Summer March 2016

Literacy Achievement in Nongraded Classrooms

Anita Therese Kreide
Loyola Marymount University

Summer July 2011

Literacy Achievement in Nongraded Classrooms

Anita Therese Kreide
Loyola Marymount University, kreidea@yahoo.com

LOYOLA MARYMOUNT UNIVERSITY

Literacy Achievement in Nongraded Classrooms

by

Anita Therese Kreide

A dissertation presented to the Faculty of the School of Education,

Loyola Marymount University,

in partial satisfaction of the requirements for the degree

Doctor of Education

2011

Loyola Marymount University
School of Education
Los Angeles, CA 90045

This dissertation written by Anita Kreide, under the direction of the Dissertation Committee, is approved and accepted by all committee members, in partial fulfillment of requirements for the degree of Doctor of Education.

Date: March 28, 2011

Dissertation Committee

[redacted]
Dr. Candace Poindexter, Committee Chair

[redacted]
Dr. Karen Huchting, Committee Member

[redacted]
Dr. Laurie Love, Committee Member

ACKNOWLEDGEMENTS

It is a pleasure to thank those who have made this thesis possible:

To my committee, it is an honor for me to have been blessed with such an amazing committee of intelligent and powerful women. You each have been a constant source of support and inspiration on my journey. I owe my deepest gratitude to my chair Dr. Poindexter, whose wisdom guided me from beginning to end. Without her advice on direction, the enormity of the topic would have prevented its completion.

To my cohort family, I have learned and grown so much over these past years. Yes, partially from the knowledge presented in the class, but entirely from the depth of experience and wisdom each of you have shared with me in class. I am a better person for having known each of you. Your spirit and drive will resonate with me as I travel through life trying to make a difference in the lives of children.

To my beloved family, this thesis would not have been possible without your constant love, encouragement, and support. To my grandparents and parents, who had no doubt I would achieve all of my goals. To my sister in-law who was always willing to help with SPSS quirks, no matter how late. To my parents, in-laws, and best friend, thank you for the unquestioned and numerous hours of babysitting. To my dear cousin who took the past year of her life, dropping everything to help us care for our newborn son. To my husband for understanding me as I am, never trying to change me and always listening to my ideas. To my children who know how to make me laugh during the most

trying of times. You are my reason for choosing this topic and this thesis is dedicated to you William, Alina, and Spencer.

In order to change an existing paradigm you do not

Struggle to try and change the problematic model.

You create a new model and make the old one obsolete.

That, in essence, is the higher service

To which we are all being called.

Buckminster Fuller

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

ABSTRACT ix

CHAPTER 1: INTRODUCTION 1
Background. 1
Statement of the Problem 4
Purpose of the Study 7
Significance of the Study 7
Theoretical Framework 10
Research Question 12
Research Design and Methodology 12
Limitations and Assumptions 13
Delimitations 14
Definition of Terms 14
Summary 16

CHAPTER 2: REVIEW OF THE LITERATURE 18
Introduction 18
History 18
- Business Interest in Graded Education 19
- Legislation and Graded Education 20
- Legislation and Nongraded Education 22

Theoretical Framework 25
- Reproduction Theory 25
- Individualistic Pedagogy 26
- Relational Pedagogy 27
- Micropolitical Perspective 28

Nongraded Education 29
- Noncognitive Rationale 29
  - Retention, Promotion, and Ability Tracking 30
  - Cooperation Versus Competition 32
  - Peer Tutoring 33
  - Environment 34
  - Socioemotional Development 35
  - Noncognitive Summary 36
- Cognitive Rationale 37
  - Definitions 37

Quantitative Research Studies 38
Literacy 44
Theory of Reading Development....45
CST/ELA Measure....48
Summary…….. 52

CHAPTER 3: METHODOLOGY....55
Introduction....55
Research Question and Hypothesis....55
Methodology....55
Measure....56
Sampling method 58
Site 58
Program....59
Population 61
Matched Control Populations 64
Procedures....65
Data Analysis....66
Summary....66

CHAPTER 4: RESULTS 67
Introduction....67
Research Question 67
Analytical Plan....67
Results 68
Entire Population 68
Matched Control Populations 72

CHAPTER 5: CONCLUSIONS AND IMPLICATIONS 76
Introduction....76
Conclusions....76
Recommendations....80
Implications....81

REFERENCES....85

LIST OF TABLES

1. CST English-Language Arts State Ranges 57

2. Comparison of Graded and Nongraded Student Populations. 62

3. Comparison of Matched Control Graded and Nongraded Student Populations 65

4. Comparison of Graded and Nongraded Students on CST/ELA Measure.....69

5. Comparison of Matched Control Graded and Nongraded Students on CST/ELA Measure from 2005-06 to 2009-10 School Years.....73

ABSTRACT

Literacy Achievement in Nongraded Classrooms

By

Anita Therese Kreide

This longitudinal quantitative study compared literacy achievement of students from second through sixth grade based on two organizational systems: graded (traditional) and nongraded (multiage) classrooms. The California Standards Test (CST) scaled and proficiency scores for English-Language Arts (ELA) were used as the study's independent variable to measure student performance. A matched control was utilized in which nongraded students were compared with graded students based on gender, ethnicity, and date of birth. Data analysis included independent samples t-test, analysis of variance (ANOVA), and effect size. Results showed that nongraded students had a significant advantage over their graded counterparts in literacy achievement ($p=0.000$). Effect size for the matched group increased with length of exposure in the nongraded program from Cohen's $d=0.49$ to $d=0.99$. It is difficult to determine if significant outcomes were the result of classroom structure or instructional strategies used in the

nongraded setting. However, a unique quality of this study involves the rare conditions and matched control design that allowed for variables to be controlled, which have yet to be simultaneously accounted for in multiage studies to date. Based on the results, this study suggested that nongraded education, by responding to the developmental nature of children in the classroom, may offer a viable alternative to the graded system. In nations such as Australia, New Zealand, Netherlands, Finland, and Canada with the highest literacy rates in the world, nongraded classrooms are common educational practice.

CHAPTER 1

INTRODUCTION

Background

The term nongraded may invoke thoughts of students who are not assigned letter grades (A-F). Instead nongraded refers to classrooms without grade level designation (1st -12th) in which students receive letter grades, are educated by the same teacher/s for three or more years, and whose classmates span three to four years in age. In contrast graded education (K-12) is the traditional and most common form of grouping students. It is based solely on chronological age and groups students into corresponding classroom levels. Originally designed to be consistent with the industrial model, in order to meet the needs of mass production during the industrial revolution, graded education has become a standard in American culture since its introduction in the 1840s by politician and educator Horace Mann (Gutek, 1986; Kasten, 1998; Osin & Lesgold, 1996; Rogoff, Paradise, Arauz, Correa-Chavez, & Angelillo, 2003; Stone, 2009).

The public graded system was established with predetermined standards, curriculum, letter grades, and retention/promotion practices in an effort to boost test scores and homogenize student outcomes (Aina, 2001). Standardization of children in school districts nationwide provides a competitive-comparative student performance data evaluation system which is required under the federal *No Child Left Behind* (NCLB) legislation, allowing for assessments of educational effectiveness and, according to Anderson (1993), fuels the American culture of competition (Eisner, 2005).

The performance data of schools is based on conjecture: the efficacy of the graded structure of schools. In fact Anderson (1993), a leading educational researcher finds it "strange that the graded school…and its relatively primitive assumptions about human development and learning, has held its ground this long" (p.10). Anderson and other researchers (Elkind, 1987; Kasten, 1998; Levenson, 1977; Williams & Strangis, 2002) agree the current graded organizational framework does not support theories of child learning and development.

In contrast, the nongraded system of education abides by the fundamental developmental learning theories of Vygotsky, Piaget, Gardner, and Bandura. As such, nongraded education responds to the vast educational research promoting the efficacy of developmentally appropriate practice (DAP).

DAP as coined in 1986 by the National Association for the Education of Young Children (NAEYC) is defined in their 2009 position statement as, "grounded both in the research on child development and learning and in the knowledge base regarding educational effectiveness… practice(s) that promote…optimal learning and development" (p. 1). Based on the nature of children, DAP is focused on biological as opposed to physical time. DAP acknowledges children grow at different rates not only in regards to physical measurements such height: but also in their mental and psychological processes (Elkind, 1987).

Nongraded education accommodates children at different levels of maturity and ability by providing differentiated individual instruction in a cooperative environment. However, in the current graded system "uniformity and efficiency become hallmarks" of

education in America (Eisner, 2005, p.17). According to Eisner, in the United States students are raced towards an academic finish line, modeling the competitive culture of society, where the winners are considered to be more intelligent. Can the values of a competitive dominant culture, which favors speed and competition, be removed from education? Will the pace of each child be enough to satisfy our educational system? Will structural practices change to meet the biological needs of children?

The No Child Left Behind Act (NCLB) advocates the need to reach all students and measures student outcomes by quantitative data provided mainly from standardized testing. Therefore, it becomes clear that in order to investigate structural models in education such as the nongraded system, quantitative research must be done utilizing standardized test scores. The focus of this dissertation aims to contribute to the nongraded program literature utilizing the cognitive domain of reading achievement in order to assess efficacy. This study's focus on literacy does not dismiss the extreme importance of the social-emotional realm in education, nor deny other factors such as culture, community, parents, teachers, socioeconomic levels, race, special education, English language learners, student health concerns, class size, and school funding to name a few. Rather, reading in particular was chosen as a focus due to its importance as a gateway skill toward student access to other subject areas (McIntyre et al., 2005) and because of No Child Left Behind (2001), which set goals for basic literacy by the end of the third grade through the Reading First Initiative (RFI).

Statement of the Problem

United States industrialization in the 1800s brought social and economic problems including increases in children requiring schooling, the need for managers and skilled labor, and a call for homogenization of a population that was considerably more heterogeneous and non-English speaking due to immigration (Gutek, 1986). As a result, business proprietors and mainstream society had a convergence of interest in perpetuating the establishment of the graded school system, in part to curb the side effects associated with the industrial revolution. Education became co-opted by industry, and mass production was replicated in the graded school system. In Gramsci's (1971) view, when a state uses age as a means to separate the education of students, its motive is reflected in the general economic conditions of the time where the immediate need for the profitable contribution of youth is preeminent.

According to Kasten (1998) children are considered commodities similar to cars in the factory model. For example, the term *superintendent* was first used in factories and quality control checks were called *promotion* or *nonpromotion*. The child proceeded from grade level to grade level, similar to a car's movement down the assembly line, where at the end non-uniform vehicles where discarded (dropped out). The danger when applying industrial methods to education is that assembly lines utilize homogeneous inputs which undergo the same processes to produce uniform outputs (Kasten, 1998; Katz, Evangelou, & Hartman, 1990; Osin, 1996). However, three injustices are present in industry's graded education model: the identical treatment of all children, the

perpetuation of age segregation, and the mismatch between the means and the goals of No Child Left Behind.

First, humans will never be homogeneous as predetermined by their unique genetic makeup of deoxyribonucleic acid (DNA) and differing environmental stimuli. No two humans are alike, so why does graded education insist on uniformity? Children are not cars but complex individuals with their own educational needs, yet governments apply simple factory analogs which run counter to current educational research. It is unjust to treat children as identical inputs and expect standardized outcomes. Where is the acknowledgement of the individual? In 1890 Charles Eliot, the president of Harvard, noted that graded education was the "worst feature of the American school" because it flew "in the face of nature" and schooling should have the "utmost possible account of individual instruction" (Hamilton, 1989, p.132). William Kilpatrick, a professor at Columbia University in 1925, cited failure in graded education because it did not acknowledge the individual in favor of the institution and stated, "As always, the easiest solution was to hold to the institution and let the individual suffer. So we did" (as cited in Osin, 1996, p.631).

The second social injustice of graded education comes in the form of the perpetuation of age segregation. In a recent paper on ageism, Hagestad and Uhlenberg (2006) define graded schools as institutional segregation, where children spend most of their time with a narrow age range of peers working in a setting exclusive of younger and older. Age segregation is an industrial phenomenon and is not consistent with anthropological, ethological, and developmental biology/psychology which all point to

the innate nature of age-stratification in learning (Konner, 1975; Pratt, 1986). Ellis, Rogoff, and Cromer (1981) have found negative effects in same age groupings including aggressive and antisocial behavior.

Finally, if the social obligation of schooling under NCLB is to reach all students without rejections, then the means of the factory scheme can never accomplish this goal. The graded industrial framework is developed to remove students from education who do not respond to standardized education. This is achieved via grade level testing, which leaves children behind who are not developmentally ready to be pushed through a system of promotion and retention without full consideration of the child's innate cognitive and social emotional needs. Where is the response to children's need for individual development? The system searches for outliers, and a bell shaped curve will always illustrate failure. Standardized testing under NCLB produces winners and losers, and as Eisner (2005) stated, "We Americans are very much interested in our position in a distribution. Doing well means, in practical terms, doing better than one's neighbors" (p.14). The age old expression where a square peg will never fit in a round hole applies to the means-end model of our current educational system. Graded systems by their very nature cannot achieve success because the means by which they are structured do not follow the goal of leaving no child behind.

Can nongraded education produce measurable outcomes? In nations such as Australia, New Zealand, Netherlands, Finland, and Canada with the highest literacy rates in the world, nongraded classrooms are common and even mandated (New Zealand and Netherlands Antilles) educational practice (Aina, 2001; Song, Spradlin, & Plucker,

2009). However in the United States, according to Song, Spradlin, and Plucker in a 2009 education policy brief, there is little research on "outcomes and benefits" (p.6) of nongraded education. Along with the lack of studies, according to Marshall and Gerstl-Pepin (2005), "Quantitative approaches are particularly powerful and useful for identifying larger, more sweepingly oppressive structures" (p. 95) like those found in graded education. The present study aims to speak the quantitative language of NCLB by using standardized test results to measure literacy in the nongraded classroom and provide data necessary to investigate alternatives to current graded educational policy.

