, Hofer, & Shimizu, 2003). If Loveless's (2013) theory is correct and teachers are returning to grouping as a means to meet the requirements of NCLB or the requirements of a state's NCLB waiver, it is paramount that educators ensure their use of grouping strategies will indeed help students succeed. In order to measure the effectiveness of grouping, it is necessary to examine the historical perspective, the consequences of grouping, and the research conducted to date in order to make an informed decision about the use of this practice. The height of the research on grouping was conducted during the 1980s and 1990s; but instructional practices and curriculum have changed in recent years under NCLB, NCLB waivers, and now ESSA. In today's modern English language arts classroom, students are ability grouped for various reasons, and the groups are more fluid than in the past (Caldwell & Ford, 2002; Gambrell, Morrow, & Pressley, 2007; Gentry, 2014). For example, teachers may form an ability group to teach a specific skill and then change the composition of the group the next day when reinforcing a different objective. Teachers may also form groups based on reading level, and then move students to another group as improvement in reading level is evident (Fountas & Pinnell, 1996). Many researchers have referred to this type of grouping as "flexible grouping" since the groups are dynamic and change based on need (Ford, 2005; Opitz, 1998). Curriculum has also changed, as 43¹ states have implemented the Common Core State Standards (CCSS) or curriculum standards aligned to the CCSS in some way (Common Core Standards, 2016). The CCSS are a set of standards that students are expected to learn; it is not a curriculum nor do the CCSS mandate how educators teach the standards.

¹ Minnesota adopted the English Language Arts standards only (www.corestandards.org).

Ability grouping, the process of organizing students in certain groups based on specific criteria, was reported as early as the 1800s (Otto, 1932). Since the 1800s the use of ability grouping as an instructional strategy has ebbed and flowed with the political and societal demands of the times. Introduction of Intelligent Quotient (IQ) testing in the early 1900s resulted in schools organizing students based on their IQ score and then the practice of ability grouping continued as a way to deal with the wave of immigrants entering the United States in the 1920s (Ansalone, 2006). The equality movement of the late 1950s until the 1970s challenged grouping as an ethical practice since minorities and those from lower socioeconomic statuses (SES) were overrepresented in the lower level ability or tracked groups (Worthy, 2010). Even with the call to alter or cease ability grouping and tracking, the practice remained. One must delve into the research to discover why a practice challenged in the 1960s and the 1970s is a common practice in the modern elementary classroom. There are researchers such as Loveless (2013) who believe the requirements of NCLB fueled the ability grouping practice.

The federal government law NCLB required each state to establish standardized testing tools it would use to measure the Adequate Yearly Progress (AYP) of its students in meeting or exceeding the state's curriculum standards over the course of 12 years. The goal of NCLB was to have each student meeting or exceeding a state's established standards by 2014. In August 2011 the United States Department of Education (USDOE) announced President Obama would allow states to apply for NCLB waivers in exchange for rigorous state-developed plans that improved educational outcomes for all students regardless of ability, race, or gender and hold teachers and principals accountable for students' performance on these assessments (NCLB Flexibility and Waivers, 2013). As

of March 2015, forty-one states and the District of Columbia were granted NCLB waivers or extensions to a previous waiver (States Granted Waiver 2014). Some of these states proposed in their waiver application to measure AYP by using a Student Growth Percentile (SGP) model, designed by statistician D. W. Betebenner, rather than a model that examines the percentage of students attaining the minimal score established by each state (O'Malley, Murphy, McClarty, Murphy, & McBride, 2011). Even under the new ESSA, states must have multiple measures to assess the state's educational performance; and one of these multiple measures must be scores from a state mandated test (ESSA, 2016).

The new SGP formula used by various states is based on students' past performance on the state assessment. The states will monitor a student's academic progress by comparing peers of similar caliber, allowing the state to monitor individual progress of students no matter how the child performs. It is the hope that a student would improve his/her performance on the state assessments within the average of those with similar scores the year prior (O'Malley et al., 2011). Proponents of this SGP model believe not only does it accurately assess students' growth, but teacher effectiveness as well. As of September 2015 there are 45 states and the District of Columbia who incorporate some form of student performance in a teacher's overall effectiveness rating (Doherty & Jacobs, 2015). One state that will serve as an example of such ties between student performance and teacher evaluation is New Jersey. Governor Christopher Christie signed Achieve NJ in 2012, which assigned each teacher a ranking—ineffective, partially effective, effective, or highly effective—and used the teacher's average SGP performance as a factor in determining effectiveness (Achieve NJ, 2012).

One methodology being examined as a technique to improve assessments of all learners is academic ability grouping in reading. Ability grouped students are divided for instruction based on their attainment of reading skills and academic performance. Students' reading ability can be based on a myriad of reading skills. However, according to the publication from the National Reading Panel, *Put Reading First* (Armbruster, Lehr, & Osborn, 2001, 2006) there are six major areas of reading instruction in kindergarten through Grade 3. The panel examined peer-reviewed research articles determining the essential skills for developing readers. The committee determined that phonemic awareness, phonological awareness, phonics instruction, fluency, vocabulary, and text comprehension are the essential areas for reading instruction in the primary grades (Armbruster et al., 2001, 2006).

Assessment of the reading skills can take place through various informal and formal evaluations. Teachers may examine students' writing to determine acquisition of graphemes, phonological awareness, and phonics. Teachers may also listen to students read aloud to determine decoding skills, miscues, and fluency. Teachers will then often ask questions to elicit the student's understanding of the text. While teachers assess reading skills informally through many disciplines, there are also formal assessments, which guide teachers in evaluating some or all of the essential reading skills identified by the National Reading Panel.

Teachers may determine levels and skill attainment by using a formalized, leveling benchmark system such as the *Diagnostic Reading Assessment 2*, (DRA2) published by Pearson or the *Fountas and Pinnell Benchmark Assessment*, published by Heinemann. In order to assess a student's skill level, the student reads a leveled text,

provided by the publisher, while the teacher records reading miscues and the student's responses to comprehension questions (Pearson, 2011). The student's responses are recorded and assessed on a rubric that examines some specific skills identified by the National Reading Panel (2001). For example, when using the DRA2 the teacher records the number of words read per minute and documents reading miscues or reading errors made by the student. Reading fluency, a skill identified in the publication *Put Reading First*, is impacted by a student's phonological awareness and knowledge of phonics (Pearson, 2011).

When the formal assessment is complete, the teacher uses a provided rubric to determine the student's independent and/or instructional reading level (Betts, 1946). The teacher tabulates scores earned in reading fluency, comprehension and reading engagement to generate the reading attainment (emerging, developing, independent, or advanced) for the level of text read (Pearson, n.d.). Teachers should find the independent reading level, the text a student can read without teacher support, and the instructional reading level, the text a student can read provided a teacher supports the reader in the process (Fountas & Pinnell, 1999).

In addition to the DRA2 or other diagnostic reading leveling systems, some districts measure reading skills based on performance on norm- and criterion-referenced tests. Some schools may administer the TerraNova, Iowa Test of Basic Skills, New Jersey Performance Assessment of State Standards (NJPASS), or individual state assessments like the New Jersey Assessment of Skills and Knowledge (NJASK). Two such tests that will serve as example, and are discussed in this study, are the NJPASS and the NJASK 3. The NJPASS, published by Houghton Mifflin Harcourt Publishing, measures second

grade reading skills under two main categories, Working With the Passage and Analyzing the Passage. Houghton Mifflin Harcourt Publishing notes the following skills are measured:

Working with the Passage includes comprehending the passage, such as recognizing the main idea and supporting details, getting information, paraphrasing meaning, and recognizing organization of text and purpose of reading. This comprises sixteen (61.5%) of the twenty-six points in reading.

Analyzing the Passage includes analyzing and evaluating the passage by using skills such as asking questions, predicting meaning, developing opinions, drawing conclusions, and interpreting conventions of print. This comprises ten (38.5%) of the twenty-six points in reading. (New Jersey Proficiency Assessment of State Standards, 2015).

Out of the six major areas of essential reading instruction discussed in *Put Reading First*, only one area, comprehension, is directly assessed. One cannot tell from the scores earned on the Working with Text or Analyzing the Passage sections, if the student was able to decode the text or if the student understood the meaning of specific words, as scores are not provided for vocabulary, fluency or phonemic awareness.

The NJASK 3 is a criterion-referenced test created by Measurement Incorporated for the state of New Jersey. Test specifications from NJASK 3 from spring of 2013 identified the skills students needed to be Proficient in reading. The test specifications state the following:

A student performing at the Proficient level demonstrates the ability to employ strategies to comprehend a variety of texts literally and inferentially and to express understanding of the text in written responses. As a proficient reader, the student recognizes the central idea, supporting details, purpose, and organization of the text as well as some literary devices. The proficient student can make connections to the text, form opinions, and draw conclusions. The proficient reader is able to synthesize ideas from the reading and to use these to analyze and extend the meaning of the text in written responses. (NJDOE, 2013)

The NJDOE NJASK 2013 Test Specification Manual provided percentages of points earned in each skill.

The NJASK 3 Reading section accounted for 60 percent of English Language Arts score with 40 percent of English Language Arts score dedicated to informational text items and 20 percent testing literature items. The literature section contained six multiple choice items and one open-ended item while the informational section comprised 12 multiple choice items and two open-ended items. (NJDOE, 2013)

As per the performance level descriptors established by the NJDOE, the NJASK 3 is heavily focused on comprehension. In order to be a proficient reader, a student must comprehend the central idea, supporting details, and purpose (NJDOE, 2014). Like the NJPASS, the NJASK 3 does not provide feedback on specific reading skills such as fluency and phonological awareness.

The informal and the three formal assessments discussed, DRA2, NJPASS, and NJASK 3, measure and assess reading skills in various ways. Formal and informal

assessments may then be used to form ability groups. It is then up to the teacher, school administration, and/or district administration to determine how these measurements are nuanced to form reading groups.

Ability grouping takes on numerous forms such as the following: tracking, within-class ability grouping, between-class ability grouping, guided reading, Joplin Plan, mixed-age classrooms, pull-out programs, and flexible ability grouping. Some researchers (Duflo, Dupas, & Kremer, 2009; Jecks, 2011; Lou, Abrami, Spence, Poulsen, Chambers, & d'Appollonia, 1996; Puzio & Colby, 2010) have found a slight to moderate positive correlation between ability grouping and achievement, while other researchers have found ability-based groups to perform not significantly different than students in heterogeneous classrooms when taking assessments (Burton, 2005; Condrón, 2005; Macqueen, 2008; Mosteller, Light, & Sachs, 1996; Matthew, Ritchotte, & McBee, 2013; Nomi, 2006; Slavin, 1987).

These contradictory results have fueled a debate on grouping (Gamoran, 2009). The results of key meta-analytical studies have found that at-risk learners do not perform as well in ability groups when compared to similar peers placed in the heterogeneous classroom (Condrón, 2005). Other researchers have determined that higher ability students perform better in a homogeneous group rather than in the heterogeneous classroom (Condrón, 2005; Kulik & Kulik, 1992). In addition, some researchers have examined the academic results in totality and have found no overall difference between grouped and ungrouped students (Kulik & Kulik, 1982; Macqueen, 2008; Nomi, 2005; Slavin, 1987), while others researchers found empirical data to support grouping (Collins & Gan, 2013; Jecks, 2011; Puzio & Colby, 2010). The research on grouping has also

furthered the debate among scholars on the ethical considerations of the practice. Advocates such as Worthy (2010) and Oakes (1985) have questioned the use of ability grouping since the lowest ability groups are often over-represented by minorities (Macqueen, 2008; Oakes, 2008). It is noted that today's ability grouping looks different than it did 20 to 30 years ago when the grouping research was at its pinnacle (Loveless, 2009; Gamoran, 2009).

Current trends in reading instruction led by experts such as Fountas and Pinnell (1996, 2010) have found benefits of guided reading groups, a form of flexible ability grouping. Guided reading is a form of ability grouping that brings a small group of readers together, typically within the heterogeneous classroom, to work on a specific skill with support from the teacher (Ford & Opitz, 2011; Fountas & Pinnell, 2010). Some teachers examine running record rubrics, such as the DRA2 or the *Fountas and Pinnell Benchmark Assessment*, to determine which of the skills identified on the rubric need remediation and what level text a student is reading independently or with teacher support. Teachers then use this information and informal assessments to form reading groups. Reading groups can meet with a teacher as many times as the teacher determines and can be switched based on the student's need. Most teachers who use guided reading as a component of their reading instruction form their ability-based reading groups with only the students in their classroom. However, heterogeneous classrooms can have many reading levels in a classroom. Firmender, Reis, and Sweeney (2013) examined over 1,000 students in Grade 3 through Grade 5 and found third grade classrooms have approximately a nine-year span in reading levels in Grade 3 and approximately an eleven-year reading span in Grades 4 and 5. This wide variety leads to more grouping for

instruction, but this then can decrease frequency of each group meeting with the teacher for guided reading. Some schools have decided to use between-grade-level guided reading to minimize the ranges in a classroom and allow teachers to spend more time focusing in on the learner's needs (Haghighat, 2009).

Statement of the Problem

Many schools throughout the nation have implemented guided reading with primary students as a way to differentiate instruction for the various abilities within the classroom (Firmender et al., 2013; Fountas & Pinnel, 2010). However, it has not been determined if the use of grouping students between grade level or within class for reading has a stronger influence on student reading achievement. Most research to date has either focused on the secondary level, permanent tracking, within-class grouping, or has provided little description of the grouping composition and instructional strategies used during the research study. Most importantly, the evidence on the influence of ability grouping has been inconclusive.

Purpose of the Study

The purpose for this correlational, explanatory, cross-sectional study was to explain the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey. This study explained the amount of variance in the NJASK 3 reading rates accounted for by grouping type, student gender, race/ethnicity, and prior reading ability

Research Questions

The objective of this study was to explain the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey. The overarching research questions answered were the following: What is the influence of grouping formation on the NJASK 3 reading scores? What are the differences in reading achievement between third graders who are grouped between the grade level and those who group for reading within the classroom?

1. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender?
2. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender and race/ethnicity?
3. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender, race/ethnicity, and prior reading ability?

Hypothesis

Null Hypothesis 1: There is no statistically significant difference in the reading achievement of third graders grouped by DRA2 levels using between-grade guided reading groups and similar third graders participating in within-classroom guided reading groups.

Design and Methodology

The purpose of this correlational, explanatory, cross-sectional study explained the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey, which had similar students in two different elementary schools, while controlling for gender, prior reading achievement, and racial/ethnic background, using a simultaneous regression analysis.

Multiple regression analysis was an appropriate research design for this study. The purpose of this study was to explain the influence of the predictor variables on the complex dependent variable and to identify the degree to which this relationship existed, and multiple regression is a tool that calculates this influence (Gay et al., 2011). This multiple regression analysis allowed the researcher to posit the influence of variables on the major composite variable.

The sample for this study consisted of 155 students from one suburban elementary school in the northeast. When these students were in second grade, they used the same reading materials and same reading grouping formation. However, in June of 2012 the district reading committee selected a new reading program for the following school year. *Good Habits, Great Readers* was being used for the first time during the 2012-2013 school year; therefore, all third grade teachers in the district received the same two-day training from the publisher of the reading program. The control school implemented the reading program within the homeroom reading class and the treatment school implemented the program amongst the grade level. Thus, all third graders within the

district were exposed to one of two different types of grouping in their third grade reading class. Guided reading was used to implement the *Good Habits, Great Readers* program at both schools. Teachers established guided reading groups based on the DRA2 results and observed reading characteristics. While the teacher worked with a guided reading group using the leveled readers from the reading program, other students in the class engaged in center or independent work. In the spring of 2013, all subjects took the NJASK 3.

The criteria for inclusion in the sample for this study were as follows: (a) assessed on the New Jersey Performance Assessment of State Standards (NJPASS) during the 2011-2012 school year, (b) assessed on the NJASK 3 during the 2012-2013 school year (c) assessed on the DRA2 in the 2012-2013 school year, and (d) never eligible for or received special education services²

Conceptual Framework

Lev Vygotsky (1978) introduced a theory about how children learn. Vygotsky believed children had a developmental age and this developmental age is based on what a child can do independently. However, Vygotsky believed children could extend their knowledge and thinking skills with adult or peer support. Vygotsky called the difference between what a child can do independently and what can be completed with support from peers or an adult as the *zone of proximal development*. Vygotsky believed children could have the same developmental level, but could have different levels of what they can understand or do with adult assistance. Vygotsky encouraged teachers to find not only what the child has mastered (developmental age) but also the zone of what the student

² Students receiving special education services in speech only were included.

can do with support. This allows teachers to examine where the child has been and how the child is maturing.

The zone of proximal development has profound implications for the classroom. This theory should drive teachers and administrators to design programs and lessons to extend the learning to continue the developmental process. Vygotsky believed students who focus on what has been mastered are delayed in reaching the next level in the developmental process. However, children who can learn new information through support are continuing the learning process. Because students have variability in prior knowledge and maturity, students would need different teacher support to work within their zone of proximal development. Vygotsky emphasized the zone of proximal development may be different for various subjects/topics.

This Vygotskian framework was the foundation for this examination of flexible grouping. Analysis from Firmender et al. (2013) found third grade classrooms have approximately a nine-year span in reading levels and approximately an eleven year reading span in Grades 4 and 5. If one applies the zone of proximal development theory, differentiated lessons based on a student's developmental age and their zone of proximal development need to be designed. One reading lesson delivered to all classroom learners may stagnate the learning process of some students. Grouping is a strategy that allows the teacher to design a lesson for the various levels within the zone of proximal development that occur in the classroom. Fountas and Pinnell (1996), leaders in the guided reading movement, suggested teachers group students by ability level to teach the necessary skills to move ahead in the reading process. Fountas and Pinnell (1996, 1999) suggested teachers use books that are at a student's instructional level, the level at which a student

can read with the support from a peer or teacher. Teaching students at their instructional reading level is grounded in the theory of the zone of proximal development.

Significance of the Study

When reviewing the studies on grouping, three overall gaps in the literature emerge: grouping in the modern classroom looks different than it did during the height of the grouping research during the 1980s and 1990s; there is a lack of rich descriptions of the grouping practices used to deliver instruction; and most interestingly, (Gentry, 2016) the data collected are inconclusive on the benefits of grouping (Gentry, 2016). Tieso (2003) eloquently expressed, “The time has come to revisit an old friend or foe... ability grouping.”

Many grouping studies conducted to date have focused on data from secondary schools (Lofton, 2013) and/or have examined grouping practices that were fixed for the school year. Grouping practices have changed in the modern classroom (Gamoran, 2009; Loveless, 2009). Current trends in elementary grouping tend to be limited to one or two subjects, typically reading and/or math (Nagel, 2001). These groups are fluid and students change group placements based on the needs of the learner (Caldwell & Ford, 2002). In addition, many studies were conducted during the 1980s and 1990s, before introduction of the Common Core State Standards (CCSS). Reading practices and materials have changed to reflect the CCSS and incorporate guided reading using leveled readers that are presented in a continuum to increase skill application, word count, vocabulary, and comprehension development as a student’s reading level improves.

Literature from past and modern times is inconclusive on the effects of grouping on student achievement (Condrón, 2005; Macqueen, 2010). Researchers Nomi (2006)

and Slavin (1987) have postulated that the inconclusive data may be caused by a lack of description on the grouping practices. For example, certain studies did not explain if the grouping was permanent, fluid, or if curriculum was differentiated for each grouping level (Gamoran, 2009). The grouping studies have also not discussed the specific materials used in the study.

Limitations

This study was limited by the population's demographics. The 155 participants were only representative of a suburban population in one state in the northeast. The participants were comprised of 34% minorities. Because there were not a significant number of participants in each minority category (Asian, Hispanic, Black, Pacific Islander, Native American and Multiracial) students were combined into one category (racial/ethnic); this study cannot be applied to one specific minority. The study was also limited by the lack of diversity in socioeconomic levels (SES). There were four students in the study who qualified for free lunch and one student who qualified for reduced lunch. Because the total number of free/reduced-price lunch students was insignificant (3% of the sample), the study cannot be applicable to students of low SES. The study was also limited by the fact that it was not repeated for another year with a different population to determine if the results were valid and reliable across multiple student populations. Since students with a mobility factor were excluded, findings may not be applicable to students who have frequent mobility. Because the special education students were provided a different setting and materials in the control school, this variable could not be assessed in this study.

In addition, the NJPASS result was used to determine the student's prior reading ability. This single assessment may not be reflective of the student's true reading ability, is not diagnostic, nor is it correlated to the NJASK 3 assessment. The NJPASS was designed to assess objectives on the NJCCCS, but the NJASK 3 in 2013 assessed the CCSS. Unfortunately, the NJASK 3 testing specifics or blueprints do not state the specific reading skills associated with questions or the percentage of test items associated with specific reading skills (*NJASK 2013 Score Interpretation Manual Grades 3-8, 2013*). Without specific test item description, the NJPASS and 2013 NJASK 3 cannot be thoroughly compared.

Furthermore, mediating variables such as the quality of teaching, administrative support, school culture, group assignment factors, and in-house professional development were not examined and may have influenced the results. Because regression and correlation were used to determine influence, the study cannot be used to determine causation. Two variables with a high correlation do not suggest that one caused the other, but this can be used to determine a possible prediction of outcomes (Gay et al., 2011).

Delimitations

The data used in this study were retrieved from a public, suburban district in New Jersey that used flexible reading grouping between students in the same grade level in one elementary school in Grade 3 and another school in the same district that used within-class grouping for reading. All data pertained to the 2011-2012 and 2012-2013 school years. The sample was composed of 78 students in the flexible grouping treatment school and 77 students in the control group. There were 73 females and 82 males. The racial demographics consisted of 102 White students and 53 minorities. Minorities

included Asians, Black or African Americans, Hispanics or Latinos, Pacific Islanders, Native Americans, or Multiracial participants. There were five participants on free/reduced-price lunch and 150 participants who did not qualify or apply for free/reduced-price lunch status. Special education students³ were not included in the study, as they did not use the same curriculum or materials between the control and treatment group. Students were divided into three ability groups based on their scores on the NJPASS assessment results to provide a general understanding of prior ability. There were 19 students who scored within the Partially Proficient range, 58 students who scored within the Proficient range, and 78 students who scored in the Advanced Proficient range.

Variables

Dependent/Outcome Variable

The dependent variable examined was the NJASK 3 reading scores.

Independent/Treatment Variables

There was one dichotomous independent variable: grouping type (between- or within-class).

Control Variables

There were three control variables: gender, race/ethnicity, and prior reading achievement as measured by second grade NJPASS reading scores.

Definition of Terms

Ability Groups – Groups of students divided into levels based on their knowledge and performance on given assessment(s), class work, teacher judgment or a student's general ability. Students are assigned a group and then stay with that group for a duration of time such as a school year, semester, or class assignment.

³ Students who were classified under special education codes as “Speech Only” were included in the study.

Between-Grade Ability Groups – Groups of students in the same grade divided into levels based on their knowledge and performance on given assessment(s), class work, teacher judgment or student’s general ability. Groups may have students from various homerooms in the same grade. There may or may not be movement between the groups.

Developmental Reading Assessment – The Developmental Reading Assessment (DRA) is a reading leveling benchmark assessment for students in kindergarten through Grade 8 developed by Josetta Beaver and Dr. Mark Carter and was purchased by Pearson Corporation in 1997. The program is used to identify each student’s reading achievement through systematic observation, recording, and evaluation of performance. These data help educators determine patterns in student reading abilities, document progress, and communicate assessment information to administrators, parents, and students. The program was revised in 2005 and called *Developmental Reading Assessment 2 (DRA2)* and, according to the Pearson Corporation, is used in over 250,000 classrooms in the United States (Pearson, 2011).

Differentiation –The instructional strategy that provides alternative content, learning activities, and assessments to accommodate the diverse needs of the learners. The goal of this technique is to have all students succeed in mastering the objectives by providing diversified materials and/or lesson plans based on students’ interests and prior performance (Tomlinson, 2003).

District Factor Groups – “The District Factor Group (DFG) is an indicator of the socioeconomic status of citizens in each district and has been used to compare the reported test results from New Jersey’s statewide testing programs across districts. The measure was first developed in 1974 using demographic variables from the 1970 United

States Census” (NJDOE, 2014, p. 16). The comparison examines education, income, unemployment, occupation, and income level of residents. “Districts were then ranked according to their score on this measure and divided into eight groups based on the score interval in which their scores were located. Eight DFGs have been created based on the 1990 United States Census data. They range from A (lowest socioeconomic districts) to J (highest socioeconomic districts) and are labeled as follows: A, B, CD, DE, FG, GH, I, J” (NJDOE, 2015).