Purpose of the Study

The purpose of this longitudinal quantitative study was to compare the California Standards Test (CST) scaled and proficiency scores for English-Language Arts (ELA) from second through sixth grade based on two organizational systems: graded (traditional) and nongraded education. Ex post facto data for the causal-comparative design, nonrandom purposive sampling included nongraded students enrolled consecutively in the same public school district, in a middle to upper middle class suburb of Southern California. The nongraded students were compared with their matched counterparts at the same school in graded classrooms on the dependant variable measurement of the California Standards Test (CST) for English-Language Arts (ELA).

Significance of the Study

This study examined the graded public education system, where children under NCLB were asked to learn more, learn sooner, and enter school prepared or "face sanctions that include summer school and retention" (Schulting, Malone, & Dodge,

2005). For those students whose chronological age does not match their developmental age, very few programs will wait without consequence until the student matures. Biology is denied in light of the public interest for production of workers in a capitalistic society. Nongraded education allows for individualized instruction responsive to the each child's developmental needs, without having to resort to retention or promotion (Anderson & Pavan, 1993; Goodlad & Anderson, 1987; Lloyd, 1999).

The research is essential because, while nongraded education programs have been reviewed in the literature, according to Song et al. in a 2009 Education Policy Brief, "little research exists on the outcomes and benefits…much of which is quite dated, and additional research from the mid-90's provides mixed results" (p.6). The current study examined nongraded program efficacy relating to student achievement in literacy. Reading proficiency, an integral component of cognitive development, is a necessary benchmark for any educational program (Horm-Wingerd, Winter, & Plotchan, 2001). Therefore, with the need to respond to children's developmental needs via nongraded education, it is essential to use reading attainment as a lens for evaluation.

Another benefit of this study is its responsiveness to the current climate of testing and accountability. According to the Obama administration, the awaiting reauthorization of NCLB will encompass "Race to the Top" reforms which include higher quality standards and assessments, attracting and retaining quality teachers and leaders, improving data systems to assess and drive instruction, recovering "struggling schools," and "improving conditions favorable to innovation and reform" ("Fact Sheet: The Race to the Top," 2009). Based on the existing direction of education reform, this study responds

to legislations' need for data driven assessment and reform. Utilizing quantitative data from standardized tests required under NCLB, the current research speaks the language of NCLB restructuring. It also investigates the potentially innovative approach to education restructuring via the nongraded education model.

Nongraded education may not be popular in California education reform, but it does present an opportunity to schools with nongraded programs. Currently districts with nongraded classrooms fall under the state classification of alternative education. Schools within this classification represent one of the few programs in California to qualify for $4.35 billion in "Race to the Top" funds ("Fact Sheet: The Race to the Top," 2009). Given the economic recession and demand for educational reform from the current administration, quantitative research on nongraded education might help to reignite interest in the field and support nongraded education through federal funding. As a result, the nationwide legislative, educational, and economic environment provides an ideal context for this study.

For teachers and educational leaders this study offers a new perspective toward reversing the popular opinion that positions graded education as the gold standard. To reform the graded system it is necessary to educate society and research structural alternatives, such as nongraded education, which are more aligned with the developmental nature of children.

The results of this study are meant to inform the School District under study as to the efficacy of their nongraded program and to inform state and federal policy makers, school districts, and parochial schools as to the possible alternatives to the graded

educational system. Through reviewing the history and research context of the current graded school system, this study examined the possible opportunities offered by nongraded education to foster individual biological growth and development unfettered by a chronological age agenda of the political system.

Theoretical Framework

The underlying theories for this research are viewed through the lens of reproduction theory, individual versus relational ideology, and micropolitical perspective. Vandewalker, as early as 1908, stated that graded education needed to be altered because "education is a process of development rather than a process of instruction" (as cited in Bryant & Clifford, 1992, p.148). Yet, over 100 years later the nature of childhood development is not modeled in the current graded system, why?

Reproduction theorists such as Bourdieu and Passeron (1990) would respond that the longer the norm of classrooms being graded is adhered to, the more it becomes a dominant social ideology. Organizational structures other than graded classrooms are easily dismissed as they do not resonate with society's social experience, making graded education a sustained dominant practice that has become legitimized over the past generation (Apple, 1980; Bourdieu & Passeron, 1990; Bowles & Gintis, 1977; Giroux, 1983).

Residing within the continuation of the graded model is the perpetuation of an individualistic ideology, where students are viewed as entities unto themselves void of interconnection with their surroundings. Individualism flowed over into the American educational system via gradation due to the industrial revolution, where people were

viewed as commodities separated from the results of their labor and their relationship to society (Marx & Bottomore, 1964). According to Kasten (1998) industrial processes were replicated in the schools, and the ways humans learned and developed became separated from education. Greene (2008) a leading sociologist contributed "negative social consequences" to the ideology of individualism (p.117).

In stark contrast to individualism is relational pedagogy present in nongraded education. Relational theories are rooted in child development where scholars such as Marx, Gramsci, Nodding, and Vygotsky believe education is indivisible from the context of societal relationships. According to Vygotsky (1978) a leading cognitive development theorist, society, culture, and history taken together provide the framework of knowledge. He theorized that children learn best from their peers of varying ages and knowledge. It is through this relational pedagogy found in nongraded classrooms that children educate each other via Vygotsky's zone of proximal development (ZPD). ZPD is the zone of a child's knowledge that is tapped into and expanded by a peer to reach a new level of knowledge that could not have been accessed by the child alone.

If relational concepts of education, including nongraded schools, are to be utilized as an option to the current educational system, the ideologically individualistic and normative system of graded schools needs to be revisited through a micropolitical perspective. Micropolitics enables the graded system to viewed at its most basic level; its historical inception. Through the exposure of graded education as a system devoid of the child developmental context and established to meet the needs of industry, a voice can be

given to alternatives that may challenge the dominant cultural practice of the graded system.

Research Question

The research question is, What effect do nongraded classrooms have on students' CST scores in literacy achievement? The hypothesis predicts students enrolled in nongraded classrooms will outperform their traditional graded peers in literacy achievement.

Research Design and Methodology

The research design utilized for the study was associational in nature, employing the causal-comparative criterion-group research design to explore the ex post facto nature of the relationship between the nongraded and traditional graded programs. The *criterion* group consisted of students grades two through six who completed the nongraded program. The *comparison* group consisted of students grades two through six from the same school who participated in traditional graded programs. The graded and nongraded comparison groups were matched based on gender, ethnicity, and date of birth. In this study the independent variable was the nongraded program and the dependant variable was the measurement of the California Standards Test (CST) for English-Language Arts (ELA).

For this longitudinal causal-comparative design, nonrandom purposive sampling was selected to include nongraded students who were consecutively enrolled in the same public school district, in a middle to upper middle class suburb in California. Since the CST for ELA testing begins in the second grade during the months of April-May, former

nongraded students in grades two through six, who completed the CST/ELA were included in the study. Students beyond sixth grade were excluded from the study because the CST, developed as a result of No Child Left Behind (NCLB), had not yet been proctored as a statewide assessment. The nongraded sample was compared with matched controls from the same school's graded program to better control for variables.

Students from grades two through six were measured based on their scores from the criterion-referenced CST in ELA, a test aligned to measure the state of California content standards. Data collected from the school district regarding student demographics and CST scores in ELA for the nongraded and graded programs were obtained with permission of the school district.

Limitations and Assumptions

Based on availability of evidence, the research in this study was limited to CST/ELA scores from a single public school in Southern California over the past five years. The placement of students into the nongraded program was made solely by the parent/legal guardians of the child and could occur from outside the school's boundaries. The nongraded program accepts students using both intra and inter district permits, possibly creating a broader range of regions included in the population. The graded program at the same school was mostly limited to students within the school's jurisdiction, although some district transfers did occur. The school site represented a predominately middle class Caucasian demographic which may limit the applicability of results. Also, access to the socioeconomic status (SES) of the subjects was not provided by the district and limited comparisons on this factor.

Within the classroom setting several restrictions were present. First, as Lloyd (1999) suggested, "Classrooms are not static entities and teachers can vary in their approach while maintaining the same education philosophy" (p.190). While variability in teacher effectiveness can exist, teachers in graded and nongraded classrooms at the school site were comparable in competency, both maintaining clear teaching credentials.

Delimitations

The data measurements in the study were delimited to the use of California Standards Test (CST) for English-Language Arts (ELA). Standardized assessment of reading achievement was chosen as it is considered a "gateway" subject. Also ELA results respond to the need for quantitative test score data required by the data driven program decision analysis standard under NCLB and continued under "Race to the Top" funding. This study was completed in only one school district. This was due in part to access of information, but also very few nongraded programs existed at the time of this study in the state of California.

The longitudinal nature of study data was also delimited by the researcher. The data timeframe was set at five years, when scores for the CST were initiated via NCLB. Other literacy measures were available pre-NCLB, but uniformity of measure was a concern of the study to maintain reliability.

Definition of Terms

The following definitions provide uniformity and are utilized in this research study.

California Standards Test (CST) *for English-Language Arts* (ELA): California criterion-referenced CST in ELA is a test aligned to measure the state of California content standards in English language. The test is given in California public schools grades two thru eleven and includes components such as word analysis, reading comprehension, literary analysis, standard English conventions, and writing strategies (*California Department of Education [CDE],* 2009).

Developmentally Appropriate Practice (DAP): A theory grounded in research to promote optimal child development based on three core components: knowledge of child development and learning, knowledge of the individual child, and knowledge of the social and cultural context of the child ("Developmentally Appropriate Practice", 2009).

Horizontal Grouping: Placement of children into a grade level based solely on chronological age (Mason & Burns, 1996).

Graded Education: A traditional school classroom where children are separated by age and placed into a corresponding grade level. Children are promoted or retained to the next grade level based on performance as referenced by standardized curriculum and testing. Generally graded students receive new teachers and classmates each year for kindergarten through twelfth grade (Lloyd, 1999).

Multi-age Education: A term often used synonymously with nongraded education. In multi-age classrooms students receive instruction from the same teachers for three or more years and are with peers who span a three to four year age range, for example, K-1-2-3, 3-4-5, and 5-6-7. Classrooms are developed with the intent of multi-age education and are not created out of economic necessity. However, the philosophy of

education which supports the developmental nature of the child is not explicit in multi-age classrooms and students consider themselves part of a certain grade (Goodland & Anderson, 1987).

Multi-grade/ combination classrooms: Created by administrations to level out class sizes and control uneven grade level enrollments. Similar to traditional graded education except two or more grades are combined into one classroom, taught separately by one teacher. Students generally do not have the same teachers more than one year (Veenman, 1995).

Nongraded Education: A classroom where students receive instruction from the same teachers for three or more years and are with peers who span a three to four year age range. The nongraded classroom is established based around a philosophy of education which supports the developmental nature of the child, where teachers make curricular adjustments to individualize instruction, and content is completed at the child's advancement rate without regard for time constraints. In nongraded instruction students are seen as members of a classroom not a grade (Anderson & Pavan, 1993).

Vertical grouping: Placement of children into classes of varying ages and grades.

Summary

This study aimed to analyze quantitative longitudinal data comparing student literacy achievement in graded versus nongraded educational systems. In Chapter 1 a brief outline is presented on the background, problem, significance, and conceptual framework of the research. Chapter 2 reviews the history and provides a critical analysis of current research available on the topic. Chapter 3 reveals the procedural components

enacted upon the quantitative data utilized for the study. Chapter 4 provides a presentation and analysis of the results of the research. Chapter 5 discusses the significance of the findings and offers direction for future research in nongraded educational studies.

CHAPTER 2

LITERATURE REVIEW

Introduction

Formal schooling as it reaches around the globe is arguably one of the most influential sociological extensions in the world (Meyer, 1977). In the United States the federal expansion of education has increased public funding and regulation of schools through the No Child Left Behind Act (2002) and the federal "Fact Sheet: Race to the Top" (2009). In these documents school districts are faced with stringent standardized regulations and testing to demonstrate student achievement, without explicitly stating educational or organizational strategies to meet individual student needs. This chapter discusses nongraded education as an option supporting the goal of increased student achievement expressed by the federal government. It presents research and historical aspects of the nongraded educational system and establishes a context and validation for the current study. This is vital because it is important to examine "the historical spaces of the past and present in which that knowledge is socially constructed" (Popkewitz, 1998, p. 535).

History

The nongraded classroom was the first hallmark of American education, in the form of one room school houses across the country. However, in the early 1800's the industrial revolution increased the gross national product and American productivity. Advanced economic growth brought with it social and economic problems. According to Gutek (1986), a leading educational historian, industrialization increased the numbers of

children in urban populations who required schooling, stimulating a need for managers and skilled labor, and called for homogenization of a population that was considerably more heterogeneous and non-English speaking due to immigration. As a result, business and mainstream society had a converged interest in perpetuating the establishment of the graded school system, as a way to maintain the uniformity required for the industrial revolution (Gutek, 1986).

Business Interest in Graded Education

The lens of commerce provides a perspective for viewing the graded classroom scheme of schools because according to Giroux (1983), "school institutions (can) only be understood through an analysis of their relationship to the state and the economy" (p. 258). The inspiration for graded education seen in the "common school" according to most historians comes from the Prussian school system, which Horace Mann and others considered superior to the American system (McClusky, 1920a, 1920b). Modeling the Prussian system of education which incorporated the "proper classification of scholars," where children were "divided according to ages and attainments, and a single teacher has the charge only of a single class," Mann in 1843 correlated the graded school system to current industrial practices (as cited in Pratt, 1986, p. 112). Mann described the similarities between graded education and the factory model saying:

> The principle of the division of labor holds good in schools, as in mechanical industry. One might as justly demand that all operations of carding, spinning and weaving be carried out in the same room, and by the same hands, as insist that children of different ages and attainments should go to the same school and be instructed by the same teacher. (1843, as cited in Pratt, 1986, p.112)

According to Gutek (1986), businessmen and professionals in the northeast were convinced of the need for graded "common schools," which they viewed as a means to provide a stable and skilled work force. They also believed the design of the graded school would teach work ethic and the "old Puritan values of hard work, punctuality, industriousness, and productivity" (Gutek, 1986, p.101). In Gramsci's (1971) view a state uses age as a means to separate the education of students. Its motive is reflected in the general economic conditions of the time, where the immediate need for the productive contribution of youth is imminent.

Legislation and Graded Education

Propelled by economic demands, the proponents of graded "common schools" were businesspersons, scholars, and politicians in Massachusetts in the early 1800's, namely Edmund Dwight, Horace Mann, and George Emerson. It is important to note that Henry Barnard and Calvin Stowe were also advocates, but their influences were felt in Connecticut and Ohio, where they brought graded education to their respective states (McClusky, 1920a). The first legislation to establish graded education was realized in Massachusetts in the 1840s with the approval of the "common school," and the creation of a central school authority via the conception of the positions of a state superintendent of public instruction and a state board of education (McClusky, 1920a). To understand the speed with which the graded school statute became legislation, from Mann's trip to Prussia in 1843 and the opening of the first model graded "common school" in 1848, it is important to follow the reasoning and financing of businessman Edmund Dwight.

Edmund Dwight, a wealthy businessman in Springfield, Massachusetts, who employed in his industries about 3000 workers, began to venture into the textile industry in both Springfield and Boston. A review of the Dwight family records showed their business profits from 1803 to 1820 tripled, "before shrinking over the next three years as the Dwight's began to invest in the textile industry" (Weil, 1998, p.1345). In fact the family's value decreased by one quarter or 26.4% of the total value of all their assets due to textile investment. During the family's financial decline, according to Frederick McClusky (1920b), Mr. Dwight enlisted the services of Horace Mann, who was secretary of the Massachusetts Board of Education and former Massachusetts Senate Representative, by:

> Personally increasing the compensation allowed to him by the state; that Dwight gave $10,000 on the condition that the state appropriate …the establishment of the first normal (common) school; and that on numerous other occasions he contributed important pecuniary aid in carrying forward the designs of the board and of its secretary. (p. 46)

Dwight also mentioned in his memoirs the offering of his house, time, and money to secure the legislation of graded schools in Massachusetts (McClusky, 1920b).

Under Dwight's financing, Mann was able to push graded schools through Massachusetts State legislature in the 1840's along with establishing a publically supported education system through taxation and a centralized graded public school system via the creation of a state superintendent of public instruction and state board of education (McClusky, 1920b). The Quincy Grammar School, the first graded or "common" school in 1848, served as a model for the reform. Within six years every school in Boston replicated the graded Quincy School and soon the graded system was

molded into the Ohio and Connecticut school systems (McClusky, 1920b). This same model is replicated throughout the United States, and includes California's Public School System. The historical development of the current educational structure is crucial because it confirms that graded education was created solely based on the needs of industry, not child developmental and/or educational research (Gutek, 1986; Kasten, 1998; Osin & Lesgold, 1996; Rogoff et.al, 2003; Stone, 2009). This paradigm runs contrary to the philosophy of nongraded education which is centered on developmentally appropriate practice for the student (Elkind, 1987; Katz et al., 1990; Schrier & Mercado, 1994).

Legislation and Nongraded Education

Legislation for graded education as seen in the common school movement of Massachusetts was very specific and spread quickly across the United States as a national model. While no law or policy since the 1900's has been enacted specifically prohibiting graded education, nongraded education has appeared as a component in the educational reform bills in Kentucky and Michigan.

In Kentucky as part of the landmark school finance case of *Rose v. Council for Better Education* (1989), nongraded education was mandated for grades K-4. The case originated in 1984 when 66 out of 178 state school districts joined together to form the Council for Better Education and entered into a law suit against Kentucky's Governor, Superintendent of Public Instruction, General Assembly, and John A Rose (President Pro Tempore of the Senate). The case made its way to the Kentucky Supreme Court in 1989, where the plaintiffs argued that Kentucky's entire school system was unconstitutional

under state constitution section 183 for failing to provide "an efficient system of common schools throughout the state." It was decided by Chief Justice, Robert F. Stephens that the public school system had indeed failed to provide an "efficient school system" which he equated with "adequate education," and declared "Every child…must be provided with an equal opportunity to have an adequate education... The children of the poor and …rich, the children who live in poor districts and …rich districts must be given the same opportunity and access to an adequate education" (p. 216). Therefore because of the "inadequacy" of the Kentucky school system and the "disparity in education (al) opportunity across the state" the massive school reform known as the Kentucky Education Reform Act (KERA) was initiated in 1990. Contained within the act were instructional reforms that included nongraded classrooms for the first four years of elementary school. The nongraded program was chosen because it contained attributes such as developmentally appropriate practices (DAP), multiage/multiability classrooms, continuous progress, authentic assessment, and met individual student needs that were not part of its previous graded program.

In Michigan new legislation, including nongraded education, was also taking place in the early 1990's. As a result of interest by districts throughout the state, in 1993 the State of Michigan through the 87th Legislature passed House Bill No. 5121 section 1278.4 adopting early elementary school initiatives, including the establishment of nongraded classrooms through the fourth grade. While the bill did not mandate nongraded education, it provided a "choice" to school districts and parents with the support of state funding. In 1994 the nongraded program became highly requested by

parents, such that an educational consultant was hired to support district demand (Fox, 1998).

In California nongraded programs are present, but very rare in the public school system. For example, only three nongraded public schools are known to exist in the Southern California area. The establishment of these schools has been a result of parent requests to their respective school districts under California Education Code Section 58502 which states, "A parent or guardian of any pupil may request the Board of Education of a school district to establish an alternative school program or programs in the district pursuant to this chapter." It is interesting to note that according to Brian Uslin, a researcher for the California Department of Education, Section 58502 which started as Assembly bill 3100 and became law on April 30, 1977 was never intended for use as a waiver by the legislature for nongraded education, but instead was joined with three other bills focused mainly on postsecondary education (personal communication, April 23, 2010). This legislation provides the only means by which nongraded education takes place in California public schools.

While education code 58500 on alternative education allows a district to provide nongraded education and provides funding to the district based on the average daily attendance of students, it does not require the district to accept the parent request waiver for nongraded instruction. Therefore, very few programs are established district wide, as nongraded schools entail the support of the district, administration, principal, teachers, parents, students, and staff at each school site.

Theoretical framework

Reproduction Theory

In an effort to understand the structural arrangements in education, this study is grounded in theories which can illuminate and reframe graded education in America. As a result it is important to first recognize the historical context of the graded education system. Business, economic, and political forces have co-opted education to the detriment of students via the replication of industrial mass production in the graded school system. Several leading educational scholars (Apple, 1980; Bourdieu & Passeron, 1990; Bowles & Gintis, 1977; Giroux, 1983) believe when systems such as graded education are perpetuated, it is the result of reproduction theory which suggests schools are merely an extension of the dominant ideology. According to Giroux (1983), with the development of reproduction theory the schools were no longer pure and isolated from their connection to politics. Instead schools "legitimized capitalist rationality and sustained dominant social practices" as illustrated in the perpetuation of graded education in America (Giroux, 1983, p.258). While reproduction theory offers an explanation to the sustainability seen in the graded education system, it does little to transcend and re-envision education.

As a result, this study seeks to reframe education by utilizing postmodernism to deconstruct the historical creation of graded education. The postmodern perspective in "Western" culture began in the 1960s and offers a critique of the epistemology of the modern era in which graded schools were conceived out of industrial society (Grieshaber & Ryan, 2006). Postmodernism draws on deconstruction to remove the assumptions and

ideological foundations behind current societal systems to offer a critique of binary oppositions such as graded versus nongraded education; according to Graue (2005), "Deconstruction enables a reading of the historically created child that shapes …social views" (p. 40).

Individualistic Pedagogy

When exercising postmodernist deconstruction, education is exposed at its core to reveal a historic pedagogy aligned with the individualistic nature of the United States during the industrial revolution. According to Marx and Bottomore (1964) the "individual life" is a "direct consequence of the alienation of man from the product of his labour…man is alienated from other men" (p.129). As a result of production "the worker has become a commodity" (p.69) and is isolated from the relational context of society. This individualistic view of industry according to Popkewitz (1998) is "historically constructed as the effects of power" (p.536) modeled in education via the separation of students by age and grade to form the graded school system. However, according to Gramsci (1971) individualistic pedagogy is not consistent with the "consciousness" of a child which "reflects social and cultural relations which are different from and antagonistic to those which are represented in the school" (p.35). Hence, from a social constructivist perspective student development cannot occur outside a societal context. As reiterated by Engels (1941), "It is not the consciousness of men that determines their being; but their social being that determines their consciousness" (p.72).

Relational Pedagogy

In order to reframe education after postmodern deconstruction, it is necessary to look beyond individualism toward a new relational pedagogy as illustrated in the nongraded system. What would education look like today if history had shifted away from industrial individualism towards a societal construct? The answer is seen in the works of scholars like Marx, Gramsci, Nodding, and Vygotsky among others who consider learning to be inseparable from the context of societal relationships. Noddings (2003) suggested that interaction with others is "ontologically basic" (p.4) and should be at the center of educational ways of knowing. She even questioned graded education when she asked, "Is something wrong with the entire school arrangement?" (p. 13).

For Vygotsky (1978), a Marxist scholar and leading cognitive development theorist, education is a socio-cultural-historical process. He stated that children learn within a "zone of proximal development," where a heterogeneous group of children such as those found in nongraded classrooms, can provide educational scaffolding to one another so new concepts can be learned which alone could not be grasped. According to Moll (1990), a foremost Vygotskian scholar, Vygotsky reminds educators school settings are a social creation and there is nothing "natural" about the current stratifications in the school system.

Vygotsky's zone of proximal development (ZPD) is integral to understanding relational pedagogy and acts as the foundation to the theoretical underpinning which supports this nongraded study. ZPD transcends from the modernist binary approach as seen in the graded system and instead reveals the dialectic nature of education. Holzman

(2006), a postmodernist scholar, stated learning as understood through the relational lens of ZPD is "both the source and the product of learning. As activity, learning and development are inseparably intertwined and emergent, best understood together as a whole" (p.114). In accepting the holistic nature of the child through ZPD, socially just learning can take place in relational pedagogy.

Critics including Langford (2010) have exhibited concern that ZPD is a child-centered pedagogy which ignores equally important issues such as gender, race, and class. However, when viewed through a postmodernist lens, Vygotsky's (1978) idea of "life space" transcended and at the same time included class, gender, and race issues. According to Vygotsky a child's "life space" is where learning happens only in the collaboration with others, establishing a zone of proximal development. Vygotsky (1978) reasoned that the collaborative child-centered space of ZPD can only occur within a socio-cultural-historical context. As a result, society, culture, and history all of which involve gender, race and class issues are acknowledged in the ways of knowing of the child. In viewing child-centered pedagogy as relational pedagogy via ZPD, a dialectic hybrid of human life is created as opposed to restrictive linear or binary models.

Micropolitical Perspective

Micropolitical assessment is a common tool employed when examining the power present in organizational structures. Using the postmodern approach to ZPD, the "norm" of graded education is challenged at the micro political level offering hope for the "marginalized, dispossessed and disempowered as they refuse modernist discourses that classify, control, and measure" (Greishaber & Ryan, 2006, p. 533). This study's focus on

micropolitics is important because according to Anderson (1999) it is at this micro level where change is thwarted via a co-opting of issues by the opposition, where the message of a movement is taken over by the resistance. In an effort to offset the co-opting of graded education reform, socio-cultural micropolitical policy analysis will allow for a re-framing of the issue. For example, instead of educators assuming students, parents and communities need to change to conform to the effective and equitable graded system of education, the focus can shift to the illumination of the oppressive structure of graded education and the activism of students, teachers, and leaders as agents who refuse to assimilate with the dominant cultural norms. It is the re-framing of politics at the sociocultural level that "give voice" in order to "reify and redefine power relationships" in an effort to reform unjust educational organization (Marshall & Gerstl-Pepin, 2005, p.108). Therefore the utilization of postmodernism to expose the injustice of graded education, the aid of the relational pedagogy of ZPD to impart a solution, and the sociocultural micro political analysis to re-frame education towards a nongraded structure offer realistic hope for change.

Nongraded Education

Noncognitive Rationale

As a part of this cognitive quantitative study, a review of noncognitive rationale serves as a foundation toward achieving a deeper understanding of the reasoning behind the nongraded education movement. It is important to note that studies in the field are dated, with the majority of research spanning the 1960s to 1990s. At the onset of 2000, research studies became limited and commentaries on multiage education appeared with

titles such as "Once-Popular Multiage Grouping Loses Steam" (Jacobson, 2003, p.1). With the enactment in 2001 of a standards-based educational reform titled No Child Left Behind (NCLB), nongraded education was held to rigid yearly graded standards (Pardini, 2005). According to Grant, at the onset of NCLB multiage programs began to wane because "teachers and administrators…felt compelled to "teach to the test"…it's easy to understand and sympathize with teachers whose jobs are on the line because their test scores are going to be published" (as cited in Pardini, 2005, p.27-28). Therefore, nongraded programs were reduced in an effort to meet the yearly performance standards of NCLB, as revealed by the sparse literature on nongraded education over the past 10 years. While nongraded education research prior to 2000 offers a significant contribution to the field, it also supports the need for the current study to fill the research gap left by NCLB.

Retention, promotion, and ability tracking. From the early literature, one of the best summaries of the nongraded educational rationale is found in a Virginia state policy brief:

> It's a change from conformity to diversity; from sequential, step-by-step approaches to self-paced and developmentally paced approaches; from age and ability grouping to multiage, multiability grouping. It means moving from the notion that the child should fit the school to a notion that the school should fit the child, from segregating special programs to integrating special programs, from competition to cooperation, and from failure-orientated to success-orientated schools. (Policy Briefs, 1991, p.3)

The underlying belief of nongraded education is that a child's chronological age should not be used as a system of classification to be equated with the cognitive, social, and emotional capabilities of the student. In the traditional graded classroom children are

homologous in age, establishing norms in terms of the student's ability. Students are compared to one another based on grade level expectations, creating a bell curve of low, average, and high performing students. This sets the stage in the graded classroom for issues of promotion, retention, and ability tracking. For researchers (Cohen, 1989; Gutièrrez & Slavin, 1992; Shepard & Smith, 1989) retention and ability tracking are two factors in particular which relate to failure in children. Ability tracking is seen in elementary groupings with students who are at the same level within a classroom, for example in math and/or reading. At the middle and high school levels students are segregated into separate classrooms which match their ability level in varying subjects.