Flexible Ability Groups – Groups of students divided into levels based on their knowledge and/or performance and reassessed to ensure students are in appropriate ability groups. Teachers determined student reassignment to new groups based on a change in performance or skill level. This strategy may be used within the classroom or between the grade level(s). There is movement between these groups (Ford, 2005).

Frustrational Reading Level – The reading level at which the text becomes too difficult for the reader; the reader has 90% or less accuracy rate on word attack, comprehension, and/or fluency (Fountas & Pinnell, 1996; Pearson, 2011).

Homogeneous Grouping – Groups established based on a common factor such as reading performance, ability or skill.

Heterogeneous Grouping – Groups are a representative sample of the population and are composed of various reading levels, abilities, or skills.

Guided Reading – a method of teaching reading in which select groups of similar students are brought together for a time to receive instruction on a specific skill using a text that is instructionally appropriate for the group. Teachers can form groups based on

skill, ability, reading level, or prior knowledge. Groups are often flexible to allow for the continuous assessment and remediation of skills (Fountas & Pinnell, 1996).

Independent Reading Level – The reading level at which a reader can read the text without teacher support and show mastery of fluency, comprehension, and word attack skills (Betts, 1946); the reader has 95% or greater accuracy rate on word attack, comprehension, and fluency (Fountas & Pinnell, 1996; Pearson, 2011).

Instructional Reading Level – The reading level best determined to teach a new skill with teacher support; the reader has a 91%-94% accuracy rate on word attack, comprehension, and fluency (Fountas & Pinnell, 1996; Pearson, 2011).

Joplin Plan – Students are assigned to reading groups based on performance. There is movement between the grades. For example, a student in third grade may move to fourth grade for reading. This fourth grade ability reading groups may have third, fourth, and/or fifth grade students working together. In order for this to occur, students in various grade levels must have reading at the same time (Carson & Thompson, 1964).

Leveled Readers – Books used during guided reading instruction that are written on a pre-determined reading level and are stratified to accommodate the various reading levels. The content of the book may support a specific skill that is used to build a reader's fluency, comprehension, or word attack skills. Length, layout, structure, organization, illustrations, words, phrases, sentences, literary features, content, and theme are all factors considered when determining the level of a reader (Fountas & Pinnell, 1999).

New Jersey Assessment of Skills and Knowledge –The New Jersey Assessment of Skills and Knowledge (NJASK) is a criterion-referenced assessment mandated by the

state of New Jersey in Grades 3-8 until the spring of 2014. Students are assessed on their attainment of the Common Core State Standards (CCSS) or New Jersey Core Curriculum Content Standards (NJCCCS). Students are rated as Advanced Proficient, Proficient, or Partially Proficient in math and language arts in Grades 3-8 (*NJASK 3 Score Interpretation Manual*, 2013).

New Jersey Proficiency Assessment of State Standards (NJ PASS) – Criterion-referenced test published by Houghton, Mifflin, and Harcourt, which assesses students' knowledge of the New Jersey Core Curriculum Content Standards. Students are rated as Advanced Proficient, Proficient, or Partially Proficient in math and language arts in Grade 2 (New Jersey Proficiency Assessment, 2015).

Tracking – Assigning students to a group of classes based on overall achievement. The grade level population is divided into high, middle, and low achievers tracks. Once assigned to a track, students will take the assigned courses designed for their designated track. There is minimal movement between tracks.

Within-Class Grouping – Forming groups within a classroom based on skills, abilities, characteristics, and/or random assignment. These groups may be subject-specific or formed for a given non-instructional reason. There may or may not be movement between the groups within the class.

Whole Class Instruction – Students are taught as one unit. Students use the same materials and participate in the same assignments. This strategy may be used in homogeneous or heterogeneous classrooms

Organization of the Study

Chapter I discusses the use of reading groups within and between the grade level and the impact this has on meeting the federal mandated acts, NCLB and/or ESSA and the state teacher evaluation rating systems. The problem is defined and definitions outlined. Chapter II presents a review of the literature defining and organizing grouping. The history of grouping is documented and the reasons teachers do or do not group are discussed. Empirical data and Vygotsky's zone of proximal development theory are reviewed as well. Chapter III explains the design methods and procedures for this study, and Chapter IV illuminates the data and statistical findings of the independent variable, reading achievement. Chapter V provides the statistical summary and the implications for educational policies and practice. The conclusion of the study in Chapter V is based on the research question: What is the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3)?

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

Since the 1800s there have been efforts by teachers to reduce the academic distance between the highest and lowest achievers in a classroom. Teachers have formed groups in classrooms to minimize the variation of the learners, targeted instruction to students' abilities, and modified learning materials. There have been various forms and titles associated with this ability grouping strategy used by teachers to reduce variability. Ability grouping has been referred to in the United States of America as tracking, achievement-based grouping, skills-based grouping, and flexible grouping, while in Europe and Australia ability grouping is called streaming or setting. The variability associated with ability grouping is not limited to titles only but extends to their forms and implementation as well.

Sorting students by academic performance level has long been at the heart of most student grouping in schools. The definition of ability groups advanced by Slavin (1987), a professor from John Hopkins University and influential researcher in the field of grouping, remains one of the most commonly referenced in current research studies. According to Slavin (1986), ability grouping is the “. . . grouping students for instruction by ability or achievement so as to reduce their heterogeneity” (p. 4). Others such as Worthy (2010), Oakes (1985), and Kulik & Kulik (1982) have offered similar definitions that all mention sorting students into groups based on ability or perceived ability. Some researchers have added to the definition of ability grouping by including information about the type of instruction used during ability grouping. Some of these researchers view

ability grouping as a form of differentiated instruction where learners receive instruction that can be presented using different modalities, curriculum, and/or levels for each group (Sondergeld & Schultz, 2008; Tomlinson, 2003). There are also researchers who define grouping in relationship to the amount of time associated with the group. For example, Fiedler, Lange, and Winebrenner (2002) discuss how ability grouping is not permanent, while Tieso (2003) describes ability grouping as being more permanent than skill grouping without providing the specifics of when the frequency of movement changes the definition from ability grouping to skills grouping. Without a detailed context of the conditions and one accepted definition, it is challenging for researchers to discuss the results of grouping studies in a more thorough and efficient manner and for practitioners to understand the results when there are various criteria, terms, and definitions associated with ability grouping.

Literature Search Procedures

In order to learn about the relevant work and documents related to the use of grouping in the classroom, searches were conducted on government reports and academic articles obtained from EBSCOhost, ERIC, the United States federal government, and the New Jersey Department of Education websites. ERIC and EBSCOhost were searched using the following key words: grouping, elementary level, reading, and flexible grouping. In addition, the key words “guided reading” and “elementary school” yielded additional research-based articles. The federal government website provided information on NCLB, ESSA, and various laws to assist students, and NJDOE’s website provided information on New Jersey’s NCLB waiver, teacher evaluation formulas, testing results, and information on district demographics

Inclusion and Exclusion Criteria for Literature Review

Studies that met the following criteria were included in this review:

1. Included a sample that consisted of Grades K-5 in a variety of combinations
2. Used experimental, quasi-experimental, correlational, and meta-analysis designs
3. Published dissertations
4. Used quantitative methodology. Only a few studies used qualitative methodology and were included in order to add to the reader's knowledge base about the complexity of the topic or the theoretical framework.
5. Published within the last 15 years unless considered seminal work that provided the foundation for later developments
6. Contained literature and reports from government reports
7. Included descriptive information that added clarity to the topic

Focus of the Review

The literature review focused on the use of grouping during reading in elementary schools. To further examine the practice of reading grouping in elementary schools, the literature review focused on describing the practice, forms of grouping, and the history of grouping in the United States. The literature review then focused on the reasons teachers may or may not use grouping in the classroom. Finally, the literature review examined the quantitative studies on grouping practices and its implications for reading achievement.

Limitations of the Review

This literature review is limited by the sparse amount of research on targeted reading at the elementary school. The vast majority of the research focuses on secondary schools and how tracking impacts students academically and emotionally. Although there is an abundance of research on grouping from the 1980s and 1990s, there is limited research from the last ten years, especially on the specific grouping practice in question in this proposed study: flexible grouping between the same grade level.

This review does not examine the social and emotional implications of grouping on students and/or teachers' perceptions. There are many qualitative studies that explore the impact of grouping on students' self-perceptions or teacher beliefs. However, this literature review does not provide an in-depth discussion on these topics. The research discussed on the social and emotional consequences is meant to share some general findings by selected researchers.

Review of Literature Topics

Structures of Grouping: Within-Class and Between-Class

Despite variation in the definition of ability grouping, there are commonalities in how ability grouping is implemented that differentiate ability grouping into distinct categories. There are two major types of ability groupings used in schools: between-class and within-class ability grouping (LeTendre et al., 2003; Slavin, 1987). Within-class ability grouping is an instructional strategy that assigns students to a specific group based on achievement level (LeTendre et al., 2003; Slavin, 1987). It is working with a homogeneous group within the general classroom (Fountas & Pinnell, 1996). Math groups, guided reading groups, group work, flexible grouping, skills grouping, random grouping,

and even cooperative learning are examples of how teachers group within the heterogeneous classroom. Within-class ability grouping is a frequently used instructional strategy at the elementary setting (Loveless, 2013). Ability grouping is most common in kindergarten through Grade 3 in reading and in Grades 4 through 6 in math (Loveless, 2013). It is also the most common form of grouping in the United Kingdom (Hallam, Ireson, Lister, Chaudbury, & Davies, 2003). There may or may not be movement between groups when using within-class ability grouping. For example, some teachers may have assigned groups for reading based on performance at the beginning of the year and maintain this group throughout the year, while other teachers change reading groups based on student performance or skill development.

The second distinct form of ability grouping, between-class ability groups, can be organized in various forms. These include homogeneous classroom assignment, subject grouping, self-contained and resource room special education, gifted classes, and the Joplin Plan for reading. Homogeneous classroom assignment is when students are assigned to a group based on their prior knowledge, cognitive ability, or teacher judgment. Between-class ability grouping can also be subject-based as is done at the secondary level when, for example, a student may be placed in an advanced placement math class but is in a general education English class. Between-class ability grouping at the elementary setting may have students receiving homogeneous instruction in one or two subjects such as in math and/or reading based on a student's performance. The Joplin Plan, which places students into reading groups across grade levels, is a form of between-class ability grouping. The Joplin Plan was first used in Joplin, Missouri, in 1952 with students in Grades 4 through 6 (Carson & Thompson, 1964.) Students in fourth grade could possibly

be attending a second grade, third grade, or a higher grade's reading class. Last, Slavin's definition of between-class ability grouping also includes gifted and special education classes (Slavin, 1988).

Flexible Grouping

One form of grouping that can be found in both within- and between-class ability grouping is flexible grouping (Tieso, 2003). Flexible grouping, when used within the classroom, is a group that is formed based on ability, skills, or interest of those students assigned to a classroom and is continually reassessed to ensure accurate placement of members in the group (Graves, Juel, Graves, & Dewitz, 2011; Haghghat, 2009; Opitz, 1998; Radencich & McKay, 1995). Flexible grouping advocates believe the students are capable of learning skills and progressing in the given subject with the acquisition of these skills (Caldwell & Ford, 2002; Condrón, 2005). Condrón (2005) believes students learn or grow at different rates. The current assignment to a group does not indicate a student's potential. Teachers must continually assess and adjust group lessons to meet the needs of the learner. This belief is the foundation of flexible grouping. It is the goal of the teacher who works with students who need support in a subject to teach the students the skills that enable them to progress to another group (Loveless, 1998; Slavin, 1988).

The needs of the students drive the flexible grouping process and lessons address these needs (Connor, Morrison, Fishman, Crowe, Otaiba, & Schatshneider, 2013; Opitz, 1998; Tomlinson 2003, 2005; Wormelli, 2007). Flexible groups are formed to target a specific skill (McCoach, O'Connell, & Levitt, 2006), and teachers can adapt the pace and content of instruction based on the instructional level of the students (Barr, 1995; Ford, 2005). Flexible grouping or skills grouping is typically subject-based and fluid (Fieder et

al., 2002; Radencich & McKay, 1995). Opitz (1988) compared the differences between ability grouping and flexible grouping to concrete and sand because of the fluidity in the flexible groups. Students are assessed and moved from the group based on performance, whereas ability grouping is determined for a set duration (Caldwell & Ford, 2002).

A majority of the flexible grouping in the elementary school is taking place during reading and/or math instruction (Nagel, 2001). Slavin (1988) recommends limiting grouping to one or two subjects, like math and/or reading; and Caldwell and Ford (2002) and Slavin (1988) suggest these groups be reassessed for group reassignment and be differentiated in pace and instruction. This differentiated instruction allows the teacher to tailor the curriculum to the students' changing needs (Caldwell & Ford, 2002; Connor et al., 2007; Hallam et al., 2004; Kulik & Kulik, 1987). According to Tieso (2005) and colleagues, one curriculum for all learners is not reasonable or effective (Ford, 2005; Tieso 2005). This differentiated curriculum allows the flexible grouping lessons to be matched to the learners' zone of proximal development (Vygotsky, 1978) for more effective instruction in an appropriate environment (Caldwell & Ford 2002; Wormelli, 2007) while having students associating with the heterogeneous group (Kilgore, 1991; Slavin, 1987).

Current trends in flexible grouping are a result of the evolution of ability grouping that has occurred since the early 1800s. Historical events, calls from the public, and federal mandates have all influenced the nuances in grouping formation and its justification over time.

Historical Background

Early schools in the United States were often a one-room building and contained numerous grade levels in the building. Historical documents indicate some teachers had to group the students by ability in order to deal with various ages and academic abilities in the classroom (Algozzine & Anderson, 2007; Valentina, 2000). Some students worked on reading their primers, while the youngest students wrote their letters on a slate board. Ability grouping was recorded as early as the 1800s (Otto, 1934). In 1848 Quincy Grammar School in Boston, Massachusetts began to change its organization toward a graded “lock and step” model (Venezky & Bregar, 1988). This “lock and step model” divided students by ages into grades and was a form of age-based homogeneous grouping as was the practice in Germany. There were two main forces in the first half of the twentieth century which continued to propel grouping as an organizational structure: the introduction of the Intelligence Quotient test and the wave of immigrants entering the United States in the 1900s.

Intelligence Quotient and Its Effect on Grouping

In 1904 Binet developed the Intelligence Quotient (IQ) assessment, which introduced stratification of intelligence levels (Ansalone, 2006). Binet believed intelligence could be measured through an assessment of memory skills, associations, visual reasoning, and other various questions and activities. It was believed that IQ was fixed and indicated one’s ability to learn and acquire new skills (Fancher & Rutherford, 2012). Students were categorized by terms such as average, below average, or above average to describe a student’s IQ in relation to his peers. The use of IQ tests played a significant role in determining a student’s track placement during the 1920s (Mirel, 1999).

In fact, Detroit gave all first graders an IQ test. The results of these tests determined if students were placed in the X track which encompassed those students in the top 25% of IQ scores, Y track with those students who scored in the middle range, or Z track comprised of those students scoring in the bottom 25% on the IQ test. Detroit's reading program became known as the XYZ Plan (Mirel, 1999). In order to deal with the perceived fixed levels of intelligences, teachers started to form ability groups within their classrooms or between the grade levels (Nagel, 2001). *The Story Hour Readers Manual* by Ida Coe and Alice Christie published by the American Book Company in 1913 was the earliest reference to ability grouping in a teaching manual and provided suggestions on how to deal with the varying abilities in the classroom (as cited in Ansalone, 2000). High school students started to be placed in tracks such as college preparation, vocational, or general education based in part on the results of IQ testing (Oakes, 1985). Slavin (1988) also found that ability grouping or streaming increased in the 1920s as result of more standardized testing being introduced into schools.

Influence of Immigration on Grouping

The Intelligence Quotient test may have accelerated grouping and tracking, but the rise in the number of immigrants made tracking and grouping more commonplace in the American school. Until the 1890s the United States was predominately Anglo-Saxon, but the next wave of immigrants brought people from Southern and Eastern Europe followed by immigrants from South America and Puerto Rico who needed to be integrated into the current school system (Weisberger, 1994). These immigrants came to the United States in search of manufacturing and skilled laborer jobs (Weisberger, 1994). The educational system needed to respond to the influx of non-English speakers and the

demand to train a work force for manufacturing and skilled labor jobs (Lucas, 1999; Oakes, 1985). Grouping was a strategy to deal with the diversity of learners entering the school systems in the early 1920s (Ansalone, 2000; Barr, 1995; Slavin, 1988; Worthy, 2010).

Worthy (2010) described how the comprehensive high school emerged as a way to group students based on their desire or believed ability to attend college or be trained to enter the work force. High schools began to track students by courses under the umbrella of a college preparation, honors, general, or vocational track. Schools started to offer such classes as woodworking, home economics, and typing and guided students to take specific classes based on assessments, IQ, or gender. Once students were placed in a track, the classes offered to the student were only those courses available to that track. This limited course offering deterred students from taking a course in a subject area that may have been an area of strength. Researchers such as Oakes (1985), Worthy (2010), and Lucas (1999) have argued that tracking was intended to keep races and children of different socioeconomic status segregated.

Cities in the United States such as Joplin, Denver, Detroit, and Winnetka developed plans to deal with various differences in ability and waves of immigrants entering the schools (Barr, 1995). Detroit, Michigan, decided to implement across-grade level grouping in 1928 and called this between-grade-levels grouping the XYZ Plan (Cushenbery, 1967; Kulik & Kulik, 1992). Students were assessed based on their IQ and then moved between the grade level based on their IQ quartile. Grouping was happening at the elementary and middle school levels, but social and political factors brewing in society would be a catalyst for change in the organizational practices in the schools.

The practice of ability grouping continued and was evident in high schools under names such as Honors, General Education, and Vocational tracks. Ability grouping was found in the elementary schools under the guise of reading groups with appealing names such as “bluebirds,” “robins,” and “eagles” (Worthy, 2010). A survey published in 1961 found that 80% of elementary schools were using ability-based reading groups at the time (Austin & Morrison, 1961). However, during the 1960s and 1970s, as citizens including minority groups vocalized their opposition to social inequality, the ability-based practices in the public school would soon be scrutinized. In 1954 the Supreme Court’s landmark decision, *Brown v. Board of Education of Topeka*, made it unconstitutional to segregate schools based on one’s color.

The Equality Movement

Despite the Supreme Court’s ruling on the unconstitutionality of schools separated by race, the country was not quick to integrate. In 1957 President Eisenhower had to send members of the United States army to ensure that students of color were permitted into the Arkansas school where only Whites attended (Kirk, 2008). Because integration was resisted in many areas of the country, the U.S. Department of Education (USDOE) wanted to investigate the topic of equality in education. *The Concept of Equality of Educational Opportunity* (Coleman, 1967), more commonly referred to as The Coleman Report, was commissioned by the U.S. Department of Education in the 1960s and was the largest study done to date on equality in the public schools.

The Coleman Report, issued ten years after the landmark *Brown v. Board of Education of Topeka* case, reported that most segregation was happening within schools, not between schools. Coleman’s (1967) findings echoed the complaints presented in

Hobson v. Hansen, a case in which the plaintiff, a civil rights activist, argued the district was resisting desegregation by maintaining ability groups (Slavin, 1988, p. 295). Hobson argued his children were given a test and placed in a rigid track with unqualified teachers and low-level curriculum. The circuit judge ruled in Hobson's favor that the tests used were biased and were used as a tool to keep Black children segregated (*Hobson v Hansen, 1967*). The U.S. Department of Education, while examining the plight of the Black students, also wanted to address the needs of those students not being served properly by the public schools. It was during this decade that programs such as Title 1, special education, and gifted and talented were developed. The U.S. Department of Education started to categorize learners and provide funding for development of programs for specific categories (Loveless, 1998). Title 1 of the Elementary and Secondary Act of 1965 provided funds to schools with low-income students with the intention of decreasing the achievement gap between low-income students and students in the middle and upper incomes (Improving Basic, 2015) and Title III of the Elementary and Secondary Act of 1965 provided funding for bilingual education (USDOE, 2015). The federal government may have started to provide funds to address inequalities in the educational system but never tackled the grouping or tracking systems used in many schools. The idea of grouping students continued to be an accepted academic practice, but this practice of grouping would again be challenged.

The 1970s were a time of social change, and many members of society such as the National Association for the Advancement of Colored People, National Governors Association, the National Education Association, and the National Council of Teachers of English began questioning the use of grouping and tracking (Hallinan, 2004; Worthy,

2010). During the 1970s, research conducted on tracking found the practice ineffective for the lowest levels (Allington, 1980; Esposito, 1973).

Scholars also discussed the ethical considerations associated with tracking and called on researchers to examine the impact on students' self-esteem (Worthy, 2010). Educators such as Rosenbaum (1976), Sorensen (1970), and Oakes (1985, 2008) argued democracy should be inclusive and prevent educational elitism through the use of a tracking system. Those who argued against tracking/grouping centered their arguments around four main themes: labeled students, maintained segregation by race, lowered expectations in the lower tracks, and minimized exposure to rich curriculum (Eder, 1981; Lou et al., 1996; Lucas & Gamoran, 2002; Rist, 1970). Advocates who promoted tracking emphasized the research that grouping allowed teachers to tailor instruction to improve achievement and improved classroom management (Ansalone & Biafora, 2004; Hallam et al., 2003; Slavin, 1987). As the discussion on the ethical consideration of the use of tracking emerged at the end of the 1970s, Oakes (1985) wrote a pivotal book on the practices of tracking in place at that time. *Keeping Track* challenged the various forms and names of homogeneous grouping such as tracking, ability grouping, or college preparation that were still being used even after the ethical debates of the 1970s and early 1980s. A study conducted by McPartland, Colidron and Braddock (1987) surveyed all the elementary schools in Pennsylvania and found that 90% of primary schools and 85% to 90% of upper elementary schools were using grouping and 70% of these schools reported grouping between the grades for one or two subjects (1987). Oakes discussed the research, which examined the impact of these forms of ability grouping on self-esteem and its perpetuation of inequality. Blacks and Hispanics were often overrepresented in the lowest

tracks and were often included in the classes with the lowest socioeconomic group (Haller & Davis, 1980; Rist, 1970). The overrepresentation of minorities, students of low socioeconomic status, and lower IQ scores in the weaker tracks/groups was believed to have been a subtle way of keeping minorities in certain roles and maintaining their placement in society (Loveless, 1998).

The research on grouping was at its peak in the 1980s. Researchers such as Slavin (1986) and Kulik and Kulik (1982) presented meta-analytical reviews of the research on heterogeneous and homogeneous grouping that had been conducted to date and these results were not conclusive that one form of grouping yielded improved academic results for all levels (Slavin, 1987, 1988; Kulik & Kulik, 1982, 1992). Some schools began to de-track based on the inconclusive evidence that grouping by ability improved students' achievement. Teachers began to teach to heterogeneous groups or reduce the ability-based classes to math or reading. However, not all schools were stopping or limiting ability-based grouping. In 1998 Loveless reported two or three groups per class were typical in the elementary school. Lucas suggested tracking was still occurring in the 1990s even after the ethical and academic debates from the 1970s and 1980s because master school schedules were difficult to change and the schools received pressure from parents to have their children academically challenged. These factors might have impeded the detracking initiative.

Accountability Era and Student Grouping

As previously discussed, there were factors which facilitated the growth of grouping in the early 21st century, and new factors emerged at the end of the same century to keep grouping in the schools. Data collected by the National Assessment of

Educational Progress (NAEP) found the frequency of ability grouping in fourth grade reading instruction increased from 28% in 1998 to 71% in 2009 (Loveless, 2013). There were two events that impacted the state of ability grouping at the elementary level at the end of the 20th and beginning of the 21st century: the passing of the No Child Left Behind Act of 2001 (NCLB) and the introduction of guided reading in the late 1990s.

NCLB required that all students be Proficient in reading and math by 2014. States began to design criterion-referenced tests to measure students' progress and submit assessment results to the United States Department of Education (USDOE). Each school would receive a score on their Adequate Yearly Progress report noting the school's progress in reaching 100% of their student population scoring at or above the proficient level determined by each state (NCLB, 2001). In 1997 New Jersey started their testing in Grades 4 and 8 under the names The Elementary School Proficiency Assessment and The Eighth Grade Proficiency Assessment, respectively, and then expanded testing to Grades 3 through 8 as required by NCLB by the 2005-2006 school year. States around the country were administering their own version of a state test in order to report scores to the U.S. Department of Education. Test results were published in the newspapers and the NCLB mandated school report card. These school report cards had to list assessment results of students by subgroups such as race and special education classification. It became paramount that students score at or above the Proficient level on the state test because parents, newspapers, and state governments compared schools based on these published test scores.

In order to increase scores, educators examined ways to maximize their efforts to help low scoring subgroups on their state assessments. Schools began to remediate those

not passing by offering remediation classes, test preparation classes, and after school tutoring clubs to improve results. Some schools also reintroduced ability-based classes for students scoring below proficient or at-risk for failing the assessment under the guise of “test prep” classes. Educators assessed data and started to focus their remediation efforts on the “bubble kids” (Booher-Jennings, 2005). Bubble kids are those students who missed the proficient level by a minimal number of points, and educators hoped with intervention these students would increase their score to the proficient cut score. As educators focused their efforts on moving students on the bubble to a passing score, the improvement by the average, above average, and gifted learner was limited (Neal & Whitmore-Schanzenbach, 2007).

In December 2015 President Obama signed the Every Student Succeeds Act (ESSA), a reauthorization of the 1965 Elementary and Secondary Education Act (Every Student, 2015). While this law does require testing of students in Grade 3 through Grade 8 every year in math and language arts and once in high school like NCLB, it shifts the ownership and accountability of those tests back to the states. Under ESSA states control the testing, benchmarks, factors of school success, and plans to remediate failing schools (Every Child, 2016).

Teacher Accountability Movement

In 2010 NCLB was reauthorized and U.S. Secretary of Education Arne Duncan announced in the summer of 2011 the U.S. Department of Education would create waiver options for states if specific criteria were met in the waiver application. According to the U.S. Department of Education, 43 states have a waiver or an extended waiver (USDOE, 2015). These states had to agree to raise standards to prepare for college and career

readiness, improve teacher effectiveness, and include teacher accountability. States applying for waivers needed to include information on how teachers were going to be held accountable for all their students' performance. This was a significant change from the original NCLB of 2001. The 2001 NCLB, as previously discussed, held schools accountable, but the waiver application of 2011 shifted the accountability to the teachers and administrators.

In order to deter teachers and administrators from focusing solely on the bubble kids, some states proposed in their waiver application to measure Adequate Yearly Progress by using a Student Growth Percentile (SGP) model, designed by statistician D. W. Betebenner, rather than a model that examined the percentage of students attaining the minimal score established by each state (O'Malley et al., 2011). This new SGP formula is based on students' past performance on the state assessment. The state monitors a student's academic progress by comparing peers of similar caliber, thus allowing the state to monitor individual progress of students no matter how the child performed (Betebenner, 2009). The SGP model also allowed states to gather data on how students performed under the guidance of their teacher. This SGP model made many teachers look at how all students performed, not just certain groups and ability grouping was one possible strategy to serve all groups of students (Loveless, 2013).

Beginning in September 2013, New Jersey began monitoring the progress of each student, not just the collection of summative school data. New Jersey, Pennsylvania, Arkansas, Iowa, Florida, Tennessee, Georgia, Colorado, and Delaware are some states using a growth model to assess the effectiveness of its teachers and students (Bonk, Copa, Gibson, Gillan, Nau, Peoples, Wang, & Woolard, 2012). Even with the reauthorization of

the Elementary and Secondary Act in 2015, teacher accountability will still be a component, but it will be up to the states to determine how to measure this (ESSA, 2015). States like New Jersey created laws which include student performance as an indicator of teacher effectiveness, and these accountability measures are still in effect (Achieve NJ, 2012).

With this emphasis on individual yearly improvement, schools must again focus on meeting the needs of all learners, not just those at risk. Ability grouping is finding its place again in the American primary school. According to a survey conducted by Chorzempa and Graham (2006), three times as many teachers reported using ability grouping when compared to a similar survey by Baumann and Heubach in 1996.

The increase in ability grouping gained popularity in the elementary schools as guided reading became a commonly used reading practice in the primary reading classrooms (Cunningham, Hall, & Cunningham, 2008; Ford & Opitz, 2011; Fountas & Pinnell, 1996). Fountas and Pinnell's influential book, *Guided Reading: Good First Teaching for All Children*, revamped the guided reading practices suggested by Betts in 1948 in the classic text, *Foundations of Reading Instruction*. Guided reading requires teachers to ability group within the classroom. Students are grouped according to skill, prior performance, or assigned reading level (Fountas & Pinnell, 1996). Teachers then meet with these ability-based guided reading groups to work on a skill during the reading period using a scaffolded lesson (Vygotsky, 1978). While the teacher works with a small group of students, the remaining students review or apply skills in literary centers (Ford & Opitz, 2011). Guided reading groups are meant to be fluid and based on a student's needs. Teachers assess students using some type of reading progress-monitoring system

like the Developmental Reading Assessment 2, observations, classwork, and reading interest behaviors. Teachers then group students based on reading level, skills, interest, book genre, and/or motivation. These groups are formed until the objective is met and then may be dismantled. Teachers are constantly reassessing students' needs and skills and changing the groups accordingly. Guided reading groups are fluid and flexible and are not rigid as the typical tracking group is. (Fountas & Pinnell, 1996).

Grouping in the Contemporary Classroom

Grouping has been evident since the colonial schools in the United States and it continues to be present in the modern classroom (Ansalone & Biafra, 2004; Chorzempa & Graham, 2006; Ford & Opitz, 2011; Le Tendre et al., 2003). Grouping use increased after the development of standardized tests and in response to the immigration peaks that emerged in the 1900s. Teachers also grouped to deal with the academic variability that existed within the classroom (Ansalone, 2000; Slavin, 1988; Worthy, 2010).

Prevalence of Grouping

In order to examine the frequency in which grouping is used in the contemporary classroom, Loveless (2013) reviewed the survey results from the National Assessment of Educational Progress, a national test taken by a sampling of the U.S. population. In 1998, 28% of fourth grade teachers reported using ability grouping, in 2005 the percentage increased to 59%, and in 2009, 71% of the teachers reported using ability grouping (Loveless, 2013). This evidence is comparative to Chorzempa & Graham (2006), who found 63% of teachers of Grades 1 through 3 were using within-class ability groups in reading. McCoach, O'Connell, and Levitt (2006), who examined the data from the Early Child Longitudinal Study, found the kindergarten teachers surveyed reported using

groups for reading once a week. The practice of grouping is not limited to the United States. Wilkinson and Townsend (2000) found grouping was a common practice in New Zealand, as 94% of teachers of Standard 3 (Grade 4) reported dividing into ability groups for reading; and Hallam, Ireson, and Davies (2004) found 50% of primary schools in England were using grouping for at least one subject. Grouping is being used in many classrooms around the globe, but why is grouping still evident in the modern classroom?

Justification for Grouping

Studies have been conducted to determine why teachers continue to group. The research discusses three main reasons teachers continue to use grouping as an instructional practice. Teachers report they continue to group because they believe grouping improves student achievement (Chorzempa & Graham, 2006; Ford, 2005; Hallam et al., 2004; Tomlinson, 2003; Wormelli, 2007), grouping is an easier way to address the range of abilities in the classroom (Ansalone & Biafora, 2004; Hallam et al., 2004; Slavin, 1988; Tomlinson & McTighe, 2006) and individual differentiated instruction is challenging to implement (Loveless, Parkas & Duffett, 2008). Each of these three reasons are examined in this paper.

Justification for Grouping: Range of Abilities

The heterogeneous classroom is composed of students of differing abilities (Firmender et al., 2013). The results of the 2009 administration of the National Assessment of Elementary Progress (NAEP) indicate there are varying abilities in a classroom. The top 10% of those tested in fourth grade were reading six grade levels beyond the lowest 10% and there were three-grade-level difference between 25% and 75% of the students (Petrilli, 2011, p. 49). Recently, Firmender et al. (2013) examined

over 1,000 students in Grades 3 through 5 and found classrooms have approximately a nine-year span in reading levels in Grade three and approximately an eleven-year reading span in Grades 4 and 5. Chorzempa and Graham (2006) surveyed 222 schools with Grades 1 through 3 and found the mean grade difference between the highest and lowest reader in the class to be 3.3 grades with a standard deviation of 1.4. Pressley (2006) and Fountas and Pinnell (1996, 1999) have noted students enter the schoolhouse at different levels and these differences continue to exist in the primary classroom. Petrilli (2011) added that variability is inevitable; but when the variability in abilities such as six reading levels in one classroom is great, “no one wins” (p. 51).

Justification for Grouping: Improves Achievement

Hallam, Ireson, and Davies (2004) surveyed 804 primary school teachers in the United Kingdom about their reasons for using grouping and discovered teachers believed grouping increased student achievement. Ansalone and Biafora (2004) surveyed 124 public, elementary teachers in the United States and found 57% of responders felt high achievers learned best with others with the same or similar ability. In order to deal with the diversity of learning levels, teachers will deliver their whole class lessons to the average group in their classroom (Hallam et al., 2003; Tomlinson, 2005; Wormelli, 2007). Ford (2005) and Slavin (1987) have postulated that teaching one lesson is perfect for some students in the classroom but ignores the needs of the other students by being either too difficult or too easy.

Teachers use grouping as a means to improve achievement of those outside the average range usually addressed during whole group instruction. Grouping allows teachers to customize a lesson for those who may be working below or above the lesson

presented to the average learners (Ford, 2005; Scigliano & Hipsky, 2010; Slavin, 1987; Tomlinson, 2003). Kulik and Kulik, (1984) and Tieso (2003) have argued philosophically that the advanced learners have a right to have their needs met too. Cogan (1995) used a sports analogy during her philosophical argument in favor of grouping. Cogan stated, “It is ironic that school systems who criticize academic grouping as a form of discrimination allow discrimination based on athletic ability” (par. 15). Results of the NAEP since the NCLB era have documented higher achieving learners’ growth has remained stagnant, while students at the lower level have shown growth (Loveless, 2013). There are researchers who advocate teaching students in ability groups so teachers can better address the needs of the learners by differentiating curriculum and adjusting pace to provide the necessary prerequisite knowledge to master grade level curriculum (Ford, 2005; Fountas & Pinnell, 1996; Tomlinson, 2003).

Vygotsky (1978) believed students learn best when learning new information with teacher support and then removing the support as students gain more independence in completing the work. Because students have variability in prior knowledge, students would need different teacher support to progress. Vygotsky (1978) called his theory the zone of proximal development. Grouping is the strategy that allows the teacher to design a lesson for the various levels within the heterogeneous classroom.

Some teachers use grouping because they believe grouping improves reading achievement in the current placement and in the students’ future reading endeavors. Lesnick, Goerge, Smithgall, and Gwymne’s (2010) longitudinal study examined how third grade reading levels were an indicator of four future events, including eighth grade reading level. Lesnick et al. (2010) found the correlation between third graders’ reading

levels and eighth grade reading achievement to be .67. In addition, Hernandez (2011) analyzed approximately 4,000 children born between 1979 and 1989 and tracked their reading and high school graduation rates. Hernandez found students who were not reading on grade level by third grade had a 16% rate of not graduating high school on time, a rate four times higher than those students reading on grade level in third grade.

Justification for Grouping: Management

With a three-year educational span in the first through third grade classrooms, (Chorzempa & Graham, 2006) teachers look for ways to organize their classroom instruction (Ansalone & Biafora, 2004; Hallam et al., 2004; Macqueen, 2010; Nomi, 2006; Slaydon, 2013; Wormelli, 2007). Tomlinson (2003) recommended differentiating instruction to meet the needs of the individual learner. Differentiated instruction requires teachers design lessons to meet the interest and/or prior knowledge of the individual learners. Teachers could have all the students in their classroom reading individual books focusing on each student's need, or teachers could offer a tiered lesson that has more than one objective to meet the varying abilities in the classroom. Tomlinson's work has foundations in Vygotsky's (1978) zone of proximal development. Tomlinson (2003) may recommend a differentiated curriculum, but some teachers have reported that differentiated instruction for all members of a class is challenging (Hertberg-Davis, 2009; Loveless et al., 2008).

In order to execute differentiated instruction, teachers need to be able to be strong lesson designers, asking questions at a variety of Bloom's Taxonomy (1956) levels and then planning independent lesson exploration that applies the objective at different levels or tiers. The teacher then must predict the time that is needed for the differentiated

activities to be completed so all students finish at approximately the same time. In order to differentiate and maintain classroom management, some teachers have used differentiated instruction by offering two or three lessons on an objective and then dividing students into groups to deliver the instruction. This classroom management strategy is used to address the various abilities represented in the heterogeneous classroom (Ansalone & Biafora, 2004; Hallam et al., 2003; Kulik & Kulik, 1982; Slavin, 1988; Slaydon, 2013; Wormelli, 2007). However, this division of the class under the guise of “differentiated instruction” is a form of ability grouping.

Negative Consequences of Grouping

While ability-based grouping clearly has its advocates, ability grouping does have its detractors. There are three main reasons some teachers and researchers have dismissed grouping as an appropriate instructional strategy. These teachers and researchers have reported ability grouping leads to an overrepresentation of minorities in lower groups, creates student labels, and hinders the learning of the at-risk learners (Oakes; 2005; Worth; 2010).

Negative Consequences of Grouping: Labeling

Labels can have a negative or positive connotation as discussed in Rist’s acclaimed article, “The Self-fulfilling Prophecy in Ghetto Education” (1970). If students are labeled “gifted” or “advanced,” students may have more opportunities to access better curriculum, materials, and teachers and benefit from the group’s title (MacIntyre & Ireson, 2002; Nomi; 2006; Oakes, 1985). However, if a student is labeled “at-risk,” “behind,” or “low,” there may be negative consequences. Students related to the assigned level of the group and teachers made assumptions about a student’s ability based on the

group association (MacIntyre & Ireson, 2002; Oakes, 1985; Worthy, 2010). This association with a group title or label is known as “trait theory” (Allport, 1937). Qualitative studies have found teachers make assumptions about students’ abilities in other subjects based on placement in another group (Oaks, 1985; Rist, 1970; Worthy, 2010). This may explain, in part, why there is often not movement between groups. Worthy (2010) found that teachers tended to keep students in the same track/group all year; this supports previous findings that found little or no movement between groups (Ford & Opitz, 2011; Nagel, 2001; Oakes, 1985; Worthy, 2010; Wruble, 2002). Even though this study does not examine the self-concept of labeling students, it is noted that this is a consideration for some teachers in deciding to use grouping as an instructional strategy (MacIntyre & Ireson, 2002; Nomi, 2006; Tieso, 2003).

Negative Consequences of Grouping: Segregation

Research has found that boys, Blacks, and Hispanics, and those students from low socioeconomic status are overrepresented in the lowest groups (Condron, 2005; Macqueen, 2008; Loveless, 2013; Oakes, 2005; Rist, 1970; Worthy, 2010). Oakes (1985), Lunn and Ferri (1970), and Worthy (2010) have argued that ability grouping is a euphemism for segregation and a way to keep certain ethnic and social groups in a set place. In fact, the National Association for the Advancement of Colored People (NAACP) has condemned tracking and ability grouping because of the overrepresentation of minorities and the belief that track placement is a reflection of a student’s current and future abilities (Loveless, 2013). Because ability grouping formation tends to be stagnant, advocates such as Oakes (1985, 2005), Slavin (1987), and Condron (2005) have discredited the practice of set ability groups and tracking as socially

unjust and destructive to the classroom community and the ideals of a democratic society. These philosophical arguments may cause teachers to ponder their decision to group.

Negative Consequences of Grouping: Hinders Achievement

Some teachers choose not to group based on the research on the detrimental effects of ability grouping for at-risk learners. Students in the lowest groups tend to have less qualified teachers (Oakes, 1985; William & Bartholomew, 2004) and receive instruction that is less creative and lacking in higher-level thinking development (Allington, 1980; Hallam et al., 2004; William & Bartholomew, 2004). The lower groups also tend to have limited resources and simplistic curricula (Oakes, 1985; Wright-Castro, Ramirez, & Duran, 2003). Studies have shown that teachers tend to spend time working on basic skills and rote memorization (Carbonaro & Gamoran, 2002; Macqueen, 2010; Worthy, 2010). Students in the lower groups are also not exposed to limited problem solving, higher-level thinking, cooperative learning, and enrichment (Applebee, Langer, Nystrand, & Gamoran, 2003). Students in these groups are also often victims of lower teacher expectations (Macqueen, 2010; Slavin, 1987; William & Bartholomew, 2004; Worthy, 2010). Worthy (2010) interviewed 25 teachers who taught various language arts ability groups. Teachers of the lower level students responded during researcher interviews that they often could not do projects or higher level thinking with the lower students. The teachers interviewed expressed the students could not handle the responsibility to complete a project or were not capable of doing higher leveled assignments. Worthy's findings were comparable to other researchers who interviewed teachers about their perceptions of different ability groups (Lleras & Rangel, 2009; Macqueen, 2010; Rubie-Davies, Hattie & Hamilton, 2006).

Qualitative studies have noted teachers describe groups based on perceived ability, placement, and home environment (Rubie-Davies et al., 2006; Worthy, 2010). This stereotyping can lead teachers to assume some students are not capable of achieving expected grade level benchmarks. This can lead to a Self-fulfilling Prophecy (Merton, 1948) or as later updated to the “Matthew effect” (Merton, 1968). The self-fulfilling prophecy suggests that an unfounded belief may influence people’s perceptions, words and actions, and cause this unfounded belief to come to fruition. The Matthew effect is based on the verse from the Book of Matthew 25:29 in the *King James Bible*. This verse is, “For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken even that which he hath.” The Matthew effect may explain why some students get placed in lower groups and remain with this group. These students will not progress because they are given the less qualified teachers and are exposed to lower level curriculum and expectations. Students facing these disadvantages may continue to remain in the lower groups; as the adage states, “the rich get richer and the poor get poorer.”

Empirical Studies

In order to complement the findings presented on the frequency of grouping and why teachers may or may not use grouping, the empirical evidence on the effects on student achievement must be reviewed. During the 1980s and 1990s grouping research was being published in peer- reviewed journals and in magazines distributed to teachers. Researchers such as Kulik and Kulik (1987, 1992), Slavin (1987), Gamoran (1992), and Lou et al. (1996) conducted studies or analyzed data during the height of grouping research. The twenty-first century brought a new group of researchers who wanted to

revisit grouping in the modern classroom. These researchers such as Jecks (2011), Lou, Abrami, and Spence (2000) and Puzio and Colby (2010) conducted meta-analyses of selected studies involving grouping, which included more recent studies. Other researchers such as Collins and Gan (2013), Lleras and Rangel (2009), Macqueen (2010), Condrón (2005), Haghghat (2009), and Duflo, Dupas and Kremer (2009) conducted quantitative/quasi-experimental studies on grouping.

The quantitative research on grouping to date has examined the practice using four lenses: meta-analysis of studies, ability-based grouping, within-class grouping, and between-the-same-grade level grouping. Some researchers have conducted meta-analyses of previous studies or examined the results of grouping practices in totality. These researchers, who examined the overall effects of grouping, analyzed data on all participants. On the other hand, some researchers selected to analyze the data by ability level to determine if students' prior reading abilities were affected by the use of grouped or ungrouped class formation. There is another body of evidence that examined grouping when used in the classroom compared to those students who received whole group instruction. Finally, there is very limited research on grouping between the same grade level. This section will examine the empirical findings through each of the four lenses.

First Lens: Comparing Homogeneous and Heterogeneous Grouping in Totality

During the height of grouping research Kulik and Kulik (1987, 1992), Slavin (1987), and Lou et al. (1996) conducted meta-analyses of studies. While each researcher may have included different studies in their meta-analyses, the findings were similar. Slavin (1987) and Kulik and Kulik (1987) found no significant difference between grouped and non-grouped student achievement when all the studies were considered.

Slavin examined 14 studies, which focused on the elementary school and Kulik and Kulik (1987) examined 51 studies from the secondary level. Even though Kulik and Kulik and Slavin examined different levels, they both determined students did not score significantly different between grouped and ungrouped students. Slavin found an effect size of .00, and Kulik and Kulik found an effect size of +.10 when examining students' academic achievement.

While some researchers focused on samples from the United States, there have been other studies that looked at international samples. Macqueen (2008) wanted to explore the data of grouped and ungrouped students in Australia. She examined the Literacy Basic Skills Test of 113 students of two primary schools in New South Wales and she found no significant difference between grouped and ungrouped students (effect size +.279). This would support the earlier work by Slavin (1987), Kulik and Kulik (1987), Mosteller, Light, and Sachs (1996) and Condrón (2005) who also did not find a significant difference between grouped and ungrouped students when comparing the results in totality (Condrón, 2005; Hallam, 2002; Hallam et al., 2004; Kulik & Kulik, 1987; Slavin, 1987). In fact, Slavin (1987) and Oakes (1985) postulated that fixed tracking widened the inequality between the high and low students.

Second Lens: Comparing Homogeneous and Heterogeneous Grouping by Ability Level

While research to date did not find a significant difference between grouped and ungrouped students, the next logical hypothesis would be, "Is grouping more effective for certain ability levels?" This next lens examines the research on grouping when a student's initial ability—high, middle, or low—is considered.

Kulik and Kulik (1992) wanted to examine grouping by ability level and did find honors or gifted students performed significantly better in a grouped classroom when compared to their ungrouped peers, (effect size +.40). Slavin (1987) commented that even though Kulik and Kulik (1992) determined higher performing students might have performed better in homogeneous groups, the students in the lower groups “dragged down” the results. Economists Iberman, Kugler, and Sacerdote (2012) found a similar impact on student performance for those students who had been evacuated to Houston after Hurricane Katrina. The higher achieving students had a positive influence on their new classmates, whereas their low-achieving students “dragged down” the achievements of classmates. Slavin’s (1987) reflections shared previously exemplify the complexity of examining data on grouping. Synthesis and analysis may be occurring through the four lenses.

Nine years after Kulik and Kulik’s (1987) and Slavin’s (1987) pivotal publications, Lou et al. (1996) conducted a meta-analysis of studies which divided the learners into three achievement groups: high, medium, and low. There were four studies that compared homogeneous to heterogeneous groupings in reading and the meta-analysis of these four studies was a mean effect size of +0.36 in favor of homogeneous ability reading groups. However, it must be noted that Lou et al.’s low achieving students did not do as well in the homogeneous ability groups (effect size = -0.60). This finding supported the claims of those who used different age samples in their studies such as Condon (2005) and Slavin (1987) that lower level students have better achievement in a heterogeneous classroom.

Macqueen (2010) also wanted to further examine the ability-based subgroups of 113 New South Wales students involved in the Literacy Basic Skills Test. Even when examining the results of students based on their predetermined reading ability, there was no significant difference between grouped and ungrouped students for high, medium, and low achieving students. This was similar to Butler (1987), who wanted to analyze if students grouped by achievement had better comprehension skills or overall reading achievement. Butler examined 186 second graders who either received reading instruction in ability groups (high, medium, low) or in heterogeneous grouping. Butler found there was no significant difference between the grouped and non-grouped second graders in overall reading; however, she did find that grouped students scored significantly better in their comprehension skills than their non-grouped peers.

In a more recent study, McCarter (2014) analyzed reading scores of third, fourth, and fifth grade public school students on the Tennessee state assessment. She examined scores from 50 schools; 30 schools used whole group instruction and the other 20 schools used ability grouping. McCarter (2014) found no significant difference between four ability levels (Below Basic, Basic, Proficient, and Advanced Proficient) when she compared scores in each ability category at each grade level. The data analysis did not account for students' prior ability; it just compared the results of students in the four ability levels between grouped and non-grouped reading classes. However, McCarter's study does support previous findings from Macqueen (2010), Matthew, Ritchotte and McBee (2103), and Butler (1987).