Proponents of graded education argue that segregation based on age and ability is acceptable because children from nongraded and graded classrooms receive the same level of education via the employment of developmentally appropriate practices (DAP), a "consistent factor" in "positive achievement outcomes" (Kinsey, 2001, p.1). While segregation of students based on aptitude may appear appropriate, it is contrary to developmentally appropriate practice and cognitive learning theory where a heterogeneous grouping of students is essential to benefit from educational scaffolding within the zone of proximal development ("Developmentally Appropriate Practice", 2009; Vygotsky, 1978).

Ability tracking still occurs at all grade levels, with students above the norm being recommended for gifted and talented education (GATE) and/or permitted to skip a grade level, whereas students who are below the norm face retention, promotion to the next grade level for which they are not prepared, and/or referral to special education testing

(Pratt, 1986). According to Elkind (1987) young children are particularly at risk for failure in kindergarten where the youngest students commonly perform below their older classmates. Achievement data from one study suggested that children entering first grade can vary in aptitude by four years (Goodlad & Anderson, 1987). As a result, developmental kindergarten (2 years), transition classes before 1st grade, retention, and screening tests such as the Gesell have all been utilized to ensure that the child is ready for the school. However, Elkind (1987) contended, "the problem is not in the child but rather in the mismatch between the child and the curriculum" (p. 2). Therefore according to researchers (Anderson & Pavan, 1993; Franklin, 1967; Goodlad & Anderson, 1987; Song et al., 2009) graded education with its rigid set of curriculum and age related standards does not respond to the developmental needs of students and abandons the needs of both high and low achievers. Does graded education's focus on age curriculum requirements affect drop-out rates? While research is limited, an early study by Chalfant (1972) showed a significantly lower drop-out rate in nongraded secondary schools, warranting future studies in this area.

Cooperation versus competition. In nongraded classrooms, students learn in an environment without reference to grade level norms. According to Goodlad and Anderson (1987) ungraded education is a process where children are met at their current level and provided a DAP via individualized instruction, so learning and not failure can take place. Each student is challenged at their own stage of development with no set time limit, so there are no ability tracking, promotion, or retention issues, and the need for student school readiness testing and transitional programs becomes irrelevant. Also,

curriculum is integrated and matched with individual student interests. Assessment is continuous, varied, and comprehensive instead of being solely test driven (Goodland & Anderson, 1987). Student performance is not compared and as a result there is no sense of high and low ability: failure is simply not recognized (Anderson & Pavan, 1993). As one researcher states, "In multiage settings children are expected and encouraged to learn at different rates and levels" (Aina, 2001, p.223). As a result nongraded classrooms foster cooperation as opposed to competition (Elkind, 1987; Katz et al., 1990; Policy Briefs, 1991).

Cooperation is also fostered by grouping children of varying ages, a combination which occurs most commonly in human nature (Kasten, 1998; Konner, 1975; Rogoff et al., 2003). In numerous research studies (Anderson, 1993; Elkind, 1987; Katz et al., 1990; Logue, 2006; Noddings, 2005; Pratt, 1986; Song et al., 2009; Veenman, 1995) comparing same age and mixed age students, mixed age groupings children were significantly more altruistic, caring, cooperative, and less aggressive. According to Pratt's (1986) review of past research, children made friendships based on similar developmental levels, not age, and as a result more relationships developed in nongraded classrooms. This is important as socially withdrawn children are known to be at risk for future psychiatric disorders (Pratt, 1986).

Peer tutoring. Another rationale behind nongraded education is that mixed age groupings allow for Vygotsky's zone of proximal development to take place. As such students at different developmental levels are able to scaffold information for one another which alone they could not grasp. According to Cohen, Kulik, and Kulik (1982) students

achieve a higher level of achievement when peer tutoring occurs among students of different ages, as opposed to same age tutoring. Peer tutoring in nongraded classrooms is a major educational tool that occurs spontaneously between students, as well as in group learning experiences designed by the teacher (Kasten, 1998). The most common criticism of multiage groupings is made by parents of developmentally advanced learners who fear their child will become a teacher, and not a student in the classroom (Kasten, 1998; Pratt, 1986). However, according to research (Elkind, 1987; Katz et al., 1990; Schrier & Mercado, 1994) the tutors gain caretaking behaviors, leadership skills, communication skills, and intellectual skills in the process of explaining their knowledge to others. At the same time the tutee not only receives the knowledge imparted, but also social skills and benefits from cooperative interaction as modeled by the tutor.

Environment. According to Maslow's (1970) hierarchy of human needs, an individual's ability to learn is impeded until their physiological, safety, and security needs are met. In classrooms children can feel threatened by the experience of a new environment, which includes teachers and peers at the onset of each school year, such that learning cannot take place until the child feels secure in their environment (Papay, Costello, Hedl, & Speilberger, 1975). According to medical research by Quas, Murowchick, Bensadoun, and Boyce (2002), during times of school transition children experience stress which causes their hypothalamic-pituitary-adrenocortical (HPA) axis, part of their neuroendrocrine system, to release high levels of the hormone cortisol. Increased activation of the HPA axis is associated with, "poor immune

functioning…negative emotionality, social inhibition, and shyness, all of which can lead to difficulty during transition" (p. 304).

Researchers believe measures to relieve student stress in order to promote learning are fostered in the nongraded classroom. Burts et al.'s (1992) study determined students in classrooms that employed developmentally appropriate practices (DAP) such as nongraded education, were less likely to experience stress when compared to non DAP classrooms. Another stress reducer in the nongraded classroom is the presence of the same instructor for two or more years. Studies have shown that students in nongraded classrooms are more secure and comfortable in their already familiar environments when they begin a school year, as opposed to their graded peers (Fu et al., 1999; Papay, Costello, Hedl, Speilberger, 1975).

According to a study by Way (1981) multiage students scored significantly higher on "Happiness and Satisfaction Factors" than their single aged student counterparts. Way concluded that teachers were able to form stronger teacher-parent connections, understand each student's development and personality at the onset, plan appropriate individualized curriculum over summer, and experience fewer discipline issues when compared to single-age classrooms. As a result nongraded educators were able to avoid typical beginning of the year introductions and instead could focus immediately on instruction. For parents, teachers, and students the nongraded classroom provided a less stressed environment representing familiarity and comfort (Anderson & Pavan, 1993; Elkind, 1987; Logue, 2006; Miller, 1994).

Socioemotional development. Whether it is a combination of rationale or separate factors such as environment, students in nongraded education appear to be more social and emotionally developed than their graded peers. In an analysis of 30 studies completed by Pratt in 1986, none showed a negative association between a nongraded student's self-concept and attitude toward school. A later meta analysis by Lloyd (1999) found a significantly positive effect in student's attitude toward school and students were "more advanced" in "interpersonal intelligence" than their peers in age-segregated or "straight" classes" (p.190). A higher self-concept and positive attitude toward the nongraded school melds with Noddings' (2005) "ethic of care" where such characteristics are visible when a school is able to "nurture the whole child" (p.10).

Noncognitive summary. According to Lloyd's 1999 meta analysis, in both the academic and affective realm of multi-age students no negative aspects were associated with nongraded education. In fact according to Kinsey's 2001 review, studies consistently reported positive outcomes in the noncognitive realm such as, "more positive attitudes towards school, greater leadership skills, greater self-esteem, and increased prosocial and fewer aggressive behaviors, compared to peers in traditional graded classrooms" (p.1). Due to these studies it is important to reconsider nongraded education as a positive alternative to the graded system, for as Pratt (1986) suggested, "conventional structures though sanctioned by a century of familiarity, must be questioned if they stimulate rivalry, aggression, and isolation, for no apparent advantage" (p.114).

Cognitive Rationale

Definitions. According to Lloyd (1999) and Veenman (1995), before a review of nongraded research can take place it is important to understand the varying types of multiage classrooms and the terminology applied. The differentiation of the many forms of multiage classrooms can better explain the significance of each study within the context of the current study. Generally nongradedness in studies is used interchangeably with the term multi-age (Anderson & Pavan, 1993; Lloyd, 1999; Veenman, 1995). In multi-age classrooms students receive instruction from the same teachers for three or more years and are with peers who span a three to four year age range, for example, K-1-2-3, 3-4-5, and 5-6-7. In both multi-age and nongraded classrooms curriculum follows state standards and students are tested by the state via No Child Left Behind (NCLB) regulations. It is important to note that classrooms are developed with the intent of multi-age education and are not created out of economic necessity (Lloyd, 1999).

While the definition of nongraded includes multiage classrooms, the differentiation between the two is developmental in nature, where teachers in nongraded classes make curricular adjustments to individualize instruction, and content is completed at the child's developmental rate without regard for time constraints. Also, in nongraded instruction students are seen as members of a classroom not a grade (Lloyd, 1999). The major distinction is pedagogical where "multi-age grouping is often the first step towards nongradedness, nongradedness today is also considered a philosophy of education that permeates the entire school organization and program" (Veenman, 1995, p. 325).

Multi-age and nongraded are not to be equated with other classroom arrangements such as combination and multi-grade classes. In both cases, generally two or more grade levels are instructed within the same classroom, curriculum is taught separately to each grade, students do not have the same teacher for more than one year, and the classroom is created out of economic necessity not "philosophical commitment." Combination and multi-grade classrooms are usually a tool used by administrators to level out class size and control uneven grade level enrollment (Lloyd, 1999).

Quantitative research studies. Due to significantly positive and consistent results in the noncognitive realm of nongraded education, Kinsey (2001) finds "inconsistent outcomes in the academic realm surprising" (p.1). It has been noted by Song, Spradlin, and Plucker (2009) in their review of nongraded research studies that, "much of the research is quite dated, and additional research from the mid-90's provides mixed results" (p. 6). According to Veenman (1995) the inconsistent research outcomes are in part the result of inconsistent definitions utilized in the studies, while Gutierrez and Slavin (1992) believe the varying results are the lack of distinction between the types of instruction offered in each program. In order to avoid a comparison of apples and oranges the studies discussed in this paper include only those which maintain characteristics similar to the current study. Therefore, only quantitative research focused on the academic achievement of graded versus multiage and/or nongraded students, where instruction of all subjects in the nongraded classrooms includes DAP will be incorporated in this literature review.

The first study of nongraded education was completed by Crosswell in 1897. He concluded that the graded system in the cities was superior to the ungraded system in rural areas of the nation based on his observation that the ungraded students were less knowledgeable on basic facts such as addition of fractions and grammar (Crosswell, 1897). Following Crosswell, several hundred studies have been completed on multi-grade, multiage, graded, and nongraded educational systems. In the 90s two best-evidence syntheses stand out as the most highly cited and inclusive literature reviews: Gutierrez and Slavin in 1992 and Veenman in 1995.

Gutierrez and Slavin (1992) focused their review on quantitative studies dating back to 1958. They compared the achievement of nongraded students on standardized tests in relation to their graded peers. In the evaluation the authors reported, "every effort was made to obtain every study ever reported" that included evaluations of "nongraded, ungraded, multiage, or Individually Guided Education programs in grades K-6" (p. 341) that utilized standardized measurements, random or matched assignment of students, and programs which were in place for a minimum of one semester. Studies meeting the researcher's criteria were then placed into categories based on their instructional methods: nongrading of one subject, nongrading of multiple subjects (comprehensive), nongrading with individualized instruction, individually guided education, and studies lacking instructional descriptions.

According to Gutierrez and Slavin (1992) 14 of the studies were considered comprehensive nongraded programs. Therefore the results from these 14 comprehensive nongraded studies are reviewed because they align with the instructional methods defined

above and are within the scope of the current study's nongraded focus. In all 14 studies not one found, "significant differences in favor of the graded plan" while "almost all of its (nongraded plan's) positive results were significant" (p. 352). Yet in the nongrading with individualized instruction, which included one-on-one programmed instruction and student activity packages, no significant differences in academic achievement were found. The researchers noted, however, that the longer the duration of the individual instruction program, the more positive the effect on academic achievement.

Due to the mixed results found by Gutierrez and Slavin (1992) based on teaching method, their analysis of nongraded programs emphasized the need for researchers to categorize studies by instruction type. One major critique of the Gutierrez and Slavin (1992) review is its exclusion of qualitative studies (Lloyd, 1999). However, it remains unclear how academic achievement can be measured without numeric assessment.

Veenman in 1995 completed the most highly cited study in the field, a meta-analysis of both "cognitive and noncognitive effects of multigrade and multi-age classes" (p. 319). The purpose of the study was to answer the question, "What are the actual effects of multigrade or multi-age teaching on student learning?" (p. 324). Veenman utilized the review method established by Gutierrez and Slavin (1992), which entailed collecting all research on the topic and applying specific methodological criteria to each study. As a result Veenman included studies that evaluated the effects of the independent variables multi-age and multigrade groupings, on the dependent variables including cognitive and noncognitive achievement. In order to narrow the range of the study to

match the scope of the current project, only Veenman's portion involving the cognitive effects of multi-age classrooms are discussed.

Using effect size which measures the strength of a relationship between the independent (multi-age) and dependent (cognition) variables, Veenman (1995) determined that no significant differences were present in academic achievement between multi-age and single-age classes. The strength of his results were in the application of strict methodological criteria as well as his inclusion of world wide studies. However, according to a major critique of his article by Mason and Burns in 1996, Veenman's "simply no worse and simply no better" (p. 307) conclusion is fraught with uncertainty because it does not account for the formation or instruction involved in each classroom. Mason and Burns argued that when class formation is taken into account, multi-age classes have a slightly negative effect on cognition. Within months of their article Veenman (1996) responded to their claim and acknowledged the need for future studies to review classroom instruction. However, he refuted the negative association between multi-age and academic achievement by showing an effect size of essentially zero and a high interval of confidence which eliminated any negative conclusion.

Veenman's (1996) statistical defense of his results is valid and this study cannot be ignored in a review of the literature. However, Veenman's conclusion of no achievement differences between multi-age and graded classrooms are inapplicable to the current study, as he does not include nongraded classrooms in his study stating, "nongradedness was generally excluded" because it represented a "philosophy of education that permeates the entire school organization and program" (p.325).

Since the last meta-analysis by Veenman in 1995, few primary research studies have been published in journals on multi-age classrooms. Three of these studies (Burns & Mason, 2002; Mariano & Kirby, 2009; Wilkinson & Hamilton, 2003) reviewed combination/multigrade classrooms, which were created out of an economic necessity, rather than an impetus that was philosophical or pedagogical. These classrooms were not similar to nongraded or even multi-age education. Two more studies, one by Aina (2001) and the other by Logue (2006), included nongraded classrooms but employed a qualitative observational lens in the analysis. Two studies by Ong, Allison and Haladyna (2000) and Fosco, Schleser, and Andal (2004) provided the most similar context to the current study. Both studies used a quantitative assessment of reading achievement in both the nongraded and single grade classrooms.

The Ong et al. (2000) study investigated achievement of Title I students in third grade in reading, writing, and mathematics in a multiage versus single-age setting. This study is similar to the current study in many ways. Schools were selected that contained both multiage and single-age classrooms, reading achievement was measured through state standardized testing, and test data was obtained without student identification in a quasi experimental ex post facto design. The results of the study showed that multiage students performed higher than single-age students in reading, writing, and math. In terms of Title I no significant difference was apparent between the performances of Title 1 students in multiage versus single-age classrooms.

Overall the Ong et al. (2000) study was well designed and implemented. However, their measure for the state standardized assessments in reading and writing

included integrated performance which required the use of a general rubric. As a result the alpha reliability for their assessments of reading and writing were .74 and .68 respectively. The researchers acknowledged the low reliability of the measures, but anticipated with statistically significance results, the low reliability would be overridden.

In the second quantitative study cited earlier by Fosco et al. (2004) differences in cognition were examined in multiage versus traditional kindergarten through second grade classrooms. The main focus of this study was to identify the cognitive developmental level differences of each child in multiage and traditional classrooms and view "the effects of their cognitive developmental level on reading ability" (p. 4). Fosco et al. developed their study in light of a previous study by Cromey in 1999 which produced significant results showing that multiage students performed at higher cognitive developmental levels when compared to their traditional peers. The Fosco et al. study also contained similarities to this current research study. The students in multiage and traditional classrooms were matched based on age and gender, the study included a longitudinal scope (3 years), and students in each group only had experiences of one type of classroom.

The results of the Fosco et al. (2004) study indicated that children in the multiage settings achieved cognition at a faster rate and were functioning at higher cognitive levels when compared to their traditional classroom peers. Also, students at higher levels of cognition scored significantly higher on the Wide Range of Achievement Test-3 (WRAT-3) a measure of reading achievement, while at the same time no achievement difference was evident between multiage and single-age students. Fosco et al. explained the lack of

variation in reading ability by the fact the WRAT-3 only measured word recognition, which can be memorized and not "sounded out" using phonics (p. 15). Therefore, true reading ability beyond memorization could not be accurately assessed, and a new reading measurement was recommended for future studies.

Literacy

Over the years educational programs have been viewed according to their ability to produce successful students in the areas of reading and math. In the context of this paper, in order to further control for variables, nongraded education will be viewed only in relation to the cognitive domain of reading achievement. This does not dismiss the extreme importance of the social-emotional realm in education, nor deny other factors such as culture, community, parents, teachers, socioeconomic levels, race, special education, English language learners, student health concerns, class size, and school funding to name a few. According to Dr. Grover Whitehurst (2003) of the U.S. Department of Education, "Reading is absolutely fundamental…the inability to be fluent consigns children to failure in school and consigns adults to the lowest strata of jobs and life opportunities" (p. 2). Reading in particular was chosen as a focus due to its importance as a gateway skill toward student access to other subject areas (McIntyre et al., 2005) and due to No Child Left Behind (NCLB, 2001), which set goals for basic literacy by the end of the third grade through the Reading First Initiative (RFI). RFI focuses on early elementary literacy, similar to the nongraded population in this study.

According to the Department of Education, RFI is the cornerstone of academic achievement under NCLB. A goal of RFI is for all third graders to achieve reading

fluency such that during their fourth grade year they are at or above proficiency (NCLB, 2001). In accordance with Wong-Ratcliff, Powell, and Holland (2010) fourth grade is important because students not fluent by the end of fourth grade are less likely to achieve fluency over their lifetime. As indicated by NCLB statistics only 23% of fourth graders in California place in the proficient-to-advanced range in language arts ("Mapping California's Educational Progress 2008," 2008). Nationally, the U.S. Department of Education 2010 "Condition of Education" reported that for 2009 only 25% of fourth graders were considered proficient or above in reading (Aud et al., 2010). This correlated with the federal report on adult literacy which found that one in every four adults in California cannot understand a newspaper article and 23% are at the lowest rate of literacy, which represents a more extreme deficiency than any other state (Baer, Kutner, & Sabatini, 2009). This rate has not increased and a recent California Progress Report headlines "California Literacy Rate Tumbles, Symptom of State's Education Ills?" (Aiello, 2010). With only 25% of fourth graders across the nation proficient in reading, U.S. literacy statistics indicate the need for educational reform and the current study.

Theory of reading development. Before 1837 reading was taught by letter sounds and blending them together to form words. Then in 1837 Horace Mann introduced the "look and say" method which involved memorization of words on sight, originally developed as a way to teach deaf children to read (Venezky, 1987). Mann's "look and say" method was then promoted in teacher training programs and culminated in the 1930's with the introduction of the basal reader. The basal reader was designed to

develop stories out of words that students were required to memorize at each grade level to encourage reading (Venezky, 1987).

In the late 1950's Flesch published *Why Johnny Can't Read,* and stated that phonics, not memorization, was the key to student literacy. Phonics as defined by O'Conner, Fulmer, Harty, and Bell (2005) is the combination of phonemic awareness that allows students "to isolate and manipulate sounds in spoken words" and phonics "the linkage of sounds to alphabet letters and letter combinations" (p. 441). Flesch (1955) believed that "look and say" methods were contrary to the nature of reading development in children. Phonics was promoted by the research of Harvard Professor Jeanne Chall (1983) in the late 1960s and again in the early 1980s when she suggested teacher training of phonics should revert back to the same approach employed before the "look and say" method was introduced, namely phonics.

Today there are two factions in the theory of reading development, those researchers who emphasize phonics, and others who promote the whole-word approach (Xue & Meisels, 2004). The whole-word reading technique is characterized by constructing meaning from the text, where in place of decoding sounds (phonics), prediction and guessing are used so reading becomes akin to learning to speak (Morrow, 1997).

Research has indicated a balance of phonics and whole-word is optimal in the area of reading acquisition. One quasi-experimental study by Ayers (1998) focused on three experimental groups of beginning readers in kindergarten: group one received direct instruction in phonics, group two received indirect literature-based instruction, and group

three received a combination of the approaches used in groups one and two. Ayers (1998) concluded that students in the group with both phonic and whole text instruction performed highest on phonics awareness tests, which provided evidence of the effectiveness of the combined approach. This result correlated with Stuart (1999) who focused on kindergarten children in whole-word instruction versus those learning phonics. Initially the phonics group surpassed the whole-word group in reading words. A year later the phonics group continued to excel in reading and spelling words, but was behind the whole-word group in comprehension. Stuart (1999) proposed that phonics instruction in the context of whole language would produce students who could read, spell, and comprehend writing.

One possible explanation for the synchrony of phonics and whole language is that both components of reading have been found to be genetically linked. Several researchers have used longitudinal studies of twins to link reading ability in word identification, reading comprehension, rapid automatized naming, and spelling to heritable traits (Davis, Knopik, Wadsworth, & DeFries, 2001; Gayan & Olson, 2003; Knopik, Alarcon, & DeFries, 1998; Olson & Byrne, 2005; Petrill et al., 2007). However, discrepancy between two seminal authors occurs in the literature where Petrill et al. (2007) contended phonological awareness is under genetic influence, while Byrne et al. (2007) concluded the opposite. Possible differences in findings could be attributed to the fact only three twin population based studies have been conducted involving early childhood development: The International Longitudinal Study, the Twins Early Development Study, and the Western Reserve Reading Project (Petrill et al., 2007).

Limited populations and the fact that previous studies of twins have used children spanning wide age ranges are just two factors that limit the scope of the results (Petrill et al., 2007). The fact that mono and dizygotic twins guaranteed that half of the identical genes, along with the longitudinal nature of the majority of current studies, offered strength to the idea that reading can be genetically controlled. However, difficulties of accounting for multivariate environmental factors in a longitudinal study are encountered when surroundings of the participants change via divorce, new teacher, new classmates, different schools, etc.

Although definitive conclusions may not be drawn as to which factors of phonic or whole-word reading are genetic, it is clear that a combined educational approach can hold the maximum outcome for students. In fact, according to a national survey of elementary teachers in 2000, 89% of teachers were already teaching direct phonics instruction in combination with literature and language-rich activities (Baumann, Hoffman, Duffy-Hester, & Moon, 2000).

CST/ELA measure. The assessment of reading in early childhood education can take many forms ranging from informal teacher observations to battery screening tests. In an effort to focus the vast amount of research on the topic of reading assessment the "national research panels identified characteristics of quality early reading assessments to inform instruction" (Horm-Wingerd et al., 2001, p. 342) in order to standardize instruments and meet the demands of federal mandates in NCLB. The components include sound assessment instruments, specific interventional skills such as phonics, continued relevance over time, large scale testing, and cost effectiveness (Horm-Wingerd

et al., 2001). According to these factors the California Standards Tests (CST) in English Language Arts (ELA) meets the NCLB mandates as an assessment measure. Therefore, this study utilized the CST/ELA for grades second through sixth as a standardized measure to indicate student reading achievement in nongraded versus graded education.

The CST in ELA is designed to measure reading and writing using a multiple choice question format. The reading/writing assessments reflect the California state content standards for every grade level, with each question intended to test a specific standard or subset thereof. The components of the reading measure for grades two through six include word analysis, vocabulary development, reading comprehension, literary response, and literary analysis. Word analysis and vocabulary development is defined by the CST as student selection of "letter patterns and know(ing) how to translate them into spoken language by using phonics, syllabication, and word parts" (*California Department of Education*, 2009, p. 5). Test examples in word analysis and vocabulary development specifically for fourth grade incorporate the ability to distinguish words with multiple meanings, as well as questions on synonyms and antonyms. For reading comprehension it is expected students can read and comprehend grade level material with questions pertaining to following multistep instructions and making predictions about reading passages. In literary response and analysis students are asked to read varying passages and respond regarding theme, plot, setting, and characters (*California Department of Education*, 2009).

The writing component of the CST in ELA for grades two through six is composed of writing strategies and written conventions. Each grade level is tested in

these two areas based on graded content standards. Written convention is defined by the CST as students writing and speaking with a command of the English language that corresponds to their grade level. For example, fourth grade test questions in this area include grammar, punctuation, compound sentences, capitalization, and spelling. The CST distinguishes writing strategies for all grades two through six as, "Students write clear and coherent sentences and paragraphs that develop a central idea. Their writing shows they consider the audience and purpose. Students progress through the stages of the writing process (i.e., pre-writing, drafting, revising, editing successive versions)" (*California Department of Education*, 2009, p. 4). Typical fourth grade questions would require the student to develop a five paragraph essay including introduction, body, and conclusion. The only variation of the CST writing component is the addition of a writing application category in the fourth and eighth grades. All fourth graders are asked to respond to a prompt as a narrative, a summary, or a response to literature in written paragraph format, as opposed to the multiple choice format in all of the other sections. This section is then measured based on a scoring rubric assessed by trained readers, whose inter-rater reliability in scoring has been verified (*California Department of Education*, 2009).

According to the California Department of Education (CDE) (2009) the CST in ELA is controlled and found statistically reliable in its test questions, test format, administration, and scoring, and thus provides a sound measure to assess reading achievement in schools. Whereas Valencia and Pearson (1987), who critique standardized testing in reading, would argue that multiple choice formats provide limited

information regarding reading ability, the CST does offer a written component in the fourth and eighth grades by which to measure student development. Also, the multiple choice format meets the requirements of NCLB as a sound assessment, with its focus on specific content level standards, including grammar, the ability to be utilized on a large scale, and its cost effectiveness (Horm-Wingerd et al., 2001).

Scores on the CST are reported as ranging from a low of 150 to a high of 600 on each subject, with corresponding achievement levels such as Advanced, Proficient, Basic, Below Basic, and Far Below Basic (*California Department of Education*, 2009). According to the CDE (2009) state level scores are based on the entire population, eliminating sampling error and allowing for valid comparisons to be made between groups such as students, schools, and districts.

The results of the CST in ELA in combination with the California High School Exit Exam (CAHSEE) are used to determine the adequate yearly progress (AYP) of schools. AYP was developed by California lawmakers in 1999 as part of the Public Schools Accountability Act (PSAA) which sought to hold districts accountable for student performance (*California Department of Education*, 2009). With the passage of NCLB in 2001 California already had the PSAA accountability system in place, which met the requirements of NCLB.

Scores used to determine the AYP of a school and district are termed the Academic Performance Index (API) (*California Department of Education*, 2009). API is represented as a range of scores between 200 (low) and 1000 (high). The previous year's API score provides the baseline upon which schools are asked to increase by 5% each

year until a statewide target of 800 is met. Every year a minimum five points' growth is required until an API of 800 is reached. According to the CDE (2009) 48% of elementary schools, 36% of middle schools, and 1% of high schools have achieved an API of 800. Schools who do not reach their target API in the areas of English and math are considered by NCLB as not making adequate yearly progress (AYP). After two consecutive years of not reaching AYP a school can be placed in year one Program Improvement (PI), which involves reallocating federal Title 1 funds, notification to parents of school's status, and allowing student transfers to non PI schools. The consequences increase each year the school remains in PI until the sixth year, when the school can be taken over by the state. When this happens the school can either be reconstructed as a charter school, managed by a paid outside source, or taken over by the state where a majority of the staff are replaced ("No Child Left Behind" 2001). In California CST scores in ELA remain an important component in the fate of PI schools, showing efficacy for the use of CST/ELA scores in assessing student reading achievement in a construct such as nongraded education.