The National Center for Educational Statistics administered the Early Childhood Longitudinal Study (ECLS), which gathers data from a nationally representative sample

of students who started in kindergarten in the 1998-1999 school year. The students came from both private and public school and their teachers, parents, and principals were surveyed during the study as well. This K cohort was then followed until they reached Grade 8 in the spring of 2007 (National Center for Educational Statistics, 2016). This large body of data has been analyzed by researchers such as McCoach et al. (2006), Condrón (2005), Nomi (2006), McCaw (2001), and Lleras and Rangel (2009) in the hopes of gaining insight into the effectiveness of grouping. Condrón (2005) examined the ECLS data focusing on the reading achievement of students grouped for reading in grade one and then heterogeneously grouped for the following grade. Condrón had a sample size of 1,909 students using the ECLS data. Condrón did not find a significant difference between the ungrouped students one year later when using grouping as a single variable. However, when groups were divided into high, medium, and low ability groups, the results showed that advanced students scored +1.25 more points, $p < .01$, average students did not score significantly different between grouped and ungrouped, and the lowest ability group scored significantly worse in grouped classes (effect size of - 2.73, $p < .05$). Nomi (2005) also used the data from the ECLS, but she focused her study on kindergarteners and first graders; and like Condrón (2005), she divided her sample into three groups of abilities: high, medium, and low. Nomi found no significant difference in any group classification in kindergarten or first grade, even for the highest ability groups. Interestingly, McCaw, Davis-Lenski, and Braun (2001) also did not find any difference in comprehension with first graders who were grouped and those first graders who received whole group instruction when examining data from the ECLS.

Lleras and Rangel (2009) examined the ECLS as well, but their exploration of the data was focused on the effects of grouping on African Americans and Hispanic students. Lleras and Rangel (2009) examined 886 African American and 750 Hispanic students who were grouped for reading until the third grade. Lleras and Rangel compared the end-of-year reading assessment given at the end of first and third grade. They divided students into just two groups, high and low, and found grouping was harmful to African Americans (effect size -4.27 , $p=.01$) and Hispanics (effect size -2.83 , $p=.05$) in lowest groups in first and third grade. Conversely, African Americans (effect size $+2.25$, $p=.05$) and Hispanics (effect size $+3.87$, $p=.001$) in the high group performed better in groups than their ungrouped peers. It is noted that Lleras and Rangel divided their population into two subgroups, whereas many other studies examining ability levels in this paper have divided students into three ability groups. Interestingly, the three researchers who used the ECLS data focusing on grouping did not find conclusive evidence in favor of grouping even when examining the data by ability groups (Condron, 2005; Lleras & Rangel, 2009; Nomi, 2006).

The research on the influence of ability grouping on student achievement when considering initial ability level (high, medium, or low) has been examined across many decades and in many grades. There is now a large body of evidence that suggests the higher achieving students do better or the same in homogeneous groups, while the most at-risk learners do better in heterogeneous groups. It is the average learner who can adapt to either grouping formation and not be significantly impacted.

Third Lens: Comparing Within-Class Grouping with Whole Group Instruction

There are teachers who divide their students into groups for reading based on ability level or needed skill. These teachers may use guided reading to teach students the skills needed to continue to grow as readers (Ford & Opitz, 2011; Fountas & Pinnell, 1996). Other teachers may teach students in a whole group setting. In this formation students are reading the same text, receiving instruction as a group by the teacher and are expected to maintain pace with the teacher. This section examines if students grouped by ability within the classroom perform significantly different than those students receiving whole group instruction.

Kulik and Kulik (1987) and Slavin (1987) included studies in their meta-analysis which used flexible, within-class grouping. Kulik and Kulik found an effect size of +.12 for within-class flexible skills grouping and Slavin (1987) found that when he examined the seven elementary schools who used within-class grouping, the effect size increased for low (+.65), medium (+.27), and high learners (+.41). Lou et al.'s (1996) meta-analysis found a small, but significant effect size (+.16) associated with using within-class ability groups as compared to whole group instruction. McCarter's (2014) study in Tennessee, which was discussed at length in the previous section, found no significant difference between the results of schools that grouped and those that did not use the practice.

Jecks (2011) also wanted to examine previous studies, which involved within-class grouping compared to whole group instruction. Jecks' meta-analysis of 11 research findings found flexible grouping yielded a small to moderate (effect size +.37) effect on reading achievement at the elementary level when using whole class instruction 75% of

the day and homogeneous grouping the other 25% of the day. Jecks found the effect size for fixed homogeneous grouping to be + .14. Jecks also examined the studies to determine if grade level had any bearing on the effect size. Jecks found a significant effect size (+.23) for flexible grouping at the primary level and effect size (+. 39) for those at the intermediate grade level of students using flexible grouping for reading. According to Jecks, flexible grouping for elementary reading had a small to moderate effect on reading achievement. In comparison to Jecks (2011), Puzio and Colby (2010) examined the reading achievement of 5,410 participants in 15 studies that examined within-class ability grouping. Puzio and Colby (2010) found students in ability groups had a significantly higher reading achievement level, (effect size of +.22, $p=.002$) than those receiving whole group instruction within the general classroom. This was the equivalent of approximately a half-year's growth (Puzio & Colby, 2010).

The ECLS data sample was used to study grouping within the classroom. McCoach et al. (2006) confined their ECLS study to a sample of kindergarteners in 620 schools and examined the use of within-class ability grouping compared to those using whole class instruction. McCoach et al. compared students' results on printed word recognition, sound identification, word reading, and reading comprehension from the fall to the spring of their first year in kindergarten. McCoach et al. found the mean gain on the item response theory to be 10.2, $SD=6.15$ for all kindergarten students. Students in the classes, which used ability grouping at least once a week increased their item response theory by 1.5 points, effect size + .377. This was significant at the $p=.05$ level, but the authors caution this study is limited by its lack of causation.

International researchers Duflo, Dupas, and Kremer (2008) studied grouping in Kenya with 5,796 students. There were 61 randomly selected first grades who were provided another teacher to split by ability and 60 classrooms who had first graders randomly assigned to a classroom taught in a whole group format. Duflo et al. (2008) found high and low levels of students benefited from being grouped by ability. They also found there was no significant difference for the medium ability grouped students in either the grouped or ungrouped first grade classroom. In examining the overall data for grouped and ungrouped students, grouped students scored +.16 SD higher than non-grouped students, and the benefits continued the following year when the students were retested at the end of second grade. The students who were grouped for reading in first grade scored +.14 SD higher in second grade than those not grouped in first grade. Evidence on the results of using within-class grouping provides some support for its usage. The next hypothesis question would be, “Does the use of within-class grouping *between* the same grade level lead to similar results as the use of within-class grouping within the homeroom?” The last lens of which grouping research can be examined addresses this last hypothesis.

Fourth Lens: Comparing Flexible and Whole Group Instruction

While the studies discussed have been focused on fixed ability group placement versus heterogeneous class performance or have compared within-class ability grouping to those classroom that do not group, there was a limited number of studies done by researchers such as Haghghat (2009) and Lofton (2013) that specifically examined between-class flexible grouping at the elementary level. Haghghat compared the results of the Arizona state assessment in four schools serving a similar population in the same

school district. One school, Oasis, selected to use between-class grouping for its improvement plan, while the other three schools continued to use whole group instruction. Haighat examined the data from three years and found no significant difference between the schools using flexible grouping when compared to the control schools. Haighat postulated that many Oasis teachers did not want to implement the program and over the course of the study there was a 60% teacher turnover rate leading to more inexperienced teachers being hired. Interestingly, when using the NCLB rating system from the Arizona Department of Education, Oasis made AYP its first year using the flexible grouping and by the third year of implementation the school had a “performing plus” label. This study did not discern if the control groups received any state accolades.

Lofton (2013) also examined skills grouping at the elementary level in a school in Tennessee. Lofton compared one school that used skills-based grouping for 90 minutes per week with its third graders and compared the results on the same assessment to three other similar schools’ third grade results. Even though this study conducted had a different design, Lofton did find skills grouping students performed better in reading, but not significantly. This supported Haghhat’s (2009) finding as well, but this limited number of studies on flexible/skills grouping and its limitations to specific populations makes it challenging to draw a conclusion from which to design programs and policies.

Theoretical Framework

This study examines grouping through Lev Vygotsky’s (1978) zone of proximal development theory. Vygotsky theorized that students were able to accomplish many tasks independently and this was related to a student’s developmental age. Students that spent time working with tasks they had already learned were staying at the current developmental level and not moving forward in their development. However, Vygotsky

believed if a student was guided and supported to learn new information, the student would be making more progress in reaching the next developmental level. He referred to the difference between what a child could complete independently and what a child could do with guided practice as the zone of proximal development.

Vygotsky theory is applicable to the classroom and relates to the previous discussion of teachers' justification of grouping. Teachers realize a student who is attending a lesson on skills or content already mastered is not extending his/her learning. For example, some students can read an early reader with ease, while another student is struggling on the first couple of pages. The student who can decode the book could be working on comprehension or prediction, while the beginner reader could be practicing sight words or phonics.

Petrilli (2011) found a three-to-six-year difference in reading levels between students taking the NAEP. If one applies the zone of proximal development theory to the wide range of abilities within the heterogeneous classroom, it may be argued that differentiated lessons based on a student's developmental age and their zone of proximal development ought to be designed. One reading lesson delivered to all classroom learners will stagnate the learning process of some students. Grouping may be the strategy that allows the teacher to design a lesson for the various levels within the Zones of Proximal Development that occur in the classroom and move students toward their next developmental level. However, this has yet to be examined empirically.

To date most studies have examined the topic of grouping through theories of equity. This study examines the topic through Vygotsky's (1978) zone of proximal development theory. Vygotsky (1978) suggests that students are best supported when

they are presented new information and skills and are “guided” through the learning process. This study examines the instructional technique of targeted grouping, which has differing curriculum, materials, and instructional strategies to work within the students’ zone of proximal development in order to determine if this approach to grouping has a greater influence on student learning than other forms of grouping.

Conclusion

The use of grouping is not a novel idea to the 21st century, nor is it an instructional practice universally endorsed by researchers from a philosophical or empirical standpoint. There have been times in history where the instructional practice of grouping has ebbed and flowed because of classroom management, accountability, immigration, and the equality movement, but the practice has remained. Considering the use of grouping has increased in the last decade (Loveless, 2013), it is a topic worth revisiting.

There are strong empirical and philosophical arguments for and against grouping that have caused the grouping debate to continue over decades. Each side has research to support the claims and merits of their belief and this inconclusive evidence will continue to be debated until modern studies can provide more consistent findings.

The next step in this examination of grouping practices must be to explore explanations for why the data have been so inconclusive. A review of the research conducted to date has established that there is a need for not only additional research but research that examines the problem through another lens and discusses the clarification in study descriptors that may assist further research.

Besides examining the grouping topic through a new theory, the studies conducted to date must be examined critically in order to continue to advance the

research on the topic. Nomi (2006) and Slavin (1987) have postulated that the inconclusive data may be caused by a lack of description on the grouping practices. For example, certain studies did not explain if the grouping was permanent, fluid, or if curriculum was differentiated for each grouping level. The grouping studies have also not discussed the specific materials and instructional strategies (Gamoran, 2009; Gentry, 2016) used in the study, and not all studies clearly defined how students were divided into ability groups. This study hopes to compare grouping practices in reading using a specific reading practice, guided reading, and provide a rich descriptor of the grouping practices used and their relative influence on student learning gains.

CHAPTER III

METHODOLOGY

Introduction

The purpose for this correlational, explanatory, cross-sectional study was to explain the influence of the type of grouping (within-class or between-class) on the 2013 third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey. The results from this study explained the amount of variance in the 2013 NJASK 3 reading rates accounted for by grouping type, student gender, race/ethnicity, and prior reading ability. The extant research to date has focused almost exclusively on fixed grouping between a grade level; there has been limited analysis on flexible grouping between the same grade level. The goal of this study was to provide new empirical evidence of the relationship between grouping strategies and reading achievement. Districts and policymakers may use this research to make better-informed decisions concerning the use of reading groups in the elementary grades.

Research Design

This non-experimental, correlational, cross-sectional study explained the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey that had similar students in two different elementary schools, while controlling for gender, prior reading achievement, and racial/ethnic background using a simultaneous regression analysis. According to Gay et al. (2011), multiple regression allows the researcher to determine if variables are correlated and the degree of that correlation (p. 361). Gay et al. (2011) add multiple regression, which also allows the

researcher to determine which independent variables, are the best predictors of the criterion variable (p. 362).

In order to best analyze the data from this non-experimental, quantitative study, simultaneous multiple regression and Pearson correlations were conducted to investigate the relationships, if any, that existed between grouping, student characteristic variables, and reading achievement on the NJASK 3. Scores for all variables were obtained for each student in the study, and these scores were correlated with the results of the NJASK 3 test. The correlations indicated the nature and extent of the relationship between two variables (Hinkle, Wiersma, & Jurs, 2003).

This study used simultaneous multiple regression models to determine which student variables (gender, race/ethnicity, and prior ability) and which school variable (grouping type) were statistically significant in predicting reading achievement on the NJASK 3. This study explained the amount of variance in the dependent variable, reading achievement, that can be explained by the grouping and student-related predictor variables.

Research Questions

1. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender?
2. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender and race/ethnicity?

3. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender, race/ethnicity, and prior reading ability?

Sample Population/Data Source

The suburban district selected for the study is a K-12 school system with five schools. There is a high school that serves approximately 750 students and four elementary schools that serve approximately 1,700 students. Two elementary schools are structured K-3, and the other two elementary schools are structured Grades 4-8. The New Jersey Department of Education (DOE) identified this district as a member of GH Factor group. New Jersey DOE categorizes a district's socioeconomic status on an eight-category scale of A-J with A being the poorest and J the wealthiest (NJDOE, 2015).

The primary elementary schools within the district have a similar student population when judged by race, percentage of students eligible for free/reduced-price lunch, gender, limited English proficiency, and special education enrollment status. In addition, the district's teachers responsible for teaching reading to the third graders during the 2012-2013 school year have a mean of 15.8 years of teaching experience. See Table 1.

Table 1

2012-2013 District Third Grade Student Demographics

	FREQUENCY	GENDER		RACE/ETHNICITY		Socio-Economic		Special Ed	LEP	TEACHER EXPERIENCE
		Female	Male	Caucasian	Minority	Free Lunch	Reduced Lunch			
Control Group	97	47	50	57	40	4	1	13	0	12.42
Treatment Group	90	37	53	64	26	4	2	13	0	19.21
TOTAL	187	84	103	121	66	8	3	26	0	15.82

In the spring of 2012, the superintendent of the district in which the study took place suggested grouping students in first grade for reading across the same grade level.

Both elementary schools in the study adopted the suggested first grade between grade level reading program, but the treatment school decided to use between same grade level reading with second and third graders. The differences were the treatment school formed ability-based reading groups amongst students in the entire third grade, whereas the control school formed ability groups amongst the students in each homeroom teacher's assigned self-contained classroom. The treatment and control schools both changed group assignments as determined by the teacher(s). The school in the control group used guided reading group instruction within the heterogeneous classrooms in Grades 2 and 3, whereas the treatment school piloted the use of flexible, between same grade level ability reading groups for guided reading instruction in Grades 2 and 3.

Teachers formed the third grade reading groups in both the control and treatment schools based on results from multiple measures: NJPASS, DRA2, teacher observations, classwork, and/or prior grade performance. The NJPASS, published by Houghton, Mifflin, Harcourt, measured second grade reading skills under two main categories, "Working With the Passage" and "Analyzing the Passage." Houghton, Mifflin, Harcourt publishers noted the following skills were measured:

Working with the Passage included comprehending the passage, such as recognizing the main idea and supporting details, getting information, paraphrasing meaning, and recognizing organization of text and purpose of reading. This comprised sixteen (61.5%) of the twenty-six points in reading.

Analyzing the Passage included analyzing and evaluating the passage by using skills such as asking questions, predicting meaning, developing opinions, drawing conclusions, and interpreting conventions of print. This comprised ten (38.5%) of

the twenty-six points in reading. (New Jersey Proficiency of State Standards, 2015).

As per the direction of the superintendent, one reading specialist assessed all third grade students in the control and treatment schools on the DRA2 in the beginning of the school year for a district analysis of both reading group formations. The DRA2 is a reading leveling system published by Pearson, which uses a rubric to guide an examiner in identifying a student's instructional and independent reading levels based on reading engagement, oral reading skills, and comprehension (Pearson, 2011). For example, the DRA2 level 28, a third grade fictional reading level, assesses reading engagement, oral reading fluency, and comprehension. Comprehension encompasses predictions, summaries, vocabulary, literal comprehension, interpretation and reflection (Pearson, 2011). In addition to the NJPASS and DRA2 results, teachers considered a student's class work and prior performance in second grade. Teachers examined a student's written responses, contributions to class discussions, oral reading skills, prior report cards, and/or feedback from teachers who had worked with the student.

Although both the control and treatment groups used the DRA2, teacher observations, previous performance, and classwork to form reading groups, the groups differed in formation and group placement process. The teachers from the control group formed groups within the heterogeneous classroom based on the classroom teacher's professional judgment on students' performance on NJPASS, DRA2, classwork, class ability, and group size. Teachers from the treatment group formed their reading groups between the same grade level. Three reading specialists in the treatment school examined all the DRA2 scores reported for Grade 3 and divided the entire grade level into flexible,

within same grade ability-based reading groups, including special education students. The groups were examined and third graders, whose placement was questioned by either school administration or homeroom teachers, were then re-examined.

In the treatment school two reading specialists and four regular education teachers instructed the third grade in flexible, between-class ability-based reading groups. The reading specialists taught the most at-risk reading learners and had smaller class size. The control group had four classroom teachers and one special education teacher work with classified students. The special education students in the control school had a smaller class size. The control group and treatment group used the same leveled reading series and received the same exact training from the publisher's trainers. The teachers in the control and treatment group used the same-leveled reading series by Pearson Corporation, *Good Habits Great Readers*. The district adopted the series in June of 2012 so all teachers in the control and treatment group had the same prior knowledge of *Good Habits, Great Readers* and attended the same two-day training session. Finally, both treatment and control schools also had an 84 minutes language arts block each day.

The third grade classes at both the control and treatment schools were selected for inclusion in the study. Students in the sample met the following criteria:

1. Assessed on the Language Arts New Jersey Proficiency Assessment of State Standards in the spring of 2012.
2. Enrolled in third grade on or before October 15, 2012.
3. Enrolled in third grade on or before June 1, 2013.
4. Received reading instruction in the general education classroom.

5. Assessed on the *New Jersey Assessment of Skills and Knowledge* in Language Arts in Grade 3 in the spring of 2013.
6. General education student⁴

There were 77 students in the control group and 78 students in the treatment group for a total of 155 students who qualified for the study. The NJDOE collected data on each student enrolled in the study on October 15, 2012, noting race, gender, special education classification, 504 status, economic status based on free/reduced-price lunch qualifications, migrants, and homeless status. These data were collected for all students in the district and used in the study to assign characteristic variables. Descriptive statistics for students in the sample are included in the table below. (See Table 2) In order to determine the influence of grouping on reading achievement between the treatment and control school, a quantitative, correlational, explanatory study was performed.

Table 2

Student Demographics For Control and Treatment Groups

	FREQUENCY	GENDER		RACE/ETHNICITY		Socio-Economic	
		Female	Male	Caucasian	Minority	Free Lunch	Reduced Lunch
Control Group	77	41	36	48	29	3	0
Treatment Group	78	32	46	54	24	1	1
TOTAL	155	73	82	102	53	4	1

⁴ Students classified “Speech only” were considered part of the general education classroom.

Instruments

In order to collect baseline data, each student's score on the previous year's NJPASS was examined. Riverside Publishing, a division of Houghton Mifflin Harcourt, scored these assessments. Students who scored between 0-20 were Basic, students who scored 21-28 were Proficient, and students who scored 29-42 were Advanced Proficient.

The NJASK 3 was used as the dependent variable to assess students' performance at the end of third grade. This assessment also measured students' achievement for the federal NCLB requirement. Measurement, Incorporated scored the NJASK 3 tests for the state of New Jersey. Students who scored at or between 150-199 were Partially Proficient, students who scored at or between 200-249 were Proficient and students who scored at or between 250-300 were Advanced Proficient. The NJASK 3 was a criterion-referenced test created by Measurement Incorporated. Test Specifications from NJASK 3 from spring of 2013 identified the skills students needed to be Proficient in reading. The test specifications state the following:

A student performing at the Proficient level demonstrates the ability to employ strategies to comprehend a variety of texts literally and inferentially and to express understanding of the text in written responses. As a proficient reader, the student recognizes the central idea, supporting details, purpose, and organization of the text as well as some literary devices. The proficient student can make connections to the text, form opinions, and draw conclusions. The proficient reader is able to synthesize ideas from the reading and to use these to analyze and extend the meaning of the text in written responses (NJASK 2013 Score Interpretation Manual, Grades 3-8, 2013).

The NJDOE NJASK 2013 Test Specification manual provided percentages of points earned in each skill. The NJASK 3 reading section accounted for 60% of English Language Arts score with 40% of the English Language Arts score dedicated to informational text items and 20% testing literature items. The literature section contained six multiple choice items and one open-ended item, while the informational section comprised 12 multiple choice items and two open-ended items. (NJASK 2013 Score Interpretation Manual Grades 3-8, 2013).

Data Collection

The researcher requested information on the 2012-2013 third graders in the district. The researcher requested the superintendent provide each third grader's demographic information: home school, gender, free/reduced-price lunch status, race, limited English proficiency status (LEP), and special education code. In addition, the researcher requested NJPASS 2 and NJASK 3 Language Arts scores. The free/reduced-price lunch status was determined by the U.S. Department of Agriculture with eligibility requirements determined July 1, 2012, based on information collected by the school district and reported by parents. The special education status was determined by the New Jersey Administrative Code 6A:14, Special Education and race/ethnicity was determined by parental choice on the student's school registration form. Parents selected from the following list of races determined by the NJDOE: American/Native Alaskan, Asian, Black, Hawaiian/Other Pacific Islander, Hispanic, and/or White. Parents who selected more than one race affiliation were considered Multiracial. Last, LEP status was based on student's eligibility for English as a Second Language classes.

The NJPASS 2 reported scores in listening, writing, reading, and an overall performance score and level. The district was requested to provide the following scores from the NJPASS 2 for each student: Reading-Working with the Passage, Reading-Analyzing the Passage, and overall performance score. The NJASK 3 reported scores in writing, reading, and an overall performance score and level. The district was requested to provide the following scores from the NJASK 3 for each student: Reading-Literature, Reading-Informational Text, and overall performance score.

Finally, one reading specialist gave the DRA2 to students in September/October 2012 and again in May/June 2013. The district was asked to provide the instructional reading level for each subject during these assessment periods.

Units of Analysis

This study utilized student level data from suburban third graders in a public district in New Jersey. Students without data for each variable and those students receiving instruction in the special education classroom were excluded from the study. The sample size used in this study met the recommendations by Green (1991) for determining minimal acceptable sample size. Green recommends a minimum sample size of $50 + 8k$; k represents the number of predictors. In this study there were four predictor variables. When Green's formula is applied, $(50 + 32 = 82)$, this study meets the requirements for sample size, as there were 155 students in the study.

Analysis Construct (Model Specification)

The dependent variable examined was the third grade NJASK 3 scores. The dependent variable was regressed on the following variables: racial demographics, gender, prior ability, and grouping status. See Table 3 for a description of the each variable

Table 3

Descriptions of Variables

VARIABLE	DEFINITION	LEVEL of MEASUREMENT	STATUS
NJASK 3 SCORE	The score on the language arts assessment given by the State of New Jersey to third graders.	Ratio	Dependent Variable
GENDER	Male or female status as selected by student's parent/guardian.	Categorical (Dummy Variable)	Control Variable
RACE/ETHNICITY	Black, Asian, Hispanic, White or Multiracial as selected by student's parents/guardian.	Categorical (Dummy Variable)	Control Variable
ABILITY	A student's reading score on the NJPASS 2.	Ratio	Control Variable
GROUPING STATUS	Whether a student received targeted reading instruction within the classroom or between the same grade level	Categorical (Dummy Variable)	Independent Variable

Data Analysis

Descriptive statistics were run for all variables. The descriptive statistics provided information on the mean, standard deviation, and range for each ratio variable. In terms of bivariate analysis, correlation between each predictor variable and the dependent variable was conducted. This determined the relationship, if any, and the strength of relationship between the variables (Gay, Mills, & Airasian, 2011). An independent sample *t*-test was conducted to determine the similarity of the independent variable, prior ability. An independent *t*-test determines the difference between two sets of independent group scores by comparing the actual difference of each group's mean scores with the difference in mean scores expected by chance (Gay et al., 2011, p. 351).