Summary

Historically graded education was established in the 1840's by Horace Mann and his supporters to meet the demand for labor during the industrial revolution in the United States (Gutek, 1986). Consequently, child developmental research was not taken into consideration upon conception of the graded factory-like model (Kasten, 1998; Osin & Lesgold, 1996; Rogoff et al., 2003; Stone, 2009). Today learning theories provided by Vygotsky (1978) and others, have set forth a better understanding of the nature of

childhood development. In fact studies presented in this review support the idea that graded education is not aligned with the innate learning of a child (Anderson & Pavan, 1993; Franklin, 1967; Goodlad & Anderson, 1987; Song et al., 2009). According to Vygotsky educational scaffolding seen in mixed age grouping is essential to learning, yet is not found in single-age classrooms. As a result, it is important to review nongraded education as an alternative to the graded system. Studies on multiage education offer significant evidence that students in nongraded classrooms consistently attain more noncognitive benefits when compared to their traditional classroom peers (Anderson, 1993; Elkind, 1987; Katz et al., 1990; Kinsey, 2001; Logue, 2006; Lloyd, 1999; Pratt, 1986; Song et al., 2009; Veenman, 1995).

While the affective realm of nongraded education shows continuous positive results, studies of the academic realm have produced mixed conclusions (Kinsey, 2001). Therefore, it is necessary to research cognitive domains of nongraded education in an effort to evaluate nongradedness as an option to the current educational system. Cognitive nongraded research is especially pertinent at the present time when, "a review of the literature reveals very little now being written …and virtually no research being done on the subject" (Pardini, 2005, p.27).

The current study aimed to fill the gap in nongraded education by contributing a quantitative study focused on the academic literacy achievement of students in graded versus nongraded settings. Literacy is a key component to student achievement in other subject areas (McIntyre et al., 2005). With NCLB's (2001) requirement for basic literacy by the end of the third grade, it is important to utilize a standardized measure such as the

CST scores for reading achievement in an effort to evaluate programs such as nongraded education which could best match NCLB's literacy goal. This chapter has provided a historical, theoretical, and research based context for the study, while the next chapter reviews the specific research question addressed and the methodology employed in this study.

CHAPTER 3

METHODOLOGY

Introduction

The purpose of this causal comparative study was to examine the associations between graded and nongraded classrooms based on scores from the California Standards Test (CST) for English-Language Arts (ELA). The study is longitudinal and quantitative in nature utilizing ex post facto data from a California public school district. This chapter describes the research question, hypothesis, methodology, measure, sampling method, site, program, population, procedures, and analysis techniques applied to the study.

Research Question and Hypothesis

The research question investigated in this quantitative study was:

What effect do nongraded classrooms have on student's CST scores in literacy achievement?

The hypothesis predicted:

Students in the nongraded program will outperform their traditional graded peers in literacy achievement. This hypothesis was based on the knowledge that students in multiage classrooms are better able to access Vygotsky's zone of proximal development via the range of ages and abilities of their peers (Anderson & Pavan, 1993; Franklin, 1967; Goodlad & Anderson, 1987; Song et al., 2009).

Methodology

The research design utilized for this study is associational in nature. It employs the causal-comparative criterion-group research design to explore the ex post facto nature

of the relationship between the nongraded and traditional graded educational groupings. Ex post facto data is historical information concerning variables once they have already exerted any effects on one another (Gay, Mills, & Airasian, 2009). The criterion group consisted of students in grades two through six who entered the nongraded program as kindergarten students and remained for the duration of the program through sixth grade. The comparison group consisted of students grades two through six from the same school campus as the criterion group, who entered the traditional graded program as kindergarten students and remained at the school through sixth grade. In this study the independent variable was the nongraded program and the dependent variable was scores from the California Standards Test (CST) for English-Language Arts (ELA).

Measure

Literacy achievement for students in grades two through six from graded and nongraded classrooms were compared by using the California Standards Test (CST) in English Language Arts (ELA) as a measure. The CST in ELA is a standardized criterion-referenced test (CRT) that identifies the level of California content standards a student can perform and provides a scaled and proficiency score based on pre-determined performance levels set by the state. Components of the ELA include word analysis, reading comprehension, literary analysis, Standard English conventions, and writing strategies.

The CST/ELA is considered a highly reliable measure with an average Cronbach's Alpha reliability value of .94 for the 2005 to 2009 school years included in the study (*California Department of Education*, 2009). Cronbach's Alpha measures the

internal consistency of the CST/ELA to ensure that the test results reflect the knowledge of the students and cannot just be accounted for by chance alone. A value of alpha close to one indicates a greater likelihood of students receiving consistent scores if they retook the test. Therefore at .94 the CST/ELA offers a high degree of reliability (Fraenkel & Wallen, 2006).

The CST is administered to students beginning in the second grade. Students are not compared to one another although two students with identical scores on the CST/ELA would be considered to have similar ability levels. According to Bond (1996) CRT standardized testing allows for uniform interpretation of scores across students with an easier match to statewide proficiency categories (far below basic to advanced) and the corresponding scale score ranges (150 to 600) for grades two through six on the CST/ELA. Scoring ranges for the CST/ELA are illustrated in Table 1.

Table 1

CST English-Language Arts State Ranges

Grade	Far Below Basic 1	Below Basic 2	Basic 3	Proficient 4	Advanced 5
2	150-261	262-291	300-349	350-401	402-600
3	150–258	259–299	300–349	350–401	402–600
4	150–268	269–299	300–349	350–392	393–600
5	150–270	271–299	300–349	350–394	395–600
6	150–267	268–299	300–349	350–393	394–600

Note: See http://www.cde.ca.gov/ta/tg/sr/documents/csttechrpt09.pdf

Sampling Method

For the causal-comparative design, nonrandom purposive sampling was utilized to select the population for this study. Purposive sampling was employed in order to select the nongraded population from their graded counterparts, which was needed to represent the independent variable in the study (Fraenkel & Wallen, 2006).

Site

This investigation is based on ex post facto data from one elementary school in a California Unified School District that requested anonymity as a condition of the study. The school site offers a very unique research environment in that both the traditional grade levels and nongraded (multiage) classrooms are offered on the same campus. In typical nongraded studies most students are compared across schools, districts, and states. Maintaining both study populations at an intra- versus inter-school site location allowed for the rare opportunity to control for several otherwise challenging variables including district student demographics (socioeconomic levels, suburban environment, and ethnicity) and identical school environment (administration, non-certificated staff, lunch program, play facilities, extracurricular activities offered, and length of instruction) factors. Although it is possible that some variability in these factors existed for both graded and nongraded study groups due to district permits allowing students to attend either program from outside the school boundaries, for example, information as to which students attended the school on permit was not provided by the district. Therefore, student demographic variables were compared between groups.

The school under study is located in California in a middle to upper middle class suburban school district and is designated a low poverty school on the state's School Accountability Report Card (SARC) as less than 25% of the students receive the free and reduced price statewide meal program. At the facility, 100% of teachers are No Child Left Behind (NCLB) compliant in that they are fully credentialed in the fields they teach (*California Department of Education*, 2009). The current average class size for the nongraded program is 26.9 students and for the graded classes are 27.2 students. The class size has increased by an average of one pupil each year since the 2006-07 school year as a result of statewide budgetary restraints. Classrooms in the nongraded program are composed of students representing a four year span in age for the kindergarten though third grade program and a two year age span for the fourth through fifth grade program. For the 2010-11 school year a new multiage program consisting of grades four through six commenced on site. Students in the nongraded program are categorized by grade level (K-3 or 4-5) for the purpose of statewide California Standards Testing. However, as is essential to nongraded classrooms, students identify themselves as members of a particular teacher's class and do not necessarily associate themselves with a specific grade level.

Program

Within the nongraded classrooms two types of programs exist. The first is a multiage kindergarten through third grade program which is similar to the traditional graded classes because parents are not required to volunteer in the classroom. The second type of nongraded program consists of a multiage kindergarten through third

grade class as well as a fourth through fifth grade classroom where enrollment is conditional on parental involvement. Parents are required to volunteer two hours per week or eight hours a month per child in the classroom or on classroom-related activities. In the nongraded classrooms where volunteerism is not mandatory and in the traditional graded classrooms parents/guardians can volunteer in the kindergarten through third grade classrooms. It is the parent/guardian who selects which of the three programs their child will enter: traditional graded (K-6), nongraded without volunteerism (K-3), or nongraded with volunteerism (K-5).

All of the nongraded classrooms on campus were taught in four connected open-walled rooms forming a quad. Students were allowed to move from room to room based on their activities which were determined by their individual lesson plans. The personalized daily plan for first through third grade was comprised of a form inside a daily folder which lists study choices. Each student then selected from their own personalized list the subjects they would review, set a goal in each subject, and planned the order of their activities for the day. It was required that students select a reading, writing, and math activity each day. Other options varied and could include Computer, Geography, Book Factory, Tile Math or Research Reports.

Instructors controlled which options were added or deleted from a student's list and new activities were taught before they were added to the list. The teacher circulated around the room to check plans and offered assistance toward goal completion. Students could not move on to their next activity without initials on their daily form from the teacher, aid, or peer tutor. Students in the class who became "experts" in particular areas

were designated as peer tutors by the instructor. Curriculum was based on California Standards but followed a hands-on approach to learning. Students in the nongraded program participated in but were not limited to small group mini lessons, writers' workshops, musical theater, science experiments, consumer education, class meetings, gardening (selecting, planting, measuring, and managing crops), author's tea, student leadership, and establishing a small business (popcorn factory) on campus. It is important to note that although the curriculum in both graded and nongraded settings follows statewide contents standards, the delivery of the curriculum as described above only occurred in the nongraded setting on the school's campus.

Population

The participants in this study included nongraded students enrolled consecutively in the same public school district, in a middle to upper middle class suburb in California. Information was collected from students in grades two through six for five consecutive years from 2005 to 2010. Years are based on school calendar years where 2006 CST scores represent the 2005-2006 school year. In order to maintain consistencies in the testing measure, students beyond sixth grade were not included in the study because the CST, developed as a result of No Child Left Behind (NCLB), had not yet been proctored as a statewide assessment. Purposive sampling was then employed to select the nongraded population from their graded peers.

In this causal-comparative design the criterion group consists of 179 students who were tested for the CST in English Language Arts (ELA) in grades two through six from the 2005-2006 through the 2009-2010 school year who remained in the nongraded

program (both with and without volunteerism) at the school site for their kindergarten through fifth grade years. Until the 2010-2011 school year the nongraded program was not available beyond the fifth grade year. However, sixth grade scores were included as it was assumed that the effects of the nongraded program would be seen into the student's sixth grade year. The comparison group consisted of 501 students who were tested for the CST in English Language Arts (ELA) in grades two through six from the 2005-2006 until the 2009-2010 school year who remained in the traditional graded program at the school site for their kindergarten through sixth grade years.

A comparison of the demographics provided by the district between both the experimental (nongraded) and the control (graded) groups is illustrated in Table 2.

Table 2

Comparison of Graded and Nongraded Student Populations

Variable		Nongraded students %	Graded students %	Difference of means
Gender	Female	49.7	47.2	-0.025
	Male	50.3	52.8	0.025
Ethnicity	White	65.4	40.9	-0.245
	Hispanic	9.5	30.7	0.212
	African American	3.9	3.2	-0.007
	Asian	14.5	19.0	0.045
	American Indian	0.6	0.6	0.000
	Declined to State	6.1	5.6	-0.005

Note: Percent is recorded only for students with CST/ELA results in grades 2-6 and does not represent the entire school or district.

The difference in the means suggests that the experimental and control groups vary on several demographics. For White students the high negative value for the difference in means at negative 0.245 suggests they are more likely to be included in the "nongraded" program's CST/ELA scores when compared to the graded population. The smaller negative values shows the possibility of a slight increase in females, African Americans, and students who declined to state their ethnicity in the nongraded program. The high positive value at 0.212 may show that the nongraded program was less likely to contain Hispanic students. The smaller positive values suggest there may be a slight decrease in males and Asian students in the nongraded population.

To assess whether the difference in the means was significant or the result of random variation, the chi-square significance test was utilized, as both the ethnicities and genders represented nominal data. Chi-square is a goodness of fit measure which tests whether the null hypothesis is correct (Gay et al., 2009). In this case the null hypothesis was that no true variation existed between the graded and nongraded population regarding gender and ethnicities that cannot be accounted for by chance alone. In terms of gender the difference between the groups was not statistically significant at the 0.05 level ($\chi^2 = .266$, $df=1$). For ethnicity the African American, American Indian, and Declined to State populations were too small to calculate significance. It can be reasoned for the American Indian population with a 0.000 difference in the means that the populations were similar. The White, Hispanic, and Asian students in each group represented a statistically significant difference between the groups at the .05 level ($\chi^2 = .000$, $df=1$) respectively. Thus, there was a significant difference in the number of

White, Hispanic, and Asian students in the graded versus the nongraded program, with more Asian and Hispanic representation in the graded population and White representation in the nongraded program.

Matched Control Population

Utilizing the entire graded and nongraded population in the statistical analysis could have resulted in errors as a result of the lack of control of three key variables. First, as Veenman (1996) and the California Department of Education (2009) have noted, criterion referenced testing measures such as the CST could differ over grade level content and grades two through six were included in the study. Second, both classifications of nongraded students were contained in the nongraded population: those classrooms with mandatory volunteerism and those classrooms without mandatory volunteerism. Finally, as seen in Table 2 both ethnicity and gender were significantly different between the graded and nongraded populations.

In an effort to account for the dissimilarities between the criterion (nongraded) and comparison (graded) groups a matched control protocol was utilized. Ethnicity and gender were controlled by matching students in the nongraded program to their counterparts in the graded program. Also taken into account was the discrepancy within the nongraded population which is composed of classrooms with and without mandatory volunteerism. In an effort to control the parent volunteerism variable, a new nongraded group was established containing only students in the kindergarten through third grade nongraded program whose parents were not required to volunteer. In this nongraded subgroup only second and third grade CST/ELA scores were available and matched to

graded students who also took the second and third grade CST/ELA tests. The second and third grade curriculum/test formats were the most closely related as the fourth grade year included a written component. Both second and third grade scaled and proficiency scores were separated for data analysis.

Therefore two groups were created composed of a criterion group of 55 nongraded non-volunteerism population and a comparison group of 55 traditional graded non-volunteerism population who completed the CST/ELA in their second and third grade year over the 2005 to 2010 school years. These two groups were matched student for student based on gender, ethnicity, and age + or – 9 months. Their demographics are presented in Table 3.

Table 3

Comparison of Matched Control Graded and Nongraded Student Populations

		Nongraded Students	Graded Students
Variable		%	%
Gender	Female	40.0	40.0
	Male	60.0	60.0
Ethnicity	White	67.3	67.3
	Hispanic	14.5	14.5
	African American	0.0	0.0
	Asian	12.7	12.7
	American Indian	0.0	0.0
	Declined to State	5.5	5.5

Procedures

Data collected regarding student demographics and CST scores in ELA for nongraded and graded students were obtained with permission of the school district.

Institutional Review Board (IRB) approval for the study was granted by the degree awarding institution. All data was coded post-facto from 2005 to 2010. Neither student names nor link to any student identifiers were provided by the school district. All student information was completely anonymous.

Data Analysis

The basic means and standard deviation were calculated from the scaled and proficiency CST/ELA scores from both the entire and matched controlled samples of nongraded and graded populations. To compare the results of each group an independent-samples *t* test and analysis of variance (ANOVA) was computed. An examination level of .05 alpha was chosen to test the null hypothesis. The null hypothesis for this study was that there were no differences in CST/ELA score gains between the nongraded and graded students. To further analyze the magnitude of the difference in the means, effect size was calculated using Cohen's d. All data with the exception of Cohen's d was analyzed using the Statistical Program for the Social Sciences (SPSS).

Summary

Chapter 3 has provided an overview of the research facility site and program including the chosen design, population, research question and hypothesis to better frame the context of the results. Therefore, based on the purpose of the study, to view student literacy achievement in nongraded versus graded classrooms, the methodology and data analysis employed were established to yield reliable and valid results.

CHAPTER 4

RESULTS

Introduction

The intent of this quantitative study was to research whether the organizational structure in graded versus nongraded (K-5) classrooms predicts literacy achievement as measured by the California Standards Test (CST) for English Language Arts (ELA) in the second through sixth grade years. This chapter presents information related to the research question, analytical plan, and results.

Research Question

What effect do nongraded classrooms have on students' CST scores in literacy achievement?

Analytical Plan

The analytical goal was to assess the effect of participation in the nongraded program on literacy achievement as measured by CST/ELA scores, compared to students in the graded program. Descriptive statistics, independent-samples *t* tests, analysis of variance (ANOVA), and effect size were calculated to achieve this goal. Sample size, means, and standard deviations were utilized to give a picture of the average score range in each population. The independent-samples *t* test was chosen as there was no pairing of scores between the graded and nongraded populations. The *t* test was conducted to learn whether the difference between the graded and nongraded means were statistically significant and therefore either accept or reject the null hypothesis. It has been suggested by researchers that when repeated *t* tests are run, as is the case in this study, statistical

problems can arise resulting in the misrepresentation of a Type I error, dealing with a possible false rejection of the null hypothesis (Gay et al., 2009). To check for this type of error a simple analysis of variance (ANOVA) was also run to corroborate the results from the *t*-test.

To further assess any significance found in the difference of the means, effect size was calculated to indicate the magnitude of significance in a relationship. Cohen's d was chosen as it was a standardized measure utilizing the descriptive statistics run in the initial analyses. The calculation for Cohen's d is the difference between the graded and nongraded means divided by the standard deviations of the groups. The effect size is linked to the sample size, the probability level accepted for significance, and the statistical power. The probability level at which an effect would be considered statistically significant was be set at α-level of .05. For this study Cohen's d is interpreted at d =.80 (large), d =.50 (medium), and d =.20 (small) effect where d =.80 would represent 80% chance of detecting an effect (Fraenkel & Wallen, 2006).

Results

Entire population

Table 4 displays the CST/ELA results for the entire graded and nongraded programs during the 2005-06 to 2009-10 school years.

Table 4

Comparison of Graded and Nongraded Students on CST/ELA Measure

		2005-2006 School Year								
		Descriptives			Independent Samples t-test (equal variances assumed)			ANOVA		Effect-size
Scores		n	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	220	3.62	1.08	-2.28	276	.024	5.18	.024	0.33
	Nongraded	58	3.98	1.10						
Scaled										
	Graded	220	356	52.4	-2.59	276	.010	6.70	.010	0.38
	Nongraded	58	376	54.4						

		2006-2007 School Year								
		Descriptives			Independent Samples t-test (equal variances assumed)			ANOVA		Effect-size
Scores		n	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	270	3.57	1.07	-2.73	343	.007	7.45	.007	0.36
	Nongraded	75	3.95	1.05						
Scaled										
	Graded	270	355	50.8	-3.35	343	.001	11.2	.001	0.43
	Nongraded	75	378	55.8						

		2007-2008 School Year								
		Descriptives			Independent Samples t-test			ANOVA		Effect-size
Scores		n	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	327	3.72	0.98	-2.43	426	.016	5.88	.016	0.27
	Nongraded	101	4.00	1.06	equal variance assumed					
Scaled										
	Graded	327	362	48.4	-2.82	148	.005	9.36	.002	0.33
	Nongraded	101	379	56.6	equal variance not assumed					

		2008-2009 School Year						ANOVA		Effect-size
		Descriptives			Independent Samples t-test (equal variances assumed)					
Scores		n	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	390	3.76	1.05	-2.97	525	.003	8.84	.003	0.30
	Nongraded	137	4.07	1.04						
Scaled										
	Graded	390	365	51.8	-3.54	525	.000	12.5	.000	0.34
	Nongraded	137	384	62.4						

		2009-2010 School Year						ANOVA		Effect-size
		Descriptives			Independent Samples t-test (equal variances assumed)					
Scores		n	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	223	3.72	0.98	-4.42	334	.000	19.5	.000	0.51
	Nongraded	113	4.21	0.95	equal variance assumed					
Scaled										
	Graded	223	364	49.0	-4.39	190	.000	21.9	.000	0.52
	Nongraded	113	393	59.8	equal variance not assumed					

Examining the means for CST proficiency scores in Table 4, with the exception of 2005-06 and 2006-07 school years, the nongraded students were within the proficient range while the graded students were within the basic range for English Language Arts (ELA). The scaled scores of the nongraded population remained from 17 to 29 points higher for all school years when compared to the graded population. However, the higher mean scores for the nongraded students fell well within the 48.4 to 62.4 range of the standard deviation of both groups for all school years.

An independent-samples *t* test found the means (with standard deviations in parentheses) of the nongraded scaled samples for the 2005-06, 2006-07, and 2008-09 school years 376 (54.4), 378 (55.8), and 384 (62.4), to be significantly higher at the .01

level ($t(276) = -2.59$, $t(343) = -3.35$, $t(525) = -3.54$) respectively, when compared to the means (with standard deviations in parentheses) of the graded scaled samples for the same school years 356 (52.4), 355 (50.8), and 365 (51.8). For the 2007-08 and 2009-10 school years an independent-samples *t* test, in which equal variance was not assumed based on the significance level of the Levene's test at less than .05, found that the nongraded scaled score means (with standard deviations in parentheses) were 379 (56.6) and 393 (59.8) which was significantly higher at the .01 level ($t(148) = -2.82$ and $t(190) = -4.39$) than the graded scaled score means of 362 (48.4) and 364 (49.0) respectively. Although significant at the .05 level, the lowest level of significance at .02 was seen in the 2005-06 school year comparing graded ($M = 3.62$, $SD = 1.08$) and nongraded ($M = 3.98$, $SD = 1.10$) proficiency scores, where nongraded students scored significantly higher yet the mean for both groups fell within the proficient range.

The difference in the means was further compared using a one-way ANOVA to control for Type I error. Calculated significances were identical to the independent-samples *t* test. The only exception was the 2007-08 scaled scores in which the *t*-test produced a significance level of .005 and the ANOVA yielded a significance level of .002 suggesting the ANOVA had reduced Type I error.

In the study both the independent-samples *t*-test and ANOVA suggested that the nongraded students scored significantly higher at the .05 level from the 2005 to 2010 school years on the proficiency and scaled scores when compared to the graded students on the same school campus. To interpret the magnitude of the difference between the means of the nongraded and graded populations regardless of statistical significance

Cohen's d was calculated to determine effect size. A small effect size (d = 0.27) in the .2 range was seen in the proficiency scores for the 2007-08 school year. The largest effect sizes were in the medium range of .5 as seen in the proficiency and scaled scores (*d* = 0.51, *d* = 0.52) during the 2009-10 school year. Between the small (.2) and medium (.5) effect size were scaled scores of d = 0.43 for 2006-07, d = 0.38 scaled scores for 2005-06, d = 0.36 proficiency score for 2006-07, d = 0.33 scaled score for 2007-09, d = 0.33 proficiency score for 2005-06, and d = 0.30 proficiency score for 2008-09.

These effect sizes ranging from small to medium suggested that one eighth to one third of the nongraded students have unique scores when compared to the graded students. Therefore, while the difference in the means between the nongraded and graded students is statistically significant at the .01 and .05 levels, the small effect sizes for the majority of the data indicated that the magnitude of the difference between the means of the groups was minimal. Only the medium effect size for the proficiency and scaled scores d = 0.51 and d = 0.52 respectively during the 2009-10 school year were of a high enough magnitude to be considered a notable finding.

Matched Control Population

As a result of the significant difference in the graded and nongraded populations, a second study group was created to better control for study variables including demographics, parent volunteerism, and cross grade level contamination of CST/ELA scores. These participants included 55 graded and 55 nongraded second and third grade students from 2005 – 2010 who were consecutively enrolled in the same school in classrooms not requiring parent volunteerism. The nongraded sample was compared with

controls matched on age, gender, and ethnicity to increase the validity regarding class type effect (Fosco et al., 2004). The identical measures/methods performed on the entire graded and nongraded populations were completed on the matched graded and nongraded comparison groups. The results are presented in Table 5.

Table 5
Comparison of Matched Control Graded and Nongraded Students on CST/ELA Measure from 2005-06 to 2009-10 School Years

		Second Grade								
		Descriptives			Independent Samples t-test (equal variances assumed)			ANOVA		Effect -size
Scores		N	M	SD	t	df	Sig 2-tailed	F	sig	d
Proficiency										
	Graded	47	3.13	1.12	-2.36	92	.020	5.57	.020	0.49
	Nongraded	47	3.66	1.07						
Scaled										
	Graded	47	335	51.2	-2.37	92	.020	5.63	.020	0.49
	Nongraded	47	362	58.6						

		Third Grade								
		Descriptives			Independent Samples t-test (equal variances assumed)			ANOVA		Effect -size
Scores		n	M	SD	T	df	Sig 2-tailed	F	Sig	d
Proficiency										
	Graded	42	2.79	1.00	-4.40	82	.000	19.4	.000	0.96
	Nongraded	42	3.76	1.03	equal variance assumed					
Scaled										
	Graded	42	314	42.5	-4.55	73	.000	20.7	.000	0.99
	Nongraded	42	366	61.5	equal variance not assumed					

Examining the means for CST proficiency scores in Table 5 the graded and nongraded second grade students were in the basic range, while graded students in the third grade were below basic and the nongraded were at basic for English Language Arts (ELA). The scaled scores for the nongraded population remained 27 and 52 points higher for the

second and third grade respectively when compared to the graded population. However, the higher mean scores for the nongraded students fell within the standard deviation of both groups for both grade levels.

An independent-samples *t* test found the means of the nongraded proficiency samples for second (M = 3.66, SD = 1.07) and third grade (M = 3.76, SD = 1.03) to be significantly higher at the .05 ($t(92) = -2.36$) and .01 level ($t(82) = -4.40$) respectively, when compared to the means of the graded proficiency samples for (M = 3.13, SD = 1.12) and third grade (M = 2.79, SD = 1.00). For the third grade an independent-samples *t* test in which equal variance was not assumed, based on the significance level of the Levene's test at less than .05, found that the nongraded scaled score means (M = 366, SD = 61.5) was significantly higher at the .01 level ($t(73) = -4.55$) than the graded scaled score means (M = 314, SD = 42.5). The second grade mean scale scores for the nongraded population (M = 362, SD = 58.6) were also significantly higher than the scaled scores for the graded population (M = 335, SD = 51.2) at the .05 level.

The difference in the means was further compared using a one-way ANOVA to control for Type I error. Unlike the whole group calculations in Table 2, the matched control populations had identical independent-samples *t* test and ANOVA calculated significances, suggesting no Type I error in the analyses. In the study both the independent-samples *t*-test and ANOVA both suggested the nongraded students scored significantly higher at the .05 level for second and third grade on the proficiency and scaled scores when compared to the graded students on the same school campus.

To interpret the magnitude of the difference between the means of the matched nongraded and graded populations Cohen's d was again calculated to determine effect size. A medium effect size (d = 0.49) in the .5 range was seen in the proficiency and scaled scores during the second grade. A large effect size of (d = 0.96 and d = 0.99) was calculated for the proficiency and scaled scores respectively during the third grade year. These effect sizes ranging from medium to large indicated that one third to a little more than one half of the nongraded students have unique scores when compared to the graded students. Therefore, the statistically significant difference in the means at .05 and .01 levels between the nongraded and graded students and a medium to large effect size indicated the magnitude of the difference was remarkable between the means of the groups. This suggested that control of CST/ELA comparisons across grade levels, gender and ethnicity congruence, and eliminating nongraded students in mandatory volunteerism classrooms maintained the significance between the means while increasing the effect size from medium to large.

CHAPTER 5

CONCLUSIONS AND IMPLICATIONS

Introduction

This longitudinal study utilized a causal comparative study approach to research the effect of student placement in a nongraded classroom on literacy achievement. This chapter reviews the results related to the research question, limitations of the study, correlations to past research, considerations for future research, and implications of the nongraded classroom.

Conclusions

What effect do nongraded classrooms have on student's CST scores in literacy achievement? The research question was dealt with in two ways. First, the entire populations of graded and nongraded students were compared on their CST/ELA scores. Second, both nongraded and graded second and third grade students were matched based on gender, ethnicity, and date of birth to control for demographic and parent volunteerism variations.