The simultaneous regression showed the influence of a predictor variable on NJASK 3 scores, the level of significance, the relationship between the variable and the percentage of variance in the dependent variable that is explained by the independent variable. Scatter plots were then used to show the dot cluster of the two variables. The slope of the dot cluster indicated if there was a positive, negative, or little to no relationship between variables (Witte & Witte, 2010).

A multiple regression was employed to conduct the multivariate analysis. Multiple regression equations allowed the researcher to observe the influence of several predictor variables on the dependent variable, NJASK. In this study gender, race/ethnicity, ability level, and grouping status served as the predictor variables. This model provided data as to how much of the variance in the reading achievement ratio could be explained by the multiple predictor variables. Variable coefficients were examined to determine the direction and strength of any possible relationships between the predictor variables and the dependent variable.

The 5% threshold was used to determine statistical significance of the predictor variable coefficients. In other words, only predictor variables that have t-statistics +/- 1.96 and *p*-values of .05 or less were considered statistically significant. The *F*-statistic was reported so as to judge the statistical significance of the entire model, again using an .05 level as the threshold.

CHAPTER IV

ANALYSIS OF THE DATA

Introduction

The purpose of this correlational, explanatory, cross-sectional study was to explore the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores of students in one suburban district in New Jersey. This study examined the amount of variance in the NJASK 3 reading rates that could be accounted for by the grouping type, student gender, race/ethnicity as indicated by the percentage of Asian, Hispanic, Black, Pacific Islander, Native American, and Multiracial students, and prior reading ability. The sample consisted of 155 third graders in one suburban district in New Jersey.

The goal of this study was to provide empirical evidence on the relationship between grouping formation and reading achievement. As per the requirements of NCLB and its revised version, ESSA, reading achievement must be assessed in Grades 3 through 8 (Every Child Succeeds Act, 2015). Educators and policymakers must consider instructional strategies that may affect reading achievement and improve the effectiveness of instruction. According to Loveless (2013), principals and other school personnel have been increasing their adoption of grouping strategies. However, there is limited empirical evidence about whether grouping between the grades or within the general classroom has a greater influence on reading achievement (Haghighat, 2009). This study was conducted to provide research-based evidence about different approaches to group formation, which might assist policymakers, educational leaders, and teachers

when designing reading programs and implementing instructional strategies to improve reading achievement.

Predictor Variables

Four predictor variables, gender, race/ethnicity, grouping status and prior ability, were included in the analysis. There was one dependent variable, New Jersey Assessment of Skills and Knowledge Grade 3 Reading Scores (NJASK 3). Variables and descriptors are presented in Table 4.

Table 4

Variable Names and Descriptors

Variable	Label	Level of Measurement	Description
Gender	0=Female 1=Male	Categorical (Dummy Variable)	The sex identified by guardians on the student's school registration form.
Race/ Ethnicity	0= Caucasian 1= Minority	Categorical (Dummy Variable)	The Race/Ethnicity identified by guardians on the student's school registration form. Minority students comprise those students identified as Black, Hispanic, Asian, or multi-racial.
Grouping Status	0=Not in Program 1=In Program	Categorical (Dummy Variable)	The Control group participated in within-class grouping and the Treatment group participated in between grade level grouping.
Prior Ability	NJPASS 2	Ratio	The student's score on the NJPASS 2 reading section taken in Grade 2.
New Jersey Assessment of Skills and Knowledge Grade 3	NJASK 3	Ratio	The student's score on the NJASK 3 reading section taken in Grade 3.

Bivariate Analysis of Descriptive Statistics

The New Jersey Department of Education requires each school district in the state of New Jersey to capture demographic statistics on October 15 of each school year. The district used for this study was asked to share its descriptive data from October 15, 2012. In addition to the data reported to the NJDOE on October 15, 2012, the district released students' testing data from the NJPASS 2 and NJASK 3. Table 5 provides descriptive, crosstabulation statistic profiles for the categorical variables.

Table 5

Descriptive Statistics for Group Participation, Gender, and Race/Ethnicity

	FREQUENCY	GENDER		RACE/ETHNICITY	
		Female	Male	Caucasian	Minority
Control Group	77	41	36	48	29
Treatment Group	78	32	46	54	24
TOTAL	155	73	82	102	53

There was a total of 155 students in the sample with 77 students in the control program and 78 students in the treatment program. Within the sample 47% of the subjects were classified as female and 53% classified as male. In addition, the sample was 66% 'Caucasian' and 34% minority. Minority students included students who were identified as Black, Hispanic, Asian, or Multiracial during the school registration process.

The control group consisted of 77 students, composed of 41 females (53%) and 36 males (47%). The control group had 48 Caucasian students (62%) and 29 minority students (38%). The treatment group consisted of 78 students, composed of 32 females (41%) and 46 males (59%). The treatment group had 54 Caucasian students (69%) and 24 minority students (31%). Minority students included students who were identified as

Black, Hispanic, Asian, or multi-racial during the school registration process (See Table 5).

The NJPASS 2 and NJASK 3 scores were also analyzed. The NJPASS 2 had 0-26 point scoring point range in reading and the NJASK 3 had scoring point range of 0-30 in reading. All subjects in the study had a mean of 17.50 on the NJPASS 2 with a standard deviation of 4.4. There was a minimum score of 6 and a maximum score of 26 on the NJPASS 2. All subjects on the NJASK 3 had a mean of 18.33 and a standard deviation of 3.94. There was a minimum score of eight and a maximum score of thirty-one on the NJASK 3 (See Table 6).

Table 6

NJPASS 2 and NJASK 3 Descriptive Statistics

Descriptive Statistics					
	<i>N</i>	Minimum	Maximum	Mean	Std. Deviation
NJPASS	155	6.00	26.00	17.5097	4.40243
NJASK	155	8.00	31.00	18.3355	3.94817
Valid <i>N</i> (listwise)	155				

When examining the NJPASS 2 scores by group, the control group had a mean of 16.85 with a standard deviation of 4.82. The treatment group had a mean of 18.15 on the NJPASS 2 with a standard deviation of 3.86 (See Table 8). An independent sample *t*-test showed that the treatment group was not significantly different than the control group on the NJPASS. Inspection of the two groups' means indicated that the average NJPASS score for the control group (M=16.85) is not significantly lower than the treatment group (M=18.15). The difference between the means is 1.30 points on a 26-point test. The effect size *d* is approximately .05, which is a typical size for effects in the behavioral sciences.

The treatment group did not differ significantly than the control group on the NJPASS ($p = .067$). The effect size, d , is approximately .05 (See Table 7).

Table 7

Independent Sample Test

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
NJPASS	Equal variances assumed	5.630	.019	1.848	153	.067	1.29670	.70176	-.08968	2.6830
	Equal variances not assumed			1.845	145.161	.067	1.29670	.70276	-.09226	2.6856

When examining the NJASK 3 scores by group, the control group had a mean of 17.63 with a standard deviation of 4.04. The treatment group had a mean of 19.02 on the NJASK 3 with a standard deviation of 3.74. Table 8 provides bivariate descriptive statistics for the NJPASS 2 and NJASK 3 variables.

Table 8

Grouping Status and NJPASS/NJASK 3 Crosstabulation

	GROUP	N	MEAN	STD DEVIATION	STD ERROR OF MEAN
NJPASS	Treatment Group	78	18.1538	3.86149	0.43723
	Control Group	77	16.8571	4.82786	0.55019
	Total/Average	155	17.5097	4.40243	0.35361
NJASK	Treatment Group	78	19.0256	3.74850	0.42443
	Control Group	77	17.6364	4.04549	0.46103
	Total/Average	155	18.3310	3.94817	0.44273

Research Questions

The overarching research question that was answered is as follows: What is the influence of the grouping type on reading achievement? The sub-questions below also guided analysis:

1. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender?
2. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender and race/ethnicity?
3. What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender, race/ethnicity, and prior reading ability?

Hypothesis

Null Hypothesis: There is no statistically significant difference in the reading achievement of third graders assigned to between the grade guided reading groups and similar third graders participating in within-classroom guided reading groups.

Results

The researcher sought to examine the influence of four predictor variables: gender, race/ethnicity, grouping status, and prior ability on the NJASK 3 reading scores.

Preliminary analyses were conducted to check for relationships between the variables. In order to determine if relationships existed between two variables, scatter plots were created and correlation statistics analyzed. Scatter plots are visual displays of data that

show a relationship between two variables (Hinkle et al., 2003). Scatter plots were created for NJASK 3 and NJPASS 2 based on the crosstabulation of gender, race/ethnicity, and grouping status (See Figures 1-3). In addition, the relationship of the reading results on the NJASK 3 and NJPASS were represented in a scatter plot (See Figure 4).

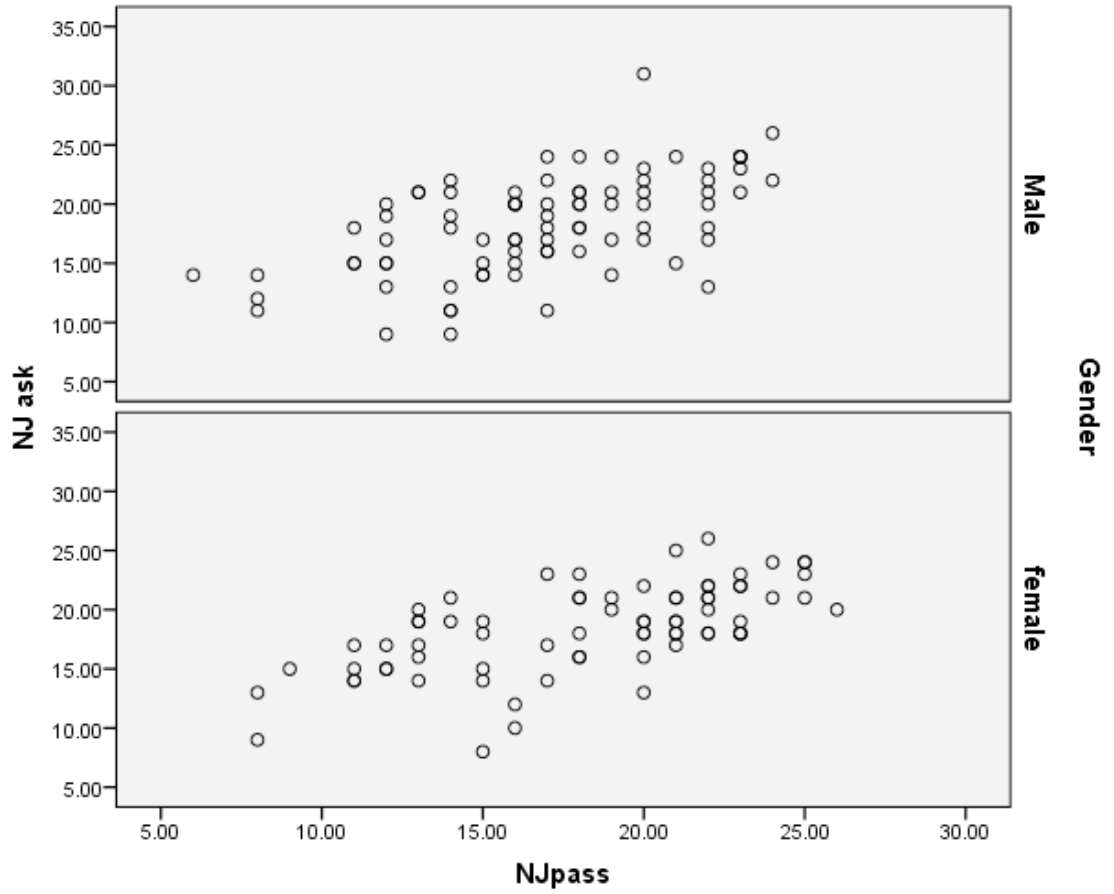


Figure 1. Scatter plot NJASK 3 and NJPASS 2 by gender.

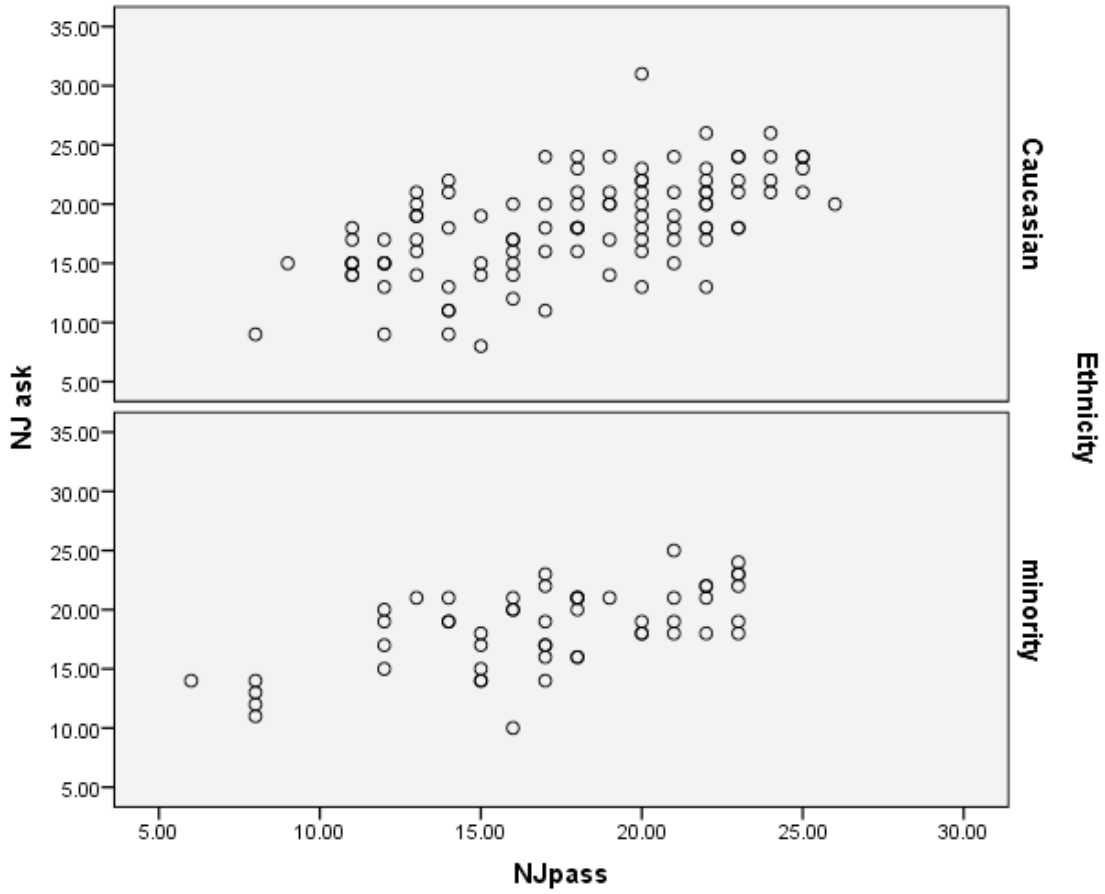


Figure 2. Scatter plot NJASK 3 and NJPASS by race/ethnicity.

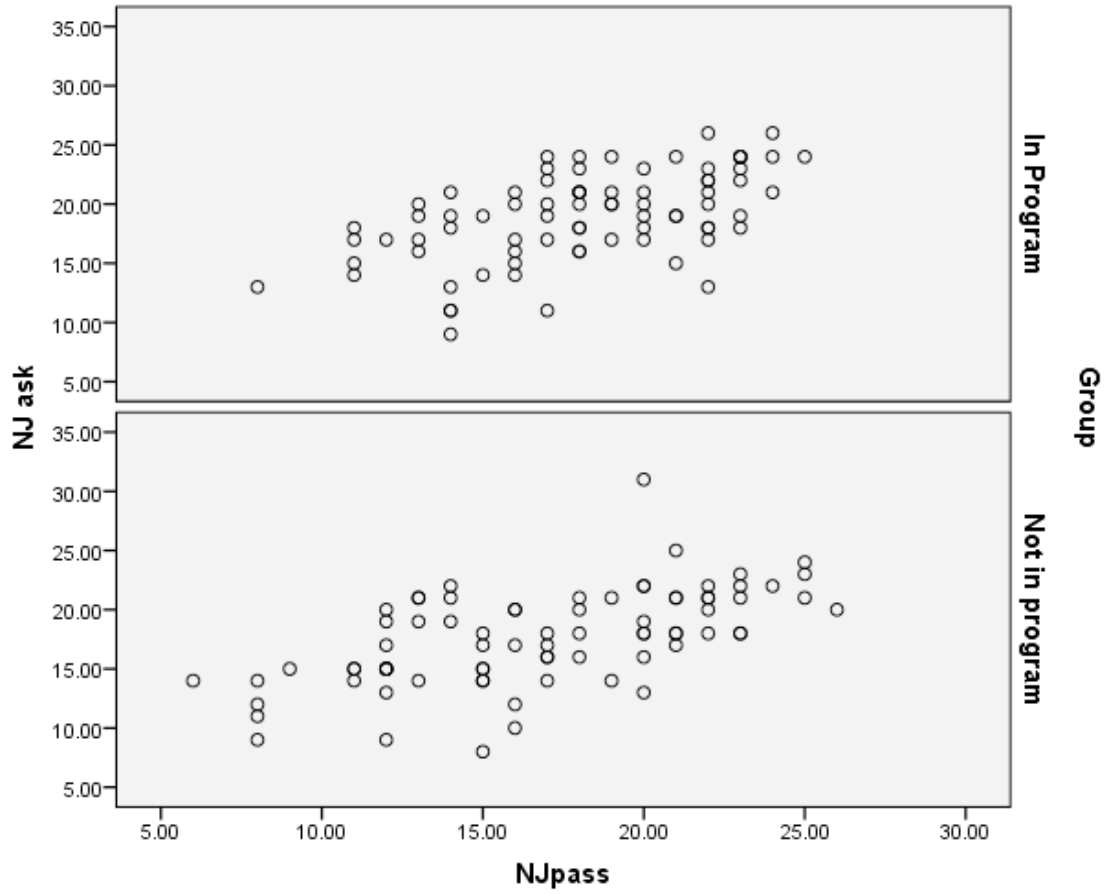
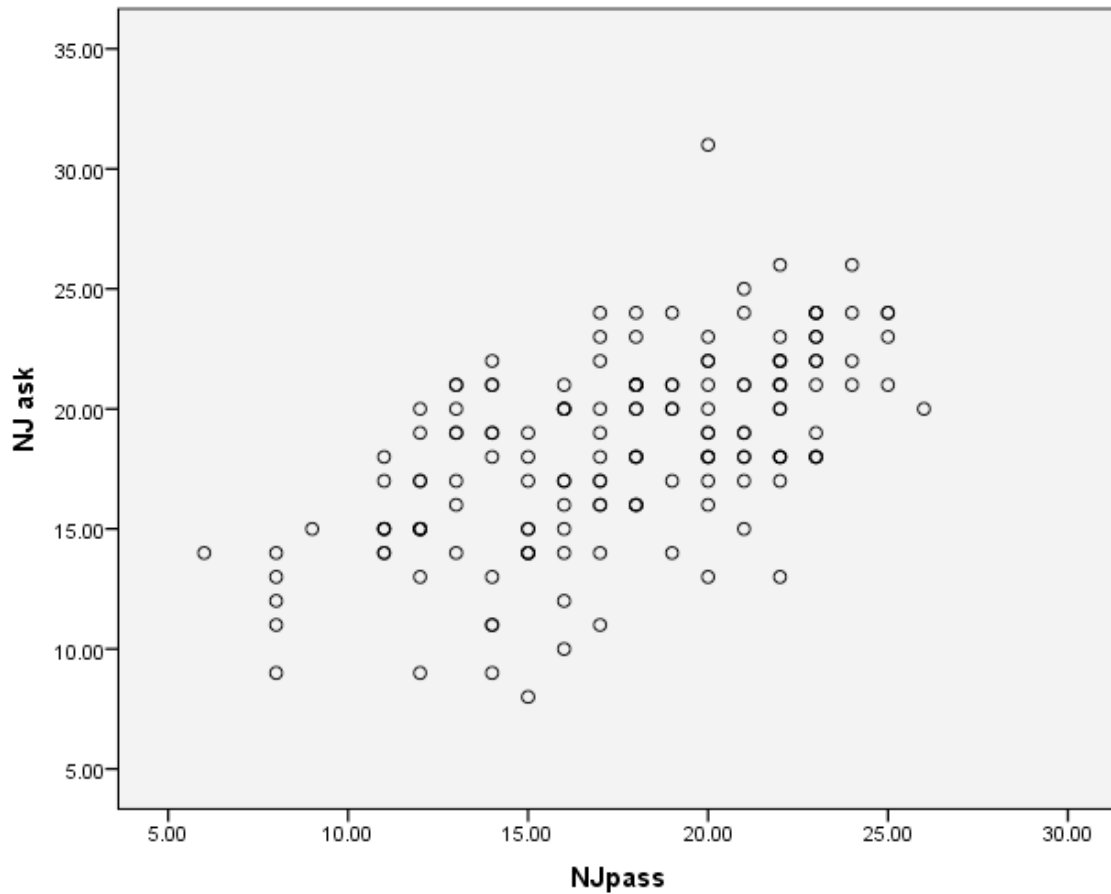


Figure 3. Scatter plot NJASK 3 and NJPASS 2 by grouping status.



$$R^2 = .345$$

Figure 4. Scatter plot NJASK 3 and NJPASS 2.

Although the scatter plots provided a visual representation of the relationship between two variables, a correlation coefficient matrix was included to gather additional, specific correlation information (See Table 9). The Pearson correlation coefficient (r) is used to determine the linear relationship that exists between two variables (Hinkle et al., 2003). The values of the correlation coefficients are between -1 and +1, which indicates a perfectly correlated negative or positive relationship and 0 meaning no correlation. Gender and race/ethnicity had a -.001, indicating little, if any negative correlation and

gender and grouping status had a weak, positive .122 correlation, again indicating little, if any, positive correlation. Gender and the NJPASS had a weak, negative -.164 correlation and gender and NJASK 3 had a -.015 negative correlation, again indicating little, if any, correlation. Race/ethnicity and grouping status had a -.073 correlation, and this indicates little or no negative correlation. Race/ethnicity and NJPASS had a weak, if any, -.099 negative correlation, and race/ethnicity and NJASK 3 had a weak, if any, .021 positive correlation. Grouping status and NJPASS had a weak .148 positive correlation and grouping status and NJASK 3 had a weak positive .177 correlation. NJPASS and NJASK had a .587 correlation. This indicates there was a moderate, positive correlation (Hinkle et al., 2003). The Pearson (r) was analyzed in those models that were significant.

Table 9

Correlation Table

		Correlations				
		Gender	RaceEthnicity	GroupingStatus	NJPASS	NJASK3
Gender	Pearson Correlation	1	-.001	.122	-.164*	-.015
	Sig. (2-tailed)		.990	.129	.041	.855
	N	155	155	155	155	155
RaceEthnicity	Pearson Correlation	-.001	1	-.073	-.099	.021
	Sig. (2-tailed)	.990		.369	.219	.791
	N	155	155	155	155	155
GroupingStatus	Pearson Correlation	.122	-.073	1	.148	.177*
	Sig. (2-tailed)	.129	.369		.067	.028
	N	155	155	155	155	155
NJPASS	Pearson Correlation	-.164*	-.099	.148	1	.587**
	Sig. (2-tailed)	.041	.219	.067		.000
	N	155	155	155	155	155
NJASK3	Pearson Correlation	-.015	.021	.177*	.587**	1
	Sig. (2-tailed)	.855	.791	.028	.000	
	N	155	155	155	155	155

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

The Pearson correlation table shows that there is a positive relationship between the predictor variable NJPASS 2 and the dependent variable NJASK 3, $r(153) = .59$, $p = .000$ (Cohen, 1988). This means that students who had relatively high NJPASS 2 reading scores were likely to have high NJASK 3 reading scores and vice versa. Using Cohen's (1988) guidelines, the effect size is moderate. The R^2 indicates that approximately 34% of the variance in NJASK 3 reading scores can be predicted from the NJPASS 2 scores. The other independent variables—gender, race/ethnicity, and grouping status—had little, if any correlation.

To increase validity, the two-variable correlational results needed to be examined further. If variables are highly correlated, this can lead to some variables appearing to be significant or insignificant when the opposite can be true (Miles, 2014). An analysis of the variance inflation factor (VIF) is one statistical tool to determine if the variables have multicollinearity. Multicollinearity exists when the combination of two or more predictor variables are highly correlated to another predictor variable. In order to determine if predictor variables have multicollinearity, the R -squared value obtained by regressing a predictor on all of the other predictors in the analysis is conducted (Miles & Shelvin, 2001). VIF scores greater than 10 are considered highly correlated (Allison, 2009). The data in Tables 10- 13 (coefficient tables) show the VIF scores are not significant; multicollinearity did not exist.

Table 10

Gender and Combined Predictor Variables Coefficients: Race/Ethnicity, Grouping Status and NJPASS

Coefficients^a

Model		Collinearity Statistics	
		Tolerance	VIF
1	RaceEthnicity	.987	1.013
	GroupingStatus	.975	1.026
	NJPASS	.970	1.031

a. Dependent Variable: Gender

Table 11

Race/Ethnicity and Combined Predictor Variables Coefficients: Grouping Status, NJPASS, and Gender

Coefficients^a

Model		Collinearity Statistics	
		Tolerance	VIF
1	GroupingStatus	.956	1.046
	NJPASS	.944	1.059
	Gender	.951	1.052

a. Dependent Variable: Race/Ethnicity

Table 12

NJPASS and Combined Predictor Variables Coefficients: Grouping Status, Gender and Race/Ethnicity

Coefficients^a

Model		Collinearity Statistics	
		Tolerance	VIF
1	GroupingStatus	.980	1.021
	Gender	.985	1.015
	RaceEthnicity	.995	1.005

a. Dependent Variable: NJPASS

Table 13

Grouping Status and Combined Predictor Variables Coefficients: Gender, Race/Ethnicity and NJPASS

Coefficients^a

Model		Collinearity Statistics	
		Tolerance	VIF
1	Gender	.973	1.028
	RaceEthnicity	.990	1.010
	NJPASS	.963	1.038

a. Dependent Variable: GroupingStatus

Multiple Regression

Once the preliminary descriptive and correlation analyses were conducted, the next step in the testing process was to determine the influence of the four predictor variables: gender, race/ethnicity, grouping status, and prior ability on the NJASK 3 reading scores. Multiple regression was used to determine this influence. Multiple regression assists the researcher in making predications using several independent/criterion variables (Witte & Witte, 2010). Using the Enter or simultaneous regression method, the four predictor variables were entered to examine the combined influence of gender, race/ethnicity, grouping formation, and/or prior ability on the NJASK 3 reading scores. In the first model, the grouping formation variable was entered, and in the second model group formation was examined while controlling for gender. In the third model, grouping was examined while controlling for gender and race/ethnicity. Finally, the fourth model examined grouping while controlling for gender, race/ethnicity, and prior ability (NJPASS). The model summaries tables showed the results of these multiple regression models (See Tables 14-17).

Table 14

Variables Entered/Removed

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Group ^b	.	Enter
2	Gender ^b	.	Enter
3	Ethnicity ^b	.	Enter
4	NJpass ^b	.	Enter

a. Dependent Variable: NJ ask

b. All requested variables entered.

Table 15

Model Summary

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.177 ^a	.031	.025	3.89886
2	.180 ^b	.033	.020	3.90894
3	.184 ^c	.034	.015	3.91942
4	.605 ^d	.366	.349	3.18611

a. Predictors: (Constant), Group

b. Predictors: (Constant), Group, Gender

c. Predictors: (Constant), Group, Gender, Ethnicity

d. Predictors: (Constant), Group, Gender, Ethnicity, NJpass

Table 16

ANOVA

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	74.788	1	74.788	4.920	.028 ^b
	Residual	2325.767	153	15.201		
	Total	2400.555	154			
2	Regression	78.020	2	39.010	2.553	.081 ^c
	Residual	2322.535	152	15.280		
	Total	2400.555	154			
3	Regression	80.911	3	26.970	1.756	.158 ^d
	Residual	2319.644	151	15.362		
	Total	2400.555	154			
4	Regression	877.861	4	219.465	21.619	.000 ^e
	Residual	1522.694	150	10.151		
	Total	2400.555	154			

a. Dependent Variable: NJ ask

b. Predictors: (Constant), Group

c. Predictors: (Constant), Group, Gender

d. Predictors: (Constant), Group, Gender, Ethnicity

e. Predictors: (Constant), Group, Gender, Ethnicity, NJpass

Table 17

Coefficients

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.636	.444		39.693	.000
	Group	1.389	.626	.177	2.218	.028
2	(Constant)	17.773	.535		33.219	.000
	Group	1.425	.633	.181	2.252	.026
	Gender	-.291	.634	-.037	-.460	.646
3	(Constant)	17.665	.591		29.884	.000
	Group	1.445	.636	.184	2.272	.025
	Gender	-.294	.635	-.037	-.462	.645
	Ethnicity	.289	.665	.035	.434	.665
4	(Constant)	8.098	1.182		6.851	.000
	Group	.677	.524	.086	1.291	.199
	Gender	.572	.526	.073	1.088	.278
	Ethnicity	.721	.543	.087	1.327	.186
	NJpass	.534	.060	.595	8.860	.000

a. Dependent Variable: NJ ask

This first regression model seeks to determine the influence of one predictor variable, grouping on the NJASK 3 reading assessment given to third graders. The r^2 of this model is .031, which means 3.1 % of the NJASK 3 reading scores is explained by grouping status. This implies 96.9% of the variance in the dependent variable, NJASK 3 reading scores, is explained by other variables not considered in this regression model. Nevertheless, this first regression model is significant at .028 when $F = 4.9$ and $df = 1, 153$. For the independent variable of grouping the beta is .177, which is significant at the .028 level when $t = 2.21$.

This second regression model seeks to determine the influence of two predictor variables, grouping and gender, on the NJASK 3 reading assessment given to third graders. The r^2 of this model is .033, which means 3.3 % of the NJASK 3 reading scores

is explained by grouping status and gender. This implies 96.7% of the variance in the dependent variable, NJASK 3 reading scores, is explained by other variables not considered in this regression model. This multiple regression model is not significant at .081 when $F = 2.55$ and $df = 2, 152$.

This third regression model seeks to determine the influence of three predictor variables, grouping, gender, and race/ethnicity on the NJASK 3 reading assessment given to third graders. The r^2 of this model is .034, which means 3.4 % of the NJASK 3 reading scores is explained by grouping status, gender, and race/ethnicity. This implies 96.6% of the variance in the dependent variable, NJASK 3 reading scores, is explained by other variables. This multiple regression model is not significant at .158 when $F = 1.75$ and $df = 3, 151$.

The fourth regression model determines the influence of four predictor variables: grouping, gender, race/ethnicity, and prior ability (NJPASS) on the NJASK 3 reading assessment given to third graders. This model summary table revealed that the multiple correlation coefficient (R) was .605 and the Adjusted R^2 was .366. This means approximately 36.6% of the variance in the NJASK 3 score can be predicted from gender, race/ethnicity, grouping status, and NJPASS 2. This also implies 63.4% of the variance in the students' NJASK scores can be explained by other variables not considered in this study. The ANOVA table (See Table 15) shows that $F = 21.619$ and is statistically significant, $p < .000$. This indicates that the predictor variables significantly combine to predict the NJASK 3 reading score. The combination of variables to predict the NJASK 3 scores from the gender, race/ethnicity, and NJPASS 2 scores was statistically significant,

$F(4, 150) = 21.619, p < .000$. However, only one predictor variable was significant contributing to this fourth model.

The t value and significance in the coefficients table (See Table 17) indicates which predictor variables are significantly contributing to the equation for predicting NJASK 3 reading scores. In this fourth model only the NJPASS 2 significantly adds to the prediction when the other three variables are already considered.

The combination of variables was statistically significant, $F(4, 150) = 21.619, p < .000$.

The beta coefficients are presented in Table 17. The students' gender, race/ethnicity, grouping status, and prior ability predicted 36.6% of the variance in the NJASK 3 when all variables were included, but prior ability was the only significant contributing predictor variable in the fourth model ($b = .595, t = 8.860, p < .000$)

Research Questions and Answers

Simultaneous multiple regression was conducted to investigate the best predictors of NJASK 3 reading scores. The regression analysis can be found in Tables 14-17.

Research Question 1: What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender?

When controlling for gender, the coefficients table (See Table 17) provides a t value and significance level to determine whether the variable significantly contributes to the equation for predicting NJASK 3 reading scores. The significance level for grouping formation is .026 when controlling for gender. Grouping was significant at the $p < .05$ level. This means students who were grouped for reading between the grade level scored approximately 1.4 points better than those grouped for reading within the class. The significance level for gender is .646, and this is not statistically significant as it is greater

than the .05 p value. Gender is not a significant predictor of NJASK 3 scores in this model, indicating the variation in achievement between the treatment and control schools could not be explained by gender.

Research Question 2: What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender and race/ethnicity?

When controlling for gender and race/ethnicity the coefficients table (See Table 17) provides a t value and significance level to determine whether the variable significantly contributes to the equation for predicting NJASK 3 reading scores. The significance level for grouping is .025 when controlling for gender and race/ethnicity. Students grouped for reading between the grade level scored approximately 1.4 points better than students grouped within the class. However, the significance level for gender is .645, and this is not statistically significant as it is greater than the .05 p value. Race/ethnicity has a .665 significance level, and this is also greater than the .05 level. Gender and race/ethnicity are not significant predictors of NJASK 3 scores in this model, indicating the variation between group performances could not be explained by race/ethnicity or gender.

Research Question 3: What is the influence of grouping type in English Language Arts instruction on NJASK 3 Language Arts scores when controlling for student gender, race/ethnicity, and prior reading ability?

The combination of variables to predict NJASK 3 reading scores from gender, race/ethnicity, grouping status, and NJPASS 2 was statistically significant, $F(4, 150) = 21.619, p < .005$. The beta coefficients are presented in Table 17. The adjusted R^2

was .366. This means approximately 36.6% of the variance in the NJASK 3 achievement can be explained by the model. According to Cohen (1988), this is a moderate effect. It is noted that the NJPASS and NJASK were significantly correlated at the .005 level ($r=.587$) and gender, race/ethnicity, and grouping formation were not correlated to NJASK reading scores. When examining the significance of the predictor variables in this model, only NJPASS was a significant contributor to the model. This means that prior ability, as measured by the NJPASS, could explain the variation between the grouping formation. Once prior ability was added to this model, grouping formation became insignificant. Gender and race continued to be insignificant predictor variables in this model.

The null hypothesis was accepted. Grouping status was not a statistically significant ($p=.199$) predictor variable with a beta of .086 and a t value of 1.291. Grouping status is not a strong predictor of the NJASK 3 score for the sample of students in this study because the beta (.086) is not close to 1. The closer the beta is to 1, the stronger the predictive power.

Summary

The predictor variables gender, race/ethnicity, grouping status, and prior reading ability were analyzed to determine their correlation to NJASK reading scores and multicollinearity. NJPASS, which measured students' prior reading ability, was the only predictive variable that was significantly correlated to the dependent variable, NJASK reading scores. When the predictive variables, gender, race/ethnicity, grouping status, and prior ability are combined, they are significant in predicting NJASK 3 scores. The combination of predictive variables accounted for 37% of the variance in the NJASK 3

reading scores. However, only prior ability, as measured by the NJPASS 2, significantly added to the prediction when the other three variables were considered. The results from this study suggest that grouping type within-class or between the grade level did not seem to have an influence on reading achievement; however, prior ability did.

The next chapter presents conclusions from this study and how these findings relate to research on homogeneous and between-class grouping and research on specific ability levels when using grouping. Chapter V will also provide recommendations for practice, policy, and future research.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

Introduction

Assessing students on their reading achievement was one of the requirements of NCLB, and the current reauthorization under the new Every Student Succeeds Act (ESSA) signed into law in December 2016. Under ESSA states must have annual testing in English Language Arts beginning in third grade to measure the target benchmark scores established by the states (Every Student, 2015). Teachers are also being evaluated based on the results of these tests.

In 2015 there were 45 states that included some form of measurable student performance in their teacher evaluation process (Doherty & Jacobs, 2015). Student performance may be measured by the percentage of students attaining the state's established proficiency level or by a student growth model. In states using the growth model, a student is compared to peers who scored similarly the prior year (Betebenner, 2009). Some states refer to this as a "value added" measure. Betebenner (2009) argues that a growth or value added model is a more just statistical model. Student growth percentile (SPG) models assess teachers on how much growth their students made while under their charge. States using a SGP model, like New Jersey, philosophically believe the SGP model allows the district or state to more fairly assess those teachers who are working with students who enter the classroom below grade level (Achieve NJ, 2016; O'Malley et al., 2011). Whether teachers are being assessed by the percentage of their students who attain a level of proficiency or by the evidence of student growth on a state

assessment, teachers in 45 states are being evaluated annually based on some component of student performance (Doherty & Jacobs, 2015).

As teachers and schools react to this method of public accountability, grouping has become one of many instructional strategies implemented with the goal of increasing the performance of all learners (Chorzempa & Graham, 2006; Loveless, 2013). Grouping use has increased in the 21st century (Chorzempa & Graham, 2006; Ford & Opitz, 2011; LeTendre et al., 2003). In 1998, 28% of fourth grade teachers reported using ability grouping, but in 2009 the percentage increased to 71% (Loveless, 2013). Loveless (2013) hypothesizes the federal mandated state testing has been a catalyst to increase the frequency of grouping.

The purpose of this correlational, explanatory, cross-sectional study was to explain the influence of the type of grouping (within-class or between-class) on the third grade New Jersey Assessment of Skills and Knowledge (NJASK 3) reading scores in one suburban district in New Jersey. This study sought to explain the amount of variance in the NJASK 3 reading rates accounted for by grouping type, student gender, race/ethnicity, and prior reading ability. The following overarching research questions guided this study: What is the influence of grouping formation on the NJASK 3 reading scores? What are the differences in reading achievement between third graders who are grouped between the grade level and those who group for reading within the classroom?

The results of the study revealed the influence, or lack of influence, grouping formation had on the NJASK 3 reading scores on the students in this sample. However, the results of this study need to be contextualized within the current research on other mitigating variables, which could not or were not examined within the confines of this

study. This chapter discusses the limitations of the study, the results by each predictor variable, and recommendations for policy, planning, and future research.

This study had significant limitations that must be considered during the discussion of the results. There were two limitations, subject diversity and the inability to account for other mitigating variables that were not captured in the models but could be strongly associated with the outcome variable, test scores. Although it is well established in the literature that socioeconomic status significantly influences student achievement, due to data limitations, it could not be considered as a predictor variable in this study (NICHHD, 2005; Washbrook & Waldfogel, 2011). The sample was screened for free/reduced-price lunch status in the treatment and control schools, but there were only five students in the sample who qualified for free/reduced-price lunch out of the 155 participants. Free/reduced-price lunch status is only one crude measure of student SES; but without other variables (parental education, for example), the study could not capture this important feature of students' background.

Sample size limitations also resulted in broad categorization of students into minority vs. White racial/ethnic groups. This study could not provide the results by specific racial subgroup in order to maintain subject confidentiality since there were only a maximum of eleven participants in each minority group reported to the state of New Jersey. This resulted in the researcher grouping all classifications of minorities into one category. Grouping minorities conflates internal variation and may not reflect the academic disparity amongst racial groups (Davis-Kean & Jager, 2014; Fryer & Levitt, 2004).

There were four predictor variables selected for this study: gender, race/ethnicity, prior ability, and grouping status. However, there are numerous mitigating variables, which may influence student achievement that were not quantified in this study. For example, in this study schools were not randomly assigned to the control and independent group. The teachers and/or principal made the determination if they were going to implement between-grade-level grouping or within-class grouping. The treatment school made a choice to implement between-class grouping and this might have influenced results. Another mitigating variable is administrative support. The depth and professional support offered to teachers in each school was not examined. Hypothetically, one school may have had more administrative observations and feedback during reading instruction, whereas another school may have had a principal who could not devote the time or did not have the knowledge to support teachers.

Professional collaboration is an additional variable that was not examined but might influence achievement. The attitudes and beliefs of teachers and their ability to collaborate may or may not have been equivalent in each school. Teachers using guided reading need to modify instruction and materials based on the instructional needs determined by the teacher through informal and formal assessments (Fountas & Pinnell, 1996). A teacher's level of experience, professional support, and prior knowledge certainly impacts her ability to diagnose and remediate reading weaknesses, thus impacting reading achievement (Brackley, 2015; Jacob, Goddard, Kim, Miller, & Goddard, 2015).

Finally, the study did not examine the frequency of group meetings or teacher contact time. The treatment school might have had more student and teacher contact time

since students were closer in academic ability. Teachers may have selected to combine groups or use whole group instruction when working on skills needed by the majority of students. This brief discussion of study limitations just highlights some glaring weaknesses in the study; however, there are numerous factors that can affect reading achievement.

Conclusions and Policy Recommendations: Gender

Researchers have highlighted the gender performance differences in the last decade as girls performed higher than boys on state, national, and international tests (Cornwell, Mustard, & Van Parys, 2013; Kingdon, Serbin, & Stack, 2016; Schwabe, McElvany, & Trendtel, 2015). This section reviews these performance trends at the state and national levels and compares these finding to the gender and grouping results from this study. Policy recommendations and considerations conclude the discussion on gender.

The state of New Jersey reports results on state assessments based on gender demographics. The results of the NJASK third and fourth grade Language Arts assessments from the year of this study, 2013, and the following year revealed 10% more females scored at the Proficient or Advanced Proficient levels established by the NJDOE than did males who took the tests (NJDOE Performance Reports, 2013, 2014). The 2015 New Jersey Language Arts PARCC results revealed that 9% more females met or exceeded standard benchmarks (NJDOE, Performance Reports, 2015). In the last three years 9% to 10% more third grade females met or exceeded benchmark standards even when two assessments companies, Measurement Incorporated and Pearson, were involved in the assessment test design, administration, and scoring.

The gender-based performance gap in New Jersey is generally reflective of national trends. The National Assessment of Educational Progress (NAEP) assesses national trends in education. NAEP is a congressionally mandated national assessment administered by the National Center for Educational Statistics every four years to students in Grades 4, 8, and 12. The NAEP began in 1969 and measures national educational trends in reading and math (National Center for Educational Statistics, 2016). The National Center for Educational Statistics releases the results to allow states, researchers, and educators to analyze longitudinal national trends from a representative sample. An examination of the NAEP results by gender performance from the fourth grade testing years from 1992-2012 indicates females have outperformed males when comparing mean reading scores from 1992 until the last testing window (National Center for Educational Statistics, 2016). The National Center for Educational Statistics (NCES) reported the gap narrowed significantly ($p < .05$) to five points during the 2012 assessment in fourth grade, whereas in 1971 the gender gap was 13 points (National Center for Educational Statistics, 2016). Females have been outperforming males on New Jersey's state assessment and NAEP (NJDOE Performance Reports, 2013, 2014; NCES, 2016). The question of whether girls and boys respond differently to grouping offers another avenue to explore these gender performance gaps.

While most of the research to date has focused on overall gender differences in reading performance, there have been limited studies on the results of reading performance when examining the influence of grouping by gender. Macqueen (2008) examined reading performance by gender while considering grouping formation. Macqueen (2008) found no significant difference in reading performance between the

male and female elementary students when using grouping. This study's descriptive statistics revealed females had an insignificant (1.4 point) higher mean score than males on the NJPASS 2 and performed the same as males on NJASK 3. While there is limited research on the influence of gender on reading achievement when using grouping as an instructional strategy, it is still worth noting that this study supported Macqueen's (2008) work. There were no appreciable gender-based differences in response to grouping in terms of reading scores. Males and females in both grouping formations, between-class or within-class grouping, performed similarly. However, researchers and policymakers must consider the large body of research on assessments at the state and national levels that found a significant difference in reading achievement between boys and girls (Sax, 2009; Schwabe, McElvany, & Trendtel, 2015).

Another way to think about the relationship between gender, achievement, and grouping is to consider same-sex educational grouping. Two meta-analytical studies on gender grouping may offer some direction to policymakers. The U.S. Department of Education authorized a study of gender-based grouping and selected Mael, Alonso, Gibson, Rogers, and Smith (2005) to review the body of quantitative research on single sex education as compared to coeducational programs. Mael et al. (2005) started with over 2,000 studies and then analyzed 40 studies, which met the study requirements. The achievement results in language arts were null; seven out of the ten studies included in the meta-analysis found girls and boys did not perform significantly different in quantifiable language arts assessments in single sex versus coeducational high school settings (Mael et al., 2005). Recently, Pahlke, Hyde, & Allison (2014) conducted another meta-analysis of 184 studies and also found little evidence to support single sex grouping

compared to mixed gender groups when examining student performance. While there may be no significant difference in academic performance in same sex groups, there are parents and students who select same sex grouping formation for other reasons. Parents may want to expose their children to a religious curriculum, increase learning opportunities, or recognize learning differences between the sexes (Liben, 2015; Sax, 2009)

Policymakers need to consider the research that grouping by gender does not appear to be a strategy that influences student achievement (Mael et al., 2005; Pahlke et al., 2014). Policymakers also need to be mindful that there was no significant difference between male and female performance on the NJASK 3 in this small study when grouping was utilized in either the classroom or between the grades. Based on the larger literature on the subject, grouping by gender should not be considered a strategy to improve students' reading achievement or decrease the gender performance gap. Policymakers need to consider other factors that decrease the gender performance gap.

Conclusions and Policy Recommendations: Race/Ethnicity

The No Child Left Behind Act of 2002 and its reauthorization under ESSA in 2015 required performance reporting of specific groups of students. NCLB required states report the performance of English Language Learners, special education students, males, females, low-income students, and minorities (NCLB, 2001). The reporting of disaggregated data allowed the federal and state governments to sanction those schools not showing improved performance for a specified group. The performance of minorities, one such disaggregated group, has been examined for numerous years. Oakes' (1985) book *Keeping Track: How Schools Structure Inequality* highlighted how minorities were

overrepresented in the lowest tracks, were subjected to lower standards and thinking skills, and were often taught by ineffective teachers (Oakes, 1985). Current research reveals minorities are still dealing with the same injustices highlighted by Oakes 25 years ago (Macqueen, 2008; MacIntyre & Ireson, 2002; Nomi, 2006; Worthy, 2010).

In order to collect information on race/ethnicity demographics as directed under the 2002 NCLB Act, New Jersey requires guardians to identify their child's race during the school registration process. Race options include White, Black or African American, Hispanic or Latino, Asian, Pacific Islander, Native American or Alaskan, or Multiracial. The results of the NJASK third and fourth grade Language Arts assessments from the year of this study, 2013, and the following year showed that more White and Asian students scored at the Proficient or Advanced Proficient levels established by the NJDOE than Hispanic/ Latino and Black/African American students. Approximately 75% of third grade White students and 86% of Asian students reached the Proficient or Advanced Proficient levels on the NJASK 3 in 2013 and 2014; by contrast, only 49% of Hispanic/Latino and 47% of Black/African American students reached the same levels (NJDOE Performance Reports, 2013, 2014). Historic trends on NJASK 3 results from 2008 to 2014 indicated Black and Hispanic students improved scores on the NJASK 3 in Language Arts, but the racial gap remained in New Jersey.

On the national level NAEP scores are disaggregated by racial/ethnic group as well. There are six reported subgroups: Caucasian, Black, Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, and Two or More Races (NCES, 2016). The National Center for Educational Statistics (NCES) reported results from the 2015 NAEP fourth grade reading assessment. When comparing the six race/ethnicities, Whites scored

significantly higher on the fourth grade NAEP assessment in 2015 compared to every race/ethnicity except Asian/Pacific Islander, who scored higher than Whites (NCES, 2016). Vanneman, Hamilton, Baldwin, Anderson, and Rahman (2009) conducted an analysis of NAEP data and identified 44 states that had a statistically significant gap in reading scores between Black and White students. The other six states did not have available data on specific racial groups to be included in this study. Even though there have been significant achievement gaps by race, the results of the NAEP reading assessment indicate Black and Hispanic students have narrowed the gap slightly (NCES, 2016).

The racial and ethnic performance disparities are not just limited to state and national assessments. Grouping and racial performance research to date has highlighted some disparities in racial performance when using grouping. Tach and Farkas (2006) examined the ECLS data and found Black and Hispanic students scored lower than White students in grouped classrooms. Lleras and Rangel (2009) also studied the ECLS data and found grouping by ability was not as effective for Black and Hispanic students when compared to similar ungrouped peers. Interestingly, Lleras and Rangel (2009) found higher achieving Black and Hispanic students performed better in ability groups than their ungrouped academic peers. Tach and Farkas (2006) and Lleras and Rangel (2009) did not stipulate the type of grouping used in these classrooms; thus, comparison to this study is challenging.

The purpose of this study was to examine the influence of grouping on the NJASK 3 reading scores when controlling for the predictor variables. One of the predictor variables, race/ethnicity, did not have a significant correlation with grouping

status and did not significantly influence the performance on the dependent variable, NJASK 3 reading scores. Again, the reader must use caution in examining the race/ethnicity results from this study. There were no more than 11 students in each minority category; thus, the four minority categories represented in this sample were combined. Grouping the minorities conflates internal variation. For example, in this study Asian students were the largest race/ethnicity in the overall minority variable. Given previous evidence of Asian students outperforming White students, the results of this subgroup combined with Black and Hispanic students who have traditionally scored lower could obscure the actual achievement patterns and response of students of different racial/ethnic backgrounds to the grouping strategy employed. While this study did not find race/ethnicity to be a significant contributor to the regression model, there is strong evidence from NAEP and NJASK that race/ethnicity must be considered in planning instructional programs and grouping decisions. The racial gap needs to be addressed and policymakers need to examine research that has proven to decrease this gap.

Policymakers should share the data on the racial performance gap that exists with parents, teachers, and community members and be transparent about where the majority of the gap comes from: conditions associated with poverty. Community members, teachers, and parents then must be part of the discussion on how to overcome the racial gap if using ability grouping to teach reading. Policymakers need to monitor the racial composition of reading groups and the fluidity between the groups. Research indicates minorities are overrepresented in the lowest groups and have less movement between groups (Macqueen, 2008; Worthy, 2010). In addition, policymakers need to make sure reading group instruction is culturally responsive (Ladson-Billings, 1994).

Nieto and Turner (2012) and Banks (2001) make recommendations that foster a culturally responsive school or group rather than an environment that may unknowingly promote “institutional racism” (Banks, 2001). To create culturally responsive reading groups, teachers must learn, and not assume, the language and literature experiences of the races/ethnicities served (Nieto & Turner, 2012). Teachers can submerge themselves into the community, learning what parents and community members do to promote language arts through songs, dance, stories, and books. Nieto (2012) suggests teachers and policymakers show respect for diversity, infuse cultural experiences into a rich curriculum that matches the learners, and set high standards. Parents also should be involved in supporting their children’s education (Nieto & Turner, 2012). In return, teachers should explain to parents the reading skills being taught and model how to infuse these skills through cultural experiences. Finally, teachers need to understand that reading is not just a cognitive process but a cultural experience as well (Compton-Lilly, 2015). Rosenblatt’s Reader-response theory (1978) recognizes the personal interaction between the text and the reader. A student with a multicultural background will bring these experiences to the text. Teachers assessing students on their reading level have to apply their knowledge of the culture when assessing students’ responses and reflections. Acceptable answers provided by a running record program like the DRA or Fountas and Pinnell benchmark system may not provide sample responses more typical for a specific race/ethnicity.

Implementing a culturally responsive reading classroom means teachers must recognize students of all races and ethnicities for their abilities and provide for alternative and/or unbiased assessments. Research has documented Black and Hispanic students are

underrepresented in gifted or advanced ability groups (Carman & Taylor, 2010; Oakes, 1995; Yoon & Gentry, 2009). Examination of previous standardized tests found some test items to be biased by giving a specific racial or socioeconomic group an unfair advantage, which led to lower test scores for other groups (McClain & Pfeiffer, 2012). Test creators are now cognizant of test bias, but informal assessments in the classroom may unknowingly be biased. Teachers need to recognize racial groups or SES groups may provide acceptable responses that may not be familiar to a teacher. In addition, policy creators and teachers need to be mindful that teachers are often involved in the identification of students for higher ability groups and the classroom environment may not highlight the giftedness of students from diverse cultural and SES groups (Mansfield, 2015).

In order to showcase talents from various cultures and racial groups, reading materials used to assess students' abilities for group placement should integrate themes like holidays, historical events, role models, and problems relating to the community of learners. Students may offer higher-level responses when they have the prior knowledge and text-to-self relationships often required during the reading experience. Policymakers should examine school demographic data to ensure that make-up of higher ability and lower ability reading groups are not disproportionate. Finally, teachers and policymakers need to acknowledge diversity is an asset during group book discussions and allows for various interpretations of the author's message. Allowing students to bring their culture and heritage into the school may decrease the reading achievement gap while making the school more enriching (Nieto & Turner, 2012).

Conclusion and Policy Recommendations: Prior Ability

In 1904 Binet developed the Intelligence Quotient (IQ) assessment, which introduced stratification of intelligence levels (Ansalone, 2006). It was believed that IQ was fixed and indicated one's ability to learn and acquire new skills (Fancher & Rutherford, 2011). Teachers and administrators began to use IQ results as a measure of a student's potential to learn and made assumptions about a student's reading level or stratified group (Fancher & Rutherford, 2011). Scholars have debated the use of an IQ test as a valid indication of one's ability or potential in the classroom or workplace (Sternberg, 1985; Duckworth, Quinn, & Tsukayama, 2012; Goleman, 2006). Sternberg (1985) has been critical of the IQ test as a sole determination of one's ability to learn. Sternberg postulates intelligence is really about how one performs within one's world by one's ability to analyze, adapt, and apply oneself. Duckworth and Seligman (2005) found measurements of self-discipline were a better predictor of student future performance than IQ. The ability to predict future performance continues to be studied today.

The IQ test may have been a formal assessment of ability, but students' potential has been predicted informally by teachers for many years (Begency, Eckert, Montarello, & Storie, 2008; Rubie-Davies et al., 2006; Worthy, 2010). Educators assumed a student's past performance was an indication of how a student would perform in the future. Many researchers have empirically examined this correlation between prior ability and future achievement (Hernandez, 2011; Shaywitz et al., 1992). The use of assessments to determine a student's current learning capability and predict future reading achievement is the basis of this next section.

Prior reading ability and future reading performance has been a specific subtopic of ability research that merits investigation. Researchers found that reading below grade level has a negative effect on future achievement (Hernandez, 2011; Lesnick, George, Smithgall, & Gwynne; 2010; Shaywitz et al., 1992;). Shaywitz et al. (1992) and Francis (1996) found 75% of children reading below grade level in third grade were still struggling with reading in ninth grade. Recently, researchers have built upon these foundational studies to further examine the predictive power of test performance on future outcomes (Hernandez, 2011; Oakhill & Cain, 2012). Researchers have examined large samples of students' performance and have predicted future reading achievement as early as first grade (Dogan, Ogut, & Kim, 2015). Oakhill and Cain (2012) and Lesnick et al. (2010) found ability levels by third grade were strong predictors of reading performance. Hernandez (2011) even correlated third grade reading performance to future reading achievement and graduation rates.

The discussion of prior ability must include research on poverty. Poverty is highly correlated to reading achievement (Hernandez, 2011). Children who live in poverty or have spent time living in poverty have a higher correlation with reading below grade level. As previously examined, reading below grade level then influences future achievement (Dogan et al., 2015; Oakhill & Cain, 2012). Hernandez (2011) referred to this comorbidity as "double jeopardy." Hernandez (2011) examined future performance based on NAEP reading scores. He found students who spent time living in poverty were less likely to be reading at established proficient levels compared to their peers not living in poverty. Hernandez found 83% of students living in poverty were not reading at the proficient levels established by the NAEP compared to 55% of peers living with

moderate to high-income families. Poverty also negatively influenced graduation and attendance rates. Hernandez compared students with similar sub-par reading levels and found 26 % of students living in poverty did not graduate compared to just 9% of peers from moderate to higher income families. Poverty matters.

Prior ability, whether based on SES or innate ability in reading at the elementary level, has an influence on future academic achievement, but this study wanted to examine the influence of prior ability when grouping was used as an instructional strategy. The next section discusses how the findings from this study complement the extant research and influence policy and practices.

There are inconsistent results on the benefits or disadvantages of grouping on the lowest readers. Mosteller et al. (1996), Condron (2005), Macqueen (2008), Hallam et al., William and Bartholomew (2004), and Lleras (2008) found that the lowest ability group scored significantly behind their ungrouped peers, whereas Collins and Gan (2013), Yadegari and Ryan (2002), Duflo et al. (2009), Puzio and Colby (2010), and Kulik (2003) found lower ability students did better in homogeneous groups. The research on higher achieving students indicated these students performed significantly better in groups than their ungrouped peers (Collins & Gan, 2013; Condron, 2005; Duflo et al., 2009; Llaras & Rangel, 2009). Finally there were researchers who found all students or those with average prior ability did the same regardless of type of grouping formation (Betts & Shkolnik, 2000; Hallinan, 1994b; Kulik & Kulik, 1982; Nomi, 2005).

This study included prior ability as a predictor variable and defined this variable as the NJPASS 2 reading score from second grade. This study found prior ability was significantly correlated ($r=.587$) to the dependent variable, NJASK 3 reading scores.

Prior reading ability was also the only significant contributor to the regression model.

This study's findings supported the scope of research on the positive correlation between prior ability and future performance (Hernandez, 2011; Oakhill & Cain, 2012; Shaywitz et al., 1992).

The research on prior ability and the results of this study have implications for policymakers. First, the research indicates prior ability affects future learning (Hernandez, 2011). It would behoove federal and state policymakers to consider these findings in their teacher evaluation formulas that are linked to student achievement. States like NJ are using a Student Growth Model, which really evaluates a teacher based on the ability cohort of each student (Betebenner, 2009). Evaluations that take into account a student's prior ability would allow a teacher to be evaluated on how much growth his/her students made compared to students of similar prior ability across the state. When selecting a quantifiable evaluation tool as required by ESSA, policymakers should consider a student growth or value added evaluation system that is based on the research on the correlation between prior ability and student performance.

Second, the research on the importance of reading on grade level by third grade is clear (Hernandez, 2011; Lesnick et al, 2010), and there is even some new research suggesting the correlation between reading levels and future academic achievement may be found even earlier than third grade (Dogan et al., 2015; Oakhill & Cain, 2013). Early intervention is paramount. Early and repeated reading intervention in Grades 1 through 3 has a significant impact on reading achievement on nationally normed reading assessments compared to those students who just received limited reading intervention (Connor, Morrison, Fishman, Crowe, Al Otaiba, & Schatschneider, 2013). It is also

important for policymakers and educators to recognize a significant factor that influences reading achievement: poverty. Policymakers and government officials need to address the poverty issues facing a country. Federal policies need to examine living wages, quality preschool programs, and affordable health care.

Finally, policymakers and educators must think about the needs of all learners when considering reading programs. This study's results indicate students who performed lower on the NJPASS performed lower on NJASK 3, and conversely students who performed well on the NJPASS were likely to perform well on the NJASK 3. This study did not find grouping within the class or between the grade level to have an influence on academic reading achievement. There is empirical evidence to suggest more advanced learners do better in homogeneous groups (Iberman, Kugler, & Sacerdote, 2012).

Paradoxically, policymakers and educators have to consider the needs of the lowest learners who benefit from peer models, more qualified teachers, enriched curriculum, and higher expectations (Kalogrides & Loeb, 2013). Policymakers need to have the difficult conversations about balancing the needs of the highest performing students who benefit from ability-based groups with the needs of the lowest performing students who may not benefit from grouping. Slavin (1988) examined this conundrum and may offer policymakers a possible recommendation that balances the needs of all learners.

Slavin (1988) suggests educators limit grouping to one or two subjects to give each ability group time to work at their instructional level but then balance the remainder of the school day with heterogeneous grouping (Jecks, 2011). In addition, administrators have to be cognizant of the research, which shows that lower level students often receive inexperienced and ineffective teachers (Oakes, 2005; Worthy, 2010). Administrators

should consider placing reading specialists or highly skilled teachers with the most at-risk learners or rotating teacher assignments. Slavin (1988), Tieso (2005), Tomlinson (2005), and Ford and Opitz (2011) also recommend differentiated curriculum for each ability group to remediate or expand skills as determined by diagnostic testing. Teachers should base their reading instruction by analyzing a student's past performance. Teachers and policymakers should look for overall reading trends in DRA and standardized assessments, and then form remediation groups based on skills. The goal of the group should be to teach skills to improve reading performance, thus improving the student's academic reading trajectory (Caldwell & Ford, 2002; Fountas and Pinnell, 1992). The research on prior ability is evident and the stakes are high; however, researchers have offered some possible remedies that policymakers need to consider.

Conclusions and Policy Recommendations: Grouping Formation

The term *grouping*, when applied to a school setting, is not adequate to provide a true understanding of the formation of two or more students; grouping may look different depending on how it is implemented. Policymakers need to be cognizant of this when examining grouping research. For example, grouping may be heterogeneous or homogeneous in composition. Grouping may take place within the classroom or between the classes; groups may be set for the year or reconfigured as needed (Tieso, 2003). The permutations of grouping formations are vast, which makes comparing research complex. Prior research conducted on grouping formation has examined grouping through three perspectives: heterogeneous versus homogeneous grouped classes, within-class grouping versus whole-group heterogeneous instruction, and between-same-grade-level grouping

versus heterogeneous grouping. The latter grouping formation, grade level reconfiguration, was the focus of this study.

In order to examine this specific grouping reconfiguration, the overarching question for this study was, “What are the differences in reading achievement between third graders who are grouped between the grade level and those who are grouped for reading within the classroom?” In order to answer this question quantitatively, a correlational analysis between variables was conducted and then those variables were entered into a regression model. The results from these statistical measurements found that the type of ability grouping was not significantly associated with reading achievement. Even though this study was limited in its subjects’ racial and SES diversity, it did support Haighighat’s (2009) findings. Haighighat found no difference in student movement between the Arizona state testing performance categories in a school in the district using flexible grouping between the grade and the control schools, except for two areas. The treatment school using between-grade-level grouping showed the most gains with students who had base line data at the “Falls Far Behind,” the lowest classification rating on the Arizona state assessment. More of the students classified as “Falls Far Behind” in the treatment school moved up to the next reporting category when compared to the control schools (Haghighat, 2009). Also, Hispanic students appeared to benefit from same grade level grouping. The treatment school had 40% more Hispanic students, who started at the “Approaching” category, improve one reporting category when using flexible grouping.

Policymakers need to consider Haghighat’s (2009) work suggesting the most at-risk learners may benefit from the flexible grouping formation. Policymakers should not

confuse flexible grouping with fixed grouping. Flexible grouping is a temporary group that comes together to reach an objective or goal (Tomlinson 2003, 2005; Wormelli, 2007). Teachers vary teaching strategies, materials, and/or curriculum in order to reach the objective (Ford, 2005; McCoach et al., 2006; Tieso, 2005). It is this type of grouping that may benefit the at-risk, minority students. Fixed grouping did not benefit the lowest learners (Condrón, 2005; Lou et al., 1996; Slavin, 1987) and should be implemented by policymakers only after careful consideration.

In closing, it is recommended policymakers identify demographic information of their community carefully, accounting for race/ethnicity, SES, and ability levels. Once this information is gathered, then policymakers need to investigate specific grouping research that is reflective of the community's demographic makeup. This research can help policymakers determine if whole-group instruction, flexible within-class grouping, or between-class grouping may be the most appropriate strategy to pursue for their learners. With any program, policymakers need to monitor implementation and conduct frequent interim assessments to determine effectiveness. Finally, policymakers have an obligation to share findings with researchers who may be able to assist in analyzing data and contribute to the research on modern day grouping practices.

Recommendations for Future Research

This research adds to the extant literature on the influence of grouping formation on students' reading performance at the elementary level. However, one study cannot lead to conclusive findings for all learners in all grouping or learning situations. In order to complement the literature, it is important to conduct future research. Research on grouping has been inconsistent, and Slavin (1987) and Nomi (2006) believe part of the

complications of examining grouping research is the number of mitigating variables that occur simultaneously in a grouping formation. Future researchers should consider analyzing additional variables that may be related to the relationship between grouping and student achievement. For example, grouping may be impacted by the support or lack of support by district administrative personnel, professional development, and materials provided to teachers. This study never examined the principals' beliefs and how this impacted the support provided to the teachers. Jacob et al. (2015) found administrators can foster teacher collaboration on a set goal, which can improve student achievement. Future researchers may want to examine the collaboration process principals instill when discussing grouping and its correlation to student achievement. In addition, future researchers need to examine the frequency of implementing guided reading and student achievement. Ford and Opitz (2008) conducted a national survey and reported most teachers reported having four reading groups and using guided reading three to five days a week with five days being the most frequent response. However, it has not been determined if the number of group meetings per week influences student reading achievement. This is an area that warrants future discussion.

This study examined the number of years of teacher experience when considering if the control and treatment school could be compared. However, this study focused on reading achievement. Teachers' training in reading may offer some insight into how well teachers are equipped to handle a specific level of ability or to remediate skills. For example, a low reader may have dyslexia, and a teacher trained in Orton Gillingham may be more effective in remediating this specific reading disability (Shaywitz, 2003). The teachers' belief in the use of grouping practice was also not examined and may be

another one of the mitigating variables to which Slavin (1987) was referring. Some teachers may or may not have believed lack of prior ability can be overcome through the use of grouping, or some teachers may have not supported the implementation. A teacher's efficacy is also a factor to be researched. Teacher efficacy is discussed in the work of Bandura (1977), who theorized one's belief in his/her ability to navigate the environment influences outcomes. A teacher's belief has been found to have an effect on student achievement (Hoy & Spero, 2005) and this is an area for future grouping research.

As previously discussed, this study was hindered by its lack of subject diversity. SES is a strong predictor of student achievement (Tienken, 2012). It would behoove future researchers to learn if flexible grouping or another grouping formation is a teaching technique capable of overcoming the income and racial achievement gap. To enrich the suggested research on flexible grouping, researchers should include data on SES and each individual racial/ethnic group's outcomes rather than use a composite measure predictor variable. Future researchers who can differentiate the race/ethnicity variation may be able to detect where academic disparity amongst racial groups exists and if a grouping formation was influential in reducing racial achievement gaps. Future researchers may also want to consider how they measure the influence of grouping on a diverse population. For example, this study examined only the influence on NJASK 3 reading scores, but there may have been other factors like special education referrals, parental feedback or student self-worth that may have been influenced by grouping formation and captured the potential benefits of flexible grouping.

Finally, the research on grouping is inconclusive. However, as Loveless (2013) highlighted, grouping is on the rise. This study did not explore the decision-making

process personnel undertook to determine whether to implement or not implement grouping in the classroom. Future researchers may want to explore qualitatively the factors personnel considered when implementing grouping and the research conducted. Researchers may consider documenting the implementation steps when incorporating grouping in the reading classroom. This may provide insight into the paradox of the inconclusive research and the increase in grouping in the modern classroom.

Summary

This study examined the influence of the predictor variables gender, race/ethnicity, prior ability, and grouping status on the NJASK 3 reading scores on third graders in one suburban district in New Jersey. The study found that only prior ability had a significant influence on the NJASK 3 reading scores.

References

- Achieve NJ. (2016, September 23). New Jersey P.L. 2012, 26. Retrieved from http://www.njleg.state.nj.us/2012/Bills/PL12/26_.PDF
- Algozzine, B., & Anderson, K. M. (2007). Tips for teaching: Differentiating instruction to include all students. *Preventing School Failure: Alternative Education for Children and Youth*, 51(3), 49-54.
- Allington, R. (1980). Teacher interrupting behavior during primary grade oral reading. *Journal of Educational Psychology*, 72(3), 371-377.
- Allison, P. D. (2009). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.
- Allport, G. W. (1937). The functional autonomy of motives. *The American Journal of Psychology*, 50(1/4), 141-156.
- Ambruster, B., Lehr, F., & Osborn, J. (2001). Put reading first: The research building blocks for teaching children to read. Kindergarten to grade 3. Washington, DC: National Institute of Health.
- Ansalone, G. (2000). Keeping on track: A reassessment of tracking in the schools. *Race, Gender & Class*, 7(3), 108-132.
- Ansalone, G. (2006). Tracking: A return to Jim Crow. *Race, Gender & Class*, 13(1/2), 144-153.
- Ansalone, G., & Biafora, F. (2004). Elementary school teachers' perceptions and attitudes to the educational structure of tracking. *Education*, 125(2), 249-258.
- Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student

- performance in middle and high school English. *American Educational Research Journal*, 40(3), 685-730.
- Austin, M., & Morrison, C. (1961). *The torch lighters: Tomorrow's teachers of reading*. Cambridge, MA: Harvard University Graduate School of Education.
- Banks, J. A. (2001). Citizenship education and diversity. *Journal of Teacher Education*, 52(1), 5-16. Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/224012034?accountid=13793>
- Banks, J. A., Cookson, P., Gay, G., Hawley, W. D., Irvine, J., Nieto, S., . . . Stephan, W. G. (2001). Diversity within unity: Essential principles for teaching and learning in a multicultural society. *Phi Delta Kappan*, 83(3), 196-203. Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/218472190?accountid=13793>
- Barr, R. (1995). What research says about grouping in the past and present and what it suggests about the future. In M. Radencich & L. McKay (Eds.) *Flexible grouping for literacy instruction in the elementary grades* (pp. 1-24). Needham Heights, MA: Allyn & Bacon.
- Baumann, J. F., & Heubach, K. M. (1996). Do basal readers deskill teachers? A national survey of educators' use and opinions of basals. *The Elementary Journal*, 96(5), 511-526.
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43.

- Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4), 42-51.
- Betts, E. (1946). *Foundations of reading instruction with emphasis on differentiated guidance*. Oxford, England: American Book Company.
- Betts, J., & Shkolnik, J. (2000). Key difficulties in identifying the effect of ability grouping on student achievement. *Economics of Education Review*, 19(1), 21-26.
- Bonk, W., Copa, J., Gibson, N., Gillan, T., Nau, J., Peoples, A., Wand, X., & Woolard, C. (2012, July). Growth models issues and advice from the states. Institute of Education Sciences. Retrieved from <http://nces.ed.gov/programs/SLDS>
- Booher-Jennings, J. (2005). Below the bubble: Educational triage and the Texas accountability system. *American Education Research Journal*, 42(2), 231-268.
- Brackley, L. (2015). A response to Elliot. *Psychology of Education Review*, 39(1), 28-31.
- Brown v. Board of Education of Topeka, 347, U.S. 483 (1954).
- Burchaninal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology*, 40(5), 415-436.
- Butler, J. (1987). Differences in achievement and time-on-task with homogeneous and heterogeneous ability grouping of second graders. Retrieved from ERIC database (ED289589)
- Caldwell, J., & Ford, M. (2002). *Where have all the bluebirds gone? How to soar with flexible grouping*. Portsmouth, NH: Heinemann.

- Camara, W. J. (2009). College admission testing: Myths and realities in an age of admissions hype. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 147-180). Washington, DC: American Psychological Association.
- Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school English. *American Educational Research Journal*, 39(4), 801-827.
- Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri Nonverbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, 54, 75-84.
- Carson, R., & Thompson, J. (1964). The Joplin plan and traditional reading groups. *The Elementary School Journal*, 45(1), 38-43.
- Chorzempa, B. F., & Graham, S. (2006). Primary-grade teachers' use of within-class ability grouping in reading. *Journal of Educational Psychology*, 98(3), 529-541.
- Cogan, K. (1995). Meeting every child's academic needs. *Christian Science Monitor*, 88, 19-21.
- Cohen, J. (1988). *Statistical power and analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence.
- Colby, K., & Puzio, G. (2010). The effects of within-class grouping on reading achievement: A meta-analytic synthesis. *Society for Research on Educational Effectiveness*. Retrieved from ERIC database. (ED514135)
- Colemn, J. S. (1967). *The concept of equality of educational opportunity*. Retrieved from ERIC database. (ED015157)

- Coleman, M. R. B. (2016). Commentary-sorting it out: Thoughts on "does sorting students improve scores? An analysis of class composition" by Courtney A. Collins and Li Gan. *Journal of Advanced Academics*, 27(2), 117-123.
- Collins, C. A., & Gan, L. (2013). *Does sorting students improve scores? An analysis of class composition* (NBER Working Paper No. 18848). New York, NY: National Bureau of Economic Research.
- Common Core State Standards: Standards in your state. (2016, May 10). Retrieved from <http://www.corestandards.org/standards-in-your-state/>
- Compton-Lilly, C. (2015). Reading lessons from Martin: A case study of one African American student. *Language Arts*, 92(6), 401-411. Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/1692493580?accountid=13793>
- Condrón, D (2005). *Stratification, skill grouping and learning to read in first grade* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No. AAT 1113411746).
- Condrón, D. (2008). An early start: Skill grouping and unequal reading gains in the elementary years. *The Sociological Quarterly*, 49, 363-94.
- Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science*, 24(8), 1408-1419.
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236-264.

- Cunningham, P. M., Hall, D. P., & Sigmon, C. (2008). *The teacher's guide to the four-blocks*. Greensboro, NC: Carson-Dellosa Christian Publishing.
- Cushenbery, D. C. (1967). The Joplin plan and cross grade grouping. Retrieved from ERIC database. (ED013708)
- 013708.
- Davis-Kean, P., & Jager, J. (2014). Trajectories of achievement within race/ethnicity: "Catching up" in achievement across time. *The Journal of Educational Research, 107*(3).
- Dogan, E., Ogut, B., & Kim, Y. Y. (2015). Early childhood reading skills and proficiency in NAEP eighth-grade reading assessment. *Applied Measurement in Education, 28*(3), 187-201.
- Doherty, K., & Jacobs, S. (2015) State of the states 2015—Connect the dots: Using teacher evaluations to inform policy and practice. Washington, DC: National Council of Teacher Quality.
- Duckorth, A. L., Quinn, P., & Tsukayama, E. (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology, 104*(2), 439–451.
- Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939-944.
- Duflo, E., Dupas, P., & Kremer, M. (2009). Can tracking improve learning? Evidence from Kenya. *Education Next, 9* (3), 64-70.
- Eder, D. (1981). Ability grouping as self-fulfilling prophecy: A micro-analysis of

- teacher-students interaction. *Sociology of Education*, 54, 151-160.
- Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for designing more effective educational environments. *Review of Educational Research*, 43, 163-179.
- Every Student Succeeds Act (ESSA), P. L. No: 114-95 114 U.S.C. (2015). Retrieved from <http://www.ed.gov/essa?src=rn>
- Fancher, R. E., & Rutherford, A. (2012). *Pioneers of psychology: A History* (4th ed.). New York, NY: W. W. Norton.
- Fiedler, E., Lange, R., & Winebrenner, S. (2002). In search of reality: Unraveling the myths about tracking, ability grouping, and the gifted. *Roeper Review*, 24(3), 108–11.
- Firmender, J. M., Reis S. M., & Sweeny, S. M. (2013). Reading comprehension and fluency levels ranges across diverse classrooms: The need for differentiated reading instruction and content. *Gifted Child Quarterly*, 57, 3-14.
- Ford, M. (2005) *Differentiation through flexible grouping: Successfully reaching all readers*. Naperville, IL: Learning Point Associates.
- Ford, M., & Opitz, M. (2011). Looking back to move forward with guided reading. *Reading Horizons*, 50(4), 225-240.
- Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all Children*. Portsmouth, NH: Heinemann.
- Fountas, I. C., & Pinnell, G. S. (1999). *Matching books to readers: A book list for guided reading, K3*. Portsmouth, NH: Heinemann.
- Francis, D. (1996). Developmental lag versus deficit models of reading disability: A

- longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88(1), 3-17.
- Fryer, R. G., & Levitt, S. D. (2004). Falling behind. *Education Next*, 4(4) Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/1237805674?accountid=13793>
- Gaffney, J. A., & Anderson, R. C. (1991). Two tiered scaffolding: Congruent processes of teaching and learning. In E. H. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices and policies* (pp.184-198). New York, NY: Teachers College Press.
- Gambrell, L., Morrow, L. M., and Pressley, M. (2007). *Best practices in literacy instruction* (3d ed.) New York, NY: Guilford Press.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, 59, 185-199.
- Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership*, 50 (2), 11-17.
- Gamoran, A., Nystrand, M., Berends, M., & Lepore, P.C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal*, 32, 687-715.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational research: Competencies for analysis and applications*. New York, NY: Pearson Education.
- Gentry, M. (2016). Commentary on "does sorting students improve scores? An analysis of class composition." *Journal of Advanced Academics*, 27(2), 124-130.
doi:<http://dx.doi.org/10.1177/1932202X16636174>

- Goleman, D. (2006). *Emotional intelligence*. New York, NY: Bantam.
- Graves, M. F., Juel, C., Graves, B. B., & Dewitz, P. (2011). *Teaching reading in the 21st century* (5th ed.), Boston, MA: Allyn and Bacon.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate behavioral research*, 26(3), 499-510.
- Guitierrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32, 19-25.
- Haghighat, I. (2009). *Improving achievement for all students through the implementation of within grade level flexible grouping*. (Doctoral Dissertation) Retrieved from ProQuest Dissertations and Theses Database. (UMI Number 3361839)
- Hallam, S. (2002). *Ability grouping in schools: A literature review*. London, England: Institute of Education, University of London.
- Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experience of different types of grouping in school. *British Educational Research Journal*, 30(4), 515-533.
- Hallam, S., Ireson, J., Lister, V., Chaudbury, I., & Davies, J. (2003). Ability grouping practices in the primary school. *Educational Studies*, 29(1), 69-83.
- Haller, E. J., & Davis, S. A. (1980). Does socioeconomic status bias the assignment of elementary school students to reading groups? *American Educational Research Journal*, 17(4), 409-418.
- Hallinan, M. (2004) The detracking movement: Why children are still grouped by ability. *Education Next*, 4(4), 72-76.
- Hallinan, M., & Sorensen, A. (1983). The formation and stability of instructional groups. *American Sociological Review*, 48, 838-851.

- Hernandez, D. (2011). Double jeopardy: How third grade reading skills and poverty influence high school graduation. Retrieved from <http://www.gradelevelreading.net/>
- Hertberg-Davis, H. (2009). Myth 7: Differentiation in the regular classroom is equivalent to gifted programs and is sufficient. Classroom teachers have the time, the skill, and the will to differentiate adequately. *Gifted Child Quarterly*, 53(4), 251-253.
- Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.)
- Hobson v. Hansen, 269 F. Supp 401 (D.C., 1967).
- Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343-356.
- Iberman, S., Kugler, A., & Sacerdote, B. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5), 2012.
- Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, 37(3), 314-332.
- Jecks, D. (2011). *Flexible skills grouping of elementary students for reading instruction: Meta-analytical review* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. AAT 3490486)
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of

- student sorting within schools. *Educational Researcher*, 42(6), 304-316.
- Kilgore, S. B. (1991). The organizational context of tracking in schools. *American Sociological Review*, 56(2), 189-203.
- Kingdon, D., Serbin, L. A., & Stack, D. M. (2016). Understanding the gender gap in school performance among low-income children A developmental trajectory analysis. *International Journal of Behavioral Development*. 0165025416631836.
- Kirk, J. (Ed.). (2008) *An epitaph for Little Rock: A fiftieth anniversary retrospective on the Central High crisis*. Fayetteville, AR: University of Arkansas Press.
- Kulik, J. A. (2003). Grouping and tracking. *Handbook of Gifted Education*, 3, 268-281.
- Kulik, J. A., & Kulik, C. L. C. (1982). Highlights from “research on ability grouping.” *Educational Leadership*, 39(8), 620-622.
- Kulik, J. A., & Kulik, C. L. C (1987). Effects of ability grouping on student achievement. *Equity and Excellence*, 23(1-2), 22-30.
- Kulik, J. A., & Kulik, C. L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, 36(2), 73-77.
- Ladson-Billings, G. (1994). What we can learn from multicultural education research. *Educational Leadership*, 51(8), 22-26.
- Lesnick, J., Goerge, R., Smithgall, C., & Gwymne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college entrance?* Chicago, IL: Chapin Hall at the University of Chicago.
- LeTendre, G. K., Hofer, B. K., & Shimizu, H. (2003). What is tracking? Cultural expectations in the United States, Germany, and Japan. *American Educational Research Journal*, 40, 43-89.

- Liben, L. S. (2015). Probability values and human values in evaluating single-sex education. *Sex Roles, 72*(9-10), 401-426.
- Lleras, C. (2008). Race, racial concentration and the dynamics of educational inequality across urban and suburban schools. *American Education Research Journal 38*(2), 215-252.
- Lleras, C., & Rangel, C. (2009). Ability grouping practices in elementary schools and African American/Hispanic achievement. *American Journal of Education, 115*, (2), 299-304.
- Lofton, K. (2013). *An examination of skills groups, traditional groups and attendance patterns in third-grade classrooms: Their effects on student achievement based on Discovery assessment test c scores* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No. AAT 3576270).
- Lou, Y. A., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class ability grouping on student achievement: An exploratory model. *The Journal of Educational Research, 94*, 101-112.
- Lou, Y. A., Abrami, P. C., Spence, J. C. Poulsen, C., Chambers, B. & d'Apollonia, S. (1996). Within class grouping: A meta-analysis. *Review of Educational Research, 66*, 423-458.
- Loveless, T. (2009). Tracking and detracking high achievers in Massachusetts' middle schools. *The Fordham Report*, Washington, DC: The Fordham Foundation.
- Loveless, T. (2013). *The resurgence of ability grouping and persistence of tracking: Part*

- II of the Brown Center Report on American Education*. Retrieved from <http://www.brookings.edu/research/reports/2013/03/18-tracking-ability-grouping-loveless>
- Loveless, T., Parkas, S., & Duffett, A. (2008) High-achieving students in the era of NCLB. Thomas B. Fordham Institute. Retrieved from ERIC database. (ED 501703)
- Lucas, S. R. (1999). *Tracking inequality*. New York, NY: Teachers College Press.
- Lucas, S., & Gamoran, A. (2002). Tracking and the achievement gap. In J. E. Chubb & T. Loveless (Eds.), *Bridging the Achievement Gap* (pp. 171-184). Washington, DC: Brookings Institute Press.
- Lunn, J. C. B., & Ferri, E. (1970). *Streaming in the primary school: A longitudinal study of children in streamed and non-streamed junior schools*. London, England: National Foundation for Educational Research in England and Wales.
- Macintyre, H., & Ireson, J. (2002). Within-class ability grouping: Placement of pupils in groups and self-concept. *British Education Research Journal*, 28(2), 249-263.
- Macqueen, S. (2008, December). *Between-class achievement grouping for literacy and numeracy: Academic outcomes for primary students*. Paper presented at Australian Association for Research Education Conference: Changing Climates Education for Sustainable Futures, Brisbane, Australia. Retrieved from <http://nova.newcastle.edu.au:8080/vital/access/manager/Repository/uon:8893>
- Macqueen, S. (2010). Primary teachers attitudes in achievement-based literacy classes. *Issues in Educational Research*, 20(2), 118-136.
- Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). Single-sex versus

- coeducational schooling: A systematic review. (Document No. 2005-01).
Washington, DC: U.S. Department of Education, Office of Planning, Evaluation
and Policy Development. Retrieved from [https://www2.ed.gov/rschstat/
eval/other/single-sex/single-sex.pdf](https://www2.ed.gov/rschstat/eval/other/single-sex/single-sex.pdf)
- Mansfield, K. C. (2015). Giftedness as property: Troubling whiteness, wealth, and gifted
education in the U.S. *International Journal of Multicultural Education*, 17(1),
143-161.
- Matthews, M., Ritchotte, J., & McBee, M. (2013). Effects of schoolwide cluster grouping
and within-class ability grouping on elementary schools' academic achievement
growth. *High Ability Studies*, 24 (2), 81-97.
- McCarter, A. K. (2014). *Ability grouping in elementary education*. (Doctoral
dissertation). Retrieved from ProQuest Dissertations and Theses database.
(Publication No. AAT3002459).
- McCaw, D. S. (2001). *Teaching reading using small flexible-skills grouping and whole
classroom instruction* (Doctoral dissertation). Retrieved from ProQuest
Dissertations and Theses database. (Publication No. AAT 3006623).
- McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United
States today: A look at state definitions, policies, and practices. *Journal of
Applied School Psychology*, 28(1), 59-88.
- McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006, July/August)). Ability grouping
across kindergarten using an early childhood longitudinal study. *Journal of
Educational Research*, 99 (6), 339-346.
- Merton, R. K. (1968). The self-fulfilling prophecy. *The Antioch Review*, 8(2),

- 193-210. Retrieved from <http://www.jstor.org/stable/460926>
- Miles, J., & Shelvin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Thousand Oaks, CA: Sage.
- Mirel, J. (1999). *The rise and fall of urban school system: Detroit 1907-1981*. Detroit, MI: University of Michigan Press.
- Mosteller, F., Light, R., & Sachs, J. (1996, December). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review*, 66(4), 797-843.
- Nagel, G. K. (2001). *Effective grouping for literacy instruction*. Boston, MA: Allyn and Bacon.
- National Center for Educational Statistics. (2016, May 12). *National Assessment of Educational Progress*. Washington, DC: Author. Retrieved from <http://nces.ed.gov/nationsreportcard/>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction--reports of the Subgroups*. Retrieved from https://lincs.ed.gov/publications/html/prfteachers/reading_first_print.html
- Neal, D., & Schanzenbach, D. W. (2007). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2), 263-283.
- New Jersey Department of Education. (2013a). *New Jersey assessment of skills and knowledge (NJASK) 2013 Score Interpretation Manual*. Trenton, NJ: Author. Retrieved from <http://www.nj.gov/education/assessment/ms/5-8/ref/SIM13.pdf>

- New Jersey Department of Education. (2013b). *New Jersey assessment of skills and knowledge (NJASK) 2013 Technical Report*. Trenton, NJ: Author. Retrieved from http://www.nj.gov/education/assessment/es/njask_tech_report13.pdf
- New Jersey Department of Education. (2014). *New Jersey assessment of skills and knowledge (NJASK) 2013-2014: A summary of core test design and administration*. Trenton, NJ: Author. Retrieved from <http://www.state.nj.us/education/assessment/es/njask/NJASKPres.pdf>
- New Jersey Department of Education (2014, August 15). *District factor groups*. Trenton, NJ: Author. Retrieved from <http://www.state.nj.us/education/schools/achievement/dfg.htm>
- New Jersey Department of Education. (2015, August 15). *School Finance: District Factor Groups (DFG) for School Districts*. Trenton, NJ: Author. Retrieved from <http://www.state.nj.us/education/finance/rda/dfg.shtml>
- New Jersey Department of Education. (2016, July 27). *New Jersey Assessment Reports (2016: July 27)*. Trenton, NJ: Author. Retrieved from <http://www.state.nj.us/education/schools/achievement/prior.htm>
- New Jersey Proficiency Assessment of State Standards. (2015, September 12). Boston, MA: Houghton, Mifflin, Harcourt. Retrieved from <http://www.hmhco.com/hmhco.com/hmh-assessments/achievementnjpass#content-information>
- NICHD Early Child Care Research Network (Ed.). (2005). *Child care and child Development: Results from the NICHD study of early child care and youth development*. New York, NY: Guilford Press, 2005.
- Nieto, S., & Turner, K. C. N. (2012). Learning local and global literacies from students

- and families. *Language Arts*, 90(1), 66-69. Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/1327227342?accountid=13793>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 20 U.S.C. 6319 (2001).
- Nomi, T. (2006). *Educational stratification in early elementary school: The casual effect of ability grouping on reading achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No. AAT 3231873).
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.
- Oakes, J. (2008). Keeping track: Structuring equality and inequality in an era of accountability. *Teacher College Record*, 110(3), 700-712.
- Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific studies of reading*, 16(2), 91-121.
- O'Malley, K. J., Murphy, S., McClarty, K, L., Murphy, D., & McBride, Y. (2011) Overview of Student Growth Models (Pearson white paper). Retrieved from http://www.pearsonassessments.com/hai/Images/tmrs/Student_Growth_WP_083111_FINAL.pdf
- Opitz, M. F. (1998). *Flexible grouping in reading: Practical ways to help all students become better readers*. New York: Scholastic Professional Books.
- Otto, H. J. (1932). *Current practices in the organization of elementary schools*. Evanston,

- IL: Northeastern University School of Education.
- Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin, 140*(4), 1042.
- PARCC states. (2016, July 1). Retrieved from <http://www.parcconline.org/about/states>
- Pearson. (2011) *Developmental reading assessment second edition, K-8 technical manual*. Retrieved from http://assets.pearsonschool.com/asset_mgr/current/20139/DRA2_Technical_Manual_2012.pdf
- Pearson. (n.d.). *DRA2 K-3 In-service Guide Update 2nd Edition*.
- Petrilli, M. (2011, Winter). All together now? Educating high and low achievers in the same classroom. *Education Next, 11*(1), 48-55.
- Pressley, Michael. (n.d.). Differentiated learning: Teaching every child. *ReadingRockets* [Web podcast]. Retrieved from <http://www.readingrockets.org/webcasts/1001#program>.
- Puzio, K., & Colby, G. T. (2010). The effects of within class ability grouping on reading achievement: A meta-analytical synthesis. Evanston, IL: Evanston Society for Research in Education Effectiveness.
- Puzio, K. & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness, 6*(4), 339-360.
doi:10.1080/19345747.2013.775683
- Radencich, M. C., & McKay, L. J. (1995). *Flexible grouping for literacy in the elementary grades*. Boston, MA: Allyn & Bacon.
- Rist, R. C. (1970). Student social class and teacher expectations: The self-

- fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40 (3), 411-451.
- Rosenbaum, J. E. (1976). *Maintaining inequality: The hidden curriculum of high school tracking*. New York, NY: Wiley.
- Rosenblatt, L (1978). *The reader, the text, the poem: The transactional theory of literary work*. Carbondale, IL: Southern Illinois University Press.
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, 76(3), 429-444.
- Sax, L. (2009). *Boys adrift: The five factors driving the growing epidemic of unmotivated boys and underachieving young men*. New York, NY: Basic Books.
- Schwabe, F., McElvany, N., & Trendtel, M. (2015). The school age gender gap in reading achievement: Examining the influences of item format and intrinsic reading motivation. *Reading Research Quarterly*, 50(2), 219. Retrieved from <http://search.proquest.com.ezproxy.shu.edu/docview/1669897630?accountid=137>
- 93
- Scigliano, D., & Hipsky, S. (2010) The three ring circus of differentiated instruction. *Kappa Delta Pi Record*, 46(2), 82-86.
- Shaywitz, S. E. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. Knopf.
- Shaywitz, S., Stuebing, K., Shaywitz, B., Fletcher, J. & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution for reading ability. *New England Journal of Medicine*, 326(3), 145-150.

- Slavin, R. E. (1986). Ability grouping and student achievement in elementary schools: A best-evidence synthesis, report 1. Center for Research on Elementary and Middle Schools. Retrieved from ERIC database. (ED 348174)
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293-336.
- Slavin, R. E. (1988) Synthesis of research on grouping in elementary and secondary schools. *Educational Leadership*, 46(1), 67-77.
- Slaydon, D. (2013). *Case study of implementation of flexible grouping in one school framed within the change based adoption model* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No. AAT 3562345).
- Smarter Balanced Assessment Consortium. (2016, July 1). Smarter balanced members and governance. Retrieved from <http://www.smarterbalanced.org/about/members/>
- Sondergeld, T. A., & Schultz, R. A. (2008). Science, standards, and differentiation: It really can be fun!. *Gifted Child Today*, 31(1), 34-40.
- Sorensen, A. 1970. Organizational differentiation of students and educational opportunity. *Sociology of Education*, 43, 355-376.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.
- Tach, L., & Farkas, G. (2006). Learning-related behavior, cognitive skills, and ability grouping when school begins. *Social Science Research*, 35(4),1048-79.
- Tienken, C. H. (2012). The influence of poverty on achievement. *Kappa Delta Pi Record*,

- 48(3), 105-107.
- Tieso, C. (2003). Ability grouping is not just tracking anymore. *Roeper Review*, 26(1), 29-36.
- Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted*, 29 (1) 60-89.
- Tomlinson, C. (2003). *Fulfilling the promise of the differentiated classroom*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Tomlinson, C. (2005). Differentiating instruction: Why bother? *Middle Ground*, 9(1),12-15. Retrieved from <http://www.amle.org/Publications/MiddleGround/August2005/Article1/tabid/339/Default.aspx>
- Tomlinson, C. A., & McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. Alexandria, VA: ASCD.
- U.S. Department of Education. (2003). Preliminary overview of programs and changes included in the No Child Left Behind Act of 2001. Retrieved from http://www.ed.gov/nclb/overview/intro/progsum/sum_pg2.html.
- U.S. Department of Education. (2004, October). Guide to education and No Child Left Behind. Retrieved from http://www2.ed.gov/nclb/overview/intro/guide/guide_pg12.html
- U.S. Department of Education (2010, March). Blueprint for reform: The reauthorization of the elementary and secondary education act. Retrieved from <http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf>

- U.S. Department of Education. (2014, February 13). States granted waiver for No Child Left Behind: Allowed to reapply for waiver.
Retrieved from <http://www.ed.gov/news/press-releases/states-granted-waivers-no-child-left-behind-allowed-reapply-renewal-2014-and-2015-school-years>
- U.S. Department of Education. (2014, February 14). NCLB flexibility and waivers.
Retrieved from <http://www2.ed.gov/nclb/freedom/local/flexibility/index.html>
- U.S. Department of Education (2014, August 3). U.S. Secretary of Education Duncan announces winners of competition to improve student assessments. Retrieved from <http://www.ed.gov/news/press-releases/us-secretary-education-duncan-announces-winners-competition-improve-student-asse>
- U.S. Department of Education. (2015, August 23). *Improving basic programs operating by local educational agencies (Title I, Part A)*.
Retrieved from <http://www2.ed.gov/programs/titleiparta/index.html>
- U.S. Department of Education. (2015, August 23). *Title III Part A programs-strengthening institutions*.
Retrieved from <http://www2.ed.gov/programs/iduestitle3a/index.html>
- U.S. Department of Education. (2015, November 20). ESEA flexibility. Retrieved from <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html>
- Valentino, K. (2000). Flexible grouping. Research articles from Houghton Mifflin.
Retrieved from <http://www.eduplace.com/science/profdev/articles/index.html>
- Vanneman, A., Hamilton, L., Anderson, J. B., & Rahman, T. (2009).
Achievement gaps: How black and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress

- (NCES 2009-455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences.
- Venezky, R. L., & Bregar, W. S. (1988). Different levels of ability in solving mathematical word problems. *The Journal of Mathematical Behavior*, 7(2), 111-134.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press.
- Waldfoegel, J., & Washbrook, E. (2011, April 26). Early years policy. *Child Development Research*. Retrieved from <https://www.hindawi.com/journals/cdr/2011/343016/>
- Weisberger, B. A. (1994). A nation of immigrants. *American Heritage*, 45(1), 75.
- Wilkinson, I. A. G., & Townsend, M. A. R. (2000). From rata to rimu: Grouping for instruction in best practice New Zealand classroom. *The Reading Teacher*, 53(6), 460-71.
- William, D., & Bartholomew, H. (2004). It's not which school you're in but what set you're in that matters: The influence of ability grouping practices on student progress in mathematics. *British Educational Research Journal*, 30(2), 279-293.
- Witte, R., & Witte, J. (2010). *Statistics*. Hoboken, NJ: John Wiley & Sons.
- Wormelli, R. (2007). *Differentiation*. Portland, ME: Stenhouse Publishing.
- Worthy, J. (2010). Only the names have been changed: Ability grouping revisited. *Urban Review*, 42(4), 271-295.
- Wright-Castro, R., Ramirez, R., & Duran, R. (2003). Latino 6th grade student perceptions of school sorting practices. Retrieved from ERIC database. (ED477546)

Wruble, R. (2002). *Great grouping strategies*. Teaching Resources.

Yadegari, S. A., Ryan, D. A. (2002). Inspiring struggling readers. *Principal*, 81(3), 26-28.

Yoon, S. Y., & Gentry, M. (2009). Racial and ethnic representation in gifted programs: Current status of and implications for gifted Asian American students. *Gifted Child Quarterly*. doi: 10.1177/0016986208330564