The entire-population results indicated that nongraded students significantly outperformed graded students for five years during 2005 to 2010, with the magnitude of the difference being most notable for the 2009-10 school year. It is interesting to note that the significance levels and effect size increased progressively from the 2005 to the 2010 school years, suggesting an increased length of time in the nongraded program contributed to higher literacy achievement compared to students in the traditional graded system. This differed from Veenman's (1995) meta-analysis of multiage groupings

which also used effect size to measure relationship strength in which he concluded no significant differences were apparent between graded and multiage populations on academic achievement.

The matched controlled population outcome demonstrated that nongraded students had a pronounced advantage over the graded students in literacy achievement. These results were in line with the Ong et al. (2000) study which found multiage students performed higher than single-age students in reading. In a similar study design, a longitudinal matched control study by Fosco, Schleser, and Andal (2004) found that multiage students achieved cognition at a faster rate, but failed to find a significant reading achievement difference between multiage and graded students. Fosco et al. explained the lack of significance in the fact that the reading instrument measured only word recognition which dealt more with memory than phonics.

This study has utilized the matched design of Fosco et al. (2004), but instead employed a more appropriate and valid measure via the CST/ELA to analyze literacy achievement. As a result the significance levels and effect size were large enough to reject the null hypothesis that the difference in CST/ELA scores between the groups could not be accounted for by chance alone. Effect size for this matched group increased from medium in the second grade to large in the third grade suggesting the benefit of length of exposure in the nongraded program. Gutierrez and Slavin (1992) noted a similar trend in nongraded students with individualized instruction programs, akin to the individualized daily plans found in the current study's nongraded program, where duration of stay in the program was associated with a more positive effect on academic

achievement. Also, in a longitudinal study on multiage programs by Pavan (1977, 1992) and Morris, Proger, and Morrell (1971) a direct relationship was found between student performance and length of duration in a multiage program.

This study represents an effort to adhere to the recommendations made by multiage researchers to offer valid content to the field of nongraded education. Both Mason and Burns (1996) and Veenman (1995), foremost researchers on multiage education, held the following concerns about future studies: creating an equivalent population as research indicated more capable students were selected into multiage classrooms, the use of education strategies found in single-grade classrooms were used in multiage settings, and the varying qualifications of teachers placed in each setting. In this investigation student selection into classes was based on parent choice, not school or teacher selection. Also, the graded and nongraded programs at the school site utilized very different instructional methods as previously described, such as self-directed daily plans and individualized instruction found in the nongraded classrooms. Finally, all of the teachers on the school campus are NCLB qualified and teachers self-selected either a graded or nongraded teaching environment.

Beyond adhering to researcher recommendations, this research offered a compelling environment and study design under which to view nongraded education. The unique environment of having both criterion and comparison groups on the same school campus and in the same district allowed for the further control of variables that could also play roles in literacy achievement such as teacher qualifications, campus climate, administration, school resources, lunch facilities, programs offered, location

demographics, length of school year, length of day, and non-certificated staff. The matched control protocol utilized for second and third grade established equal demographic populations, eliminated volunteerism as a variable, and viewed the CST/ELA measure within subject and grade as suggested by Veenman (1996). As a result the data indicated a substantial increase in the literacy achievement for students in the nongraded program. The matched control model also looks for the first time at parent volunteerism and revealed that volunteerism may not contribute to literacy achievement in the nongraded classroom.

The results of this study are representative of the rare environment in which they were conducted, as opposed to a representation of all nongraded classrooms. Student demographics contained a high level of white students, a low percentage of other ethnicities and a lack of African American students. Also the school site contained 100% NCLB qualified educators in a middle to high socioeconomic region of California. Therefore, possible inclusion of a greater variation in all demographics would have enhanced the generalizability of this study.

In all multiage studies it is difficult to determine whether significant student outcomes are the result of the classroom structure, the instructional strategies utilized in the nongraded setting, or a combination of both. However, a unique quality of this study was the rare conditions and matched control design that allowed for variables (volunteerism, campus climate, administration, school resources, lunch facilities, programs offered, length of school year, length of day, non-certificated staff) to be controlled that have yet to be simultaneously accounted for in multiage studies to date.

Elimination of so many factors allowed for a more straightforward picture of nongraded education's effect on student achievement.

Recommendations

Following are recommendations based on results of this study:

1. Future studies would benefit from eliminating comparisons between graded and nongraded classrooms and instead compare nongraded settings with different instructional methods, in an effort tease out the effects of organizational structure versus instruction.
2. Future studies should include a broader range of student demographics, while at the same time controlling for population congruence using a matched controlled research design.
3. Further review of the role of parent volunteerism and its relationship to academic achievement in the nongraded classroom should be undertaken.
4. With positive results from nongraded education in the noncognitive arena, studies would benefit from investigating the link between cognitive and socioemotional realms as a possible causal factor in academic achievement.
5. Further longitudinal studies should be conducted to learn how nongraded students performed through middle school and high school, along with their rates of entry into college when compared to their graded counterparts.
6. The term "nongraded" generally evokes the idea that students are not being assessed with a grade. Programs might benefit from establishing a new term

for the individualized and developmentally appropriate curriculum which are hallmarks of the nongraded system.

Implications

The current graded organizational structure in school needs to be revisited based on two important factors. First, the graded system is not aligned with educational theories of learning (Kasten, 1998; Osin & Lesgold, 1996; Rogoff et.al, 2003; Stone, 2009). Kasten stated the current graded system is equivalent to producing cars in the factory model. For example, the term superintendent was first used in factories and quality control checks (likened to standardized tests) were called promotion or nonpromotion. The child proceeding from grade level to grade level is similar to the progression of a car down the assembly line, where at the end those vehicles not assembled uniformly are discarded (dropping out). Children are not cars but complex individuals which Kasten (1998) argues are complex to educate, yet governments apply simple factory models which run counter to current educational research.

The results of this study suggest that nongraded education, by responding to the developmental pacing of children in the classroom, may offer a viable alternative to the graded system. By definition nongradedness is designed to provide developmentally appropriate practice (DAP), which is based on the science of child development. DAP means teachers respond to individual student needs, offering child-centered education whose strategies include constructivism, cooperative learning, and integrated curriculum, which "allows all students to participate at their own developmental level" (Williams & Strangis, 2002, p.11).

The second reason to review options to the organizational structure of education is the presence of age segregation in the graded classroom. In a recent paper on ageism, Hagestad and Uhlenberg (2006) found two types of age segregation in traditional schools; institutional and spatial. Institutional segregation is observed when children spend most of their time with a narrow age range of peers in settings exclusive of young and old. Spatial segregation is defined by historian Philippe Ariès who viewed urban life as sectioned, where each day children move to their islands (schools), adults to their islands (work), and the elderly remain on their islands (retirement communities, nursing homes, etc.) (as cited in Hagestad & Uhlenberg, 2006). According to Pratt (1986), age segregation is a "relatively recent phenomenon, and one which runs counter to the pattern of upbringing of the young which previously existed for millions of years" (p. 111). Anthropology, ethology, developmental biology and psychology all point to the innate nature and necessity of age-stratification in learning models (Konner, 1975). In fact in primate populations, tribal societies, medieval Europe, colonial society, and other countries age-stratified cultures represent the norm.

This separation by age is harmful because according to Pettigrew (1998) it establishes an "us versus them" mentality, making it difficult to understand others of varying ages, creating stereotypes and discrimination and increasing the number of "isms" in society. Therefore, no "logic or reason for grouping students in age-segregated ways" exists as grade levels were not "designed to suit the needs of children" but instead developed solely for economic reasons (Kasten, 1998, p. 2).

Due to the historical graded educational structures, individuals for the past hundred years in the United States have progressed through age-segregated schools, making the graded system appear appropriate and universal to the point that mainstream society no longer questions its legitimacy (Gramsci, 1971). Yet, graded education is not universal, even though many educators cannot envision schools without grade levels (Connell, 1987; Pratt, 1986). Multiage schooling is on the decline in the United States due to the grade level restrictions of No Child Left Behind, while in Europe, Canada, and parts of Asia it is on the rise (Pardini, 2005). In Ireland, multiage educators were able to eliminate standardized testing and instead assess students based on individual curriculum not linked to a grade level benchmark (Mulryan-Kyne, 2005). In both England and New Zealand nongraded teachers have individual academic plans for each child and students progress at their own pace, unaware of retention and promotion practices. Children proceed to new levels of learning by achievement rather than by age (Connell, 1987). In nations such as Australia, New Zealand, Netherlands, Finland, and Canada with the highest literacy rates in the world, nongraded classrooms are common and even mandated (New Zealand and Netherlands Antilles) educational practice (Aina, 2001; Song et al., 2009). According to the most current 2009 study by the global Organisation for Economic Co-operation and Development (OCED), Australia, New Zealand, Netherlands, Finland, and Canada significantly outperformed the United Stated on reading, math, and science ("PISA 2009 Results: Executive Summary," 2009)

Since there can be no magic formula or one size fits all standardized education, nongraded education provides options to the traditional school system. The goal of this

study has been to shed light on alternatives, utilizing research to determine which educational structure leads to the highest level of student success.

REFERENCES

Aiello, D. (2010, January 11). California literacy rate tumbles, symptom of State's education ills? *California Progress Report*. Retrieved from http://www.californiaprogressreport.com/site

Aina, O. E. (2001). Maximizing learning in early childhood multiage classrooms: Child, teacher, and parent perceptions. *Early Childhood Education Journal, 28*(4), 219-224.

Anderson, G. L. (1999). The politics of participatory reforms in education. *Theory into Practice*, *38*(4), 191-195.

Anderson, R. H. (1993). The return of the nongraded classroom. *Principal*, *72*(3), 9-12.

Anderson, R. H., & Pavan, B. N. (1993). *Nongradedness: helping it to happen*. Lancaster, PA: Technomic Pub.

Apple, M. W. (1980). The other side of the hidden curriculum: Correspondence theories and the labor process. *Journal of Education*, *162*(1), 47-66.

Aud, S., Hussar, W., Planty, M., Snyder, T., Bianco, K., Fox, M., ... Ferguson, K. (2010). *The condition of education.* (pp. 1-429) (United States, Department of Education, Institute for Education Sciences) (T. Nachazel & G. Hannes, Eds.). Washington, DC: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Ayers, L. R. (1998). Phonological awareness training of kindergarten children: Three treatments and their effects. In C. Weaver (Ed.), *Reconsidering a balanced approach to reading* (pp. 209-269). Urbana, IL: National Council of Teachers of English.

Baer, J. D., Kutner, M. A., & Sabatini, J. P. (2009). *Basic reading skills and the literacy of America's least literate adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) supplemental studies* (United States Department of Education, National Center for Education Statistics). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. of Education.

Baumann, J. F., Hoffman, J. V., Duffy-Hester, A. M., & Moon, J. (2000). The first R yesterday and today: U.S. elementary reading instruction practices reported by teachers and administrators. *Reading Research Quarterly*, *35*, 338-377.

Bond, Linda A. (1996). Norm- and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 5(2).

Bourdieu, P., & Passeron, J. (1990). *Reproduction in education: Society and culture* (2nd ed.). London, England: Sage Publications.

Bowles, S., & Gintis, H. (1977). *Schooling in capitalistic America: Educational reform and the contradictions of economic life*. New York: Basic Books.

Bryant, D. M., & Clifford, R. M. (1992). 150 years of kindergarten: How far have we come? *Early Childhood Research Quarterly*, *7*, 147-154.

Burns, R. B., & Mason, D. A. (2002). Class composition and student achievement in elementary schools. *American Educational Research Journal*, *39*(1), 207-233.

Burts, D. C., Hart, C. H., Charlesworth, R., Fleege, P. O., Mosley, J., & Thomasson, R. H. (1992). Observed activities and stress behaviors of children in developmentally appropriate and inappropriate kindergarten classrooms. *Early Childhood Research Quarterly*, *7*, 297-318.

Byrne, B., Samuelsson, S., Wadsworth, S., Hulslander, J., Corley, R., Defries, J. C., ... Olson, R. K. (2007). Longitudinal twin study of early literacy development: Preschool through grade 1. *Reading and Writing*, *20*, 77-102.

California Department of Education (CDE). (2009, April 20). Retrieved from http://www.cde.ca.gov

Chalfant, L. S. (1972). A three year comparative study between students in a graded and nongraded secondary school. *Dissertation Abstracts International*, *33*, 3178A.

Chall, J. S. (1983). *Learning to read: the great debate*. New York, NY: McGraw-Hill.

Cohen, D. L. (1989, December). First stirrings of a new trend: Multiage classrooms gain favor. *Education Week*, *6*(1), 13-15.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *Educational Research Journal*, *19*, 237-248.

Connell, D. (1987). The first 30 years were the fairest: Notes from the kindergarten and ungraded primary. *Young Children*. *42*, 30-39.

Cromey, A. (1999). *Impact of multi-age programming on social competency in five to seven year old children* (Unpublished doctoral dissertation). Illinois Institute of Technology, Institute of Psychology, Chicago, Illinois.

Crosswell, T. R. (1897). Courses of study in the elementary schools of the United States. *Pedagogical Seminary*, *4*(3), 294-334.

Davis, C. J., Knopik, V. S., Wadsworth, S. J., & DeFries, J. C. (2001). Genetic and environmental influences on rapid naming and reading ability: A twin study. *Annals of Dyslexia*, *51*, 241-258.

Developmentally appropriate practice in early childhood programs serving children from birth through age 8 (Publication). (2009). Retrieved from National Association for the Education of Young Children website: www.naeyc.org/dap

Eisner, E. (2005). Back to whole. *Educational Leadership*, *63*(1), 14-18.

Elkind, D. (1987). Multiage grouping. *Young Children, 43*(2), 2.

Ellis, S., Rogoff, B., & Cromer, C. (1981). Age segregation in children's social interactions. *Developmental Psychology*, *17*(4), 399-407.

Engels, F. (1941). *Ludwig Feuerbach and the outcome of classical German philosophy*. New York, NY: International.

Fact sheet: The race to the top. (2009). Retrieved from http://www.whitehouse.gov/the-press-office/fact-sheet-race-top

Flesch, R. F. (1955). *Why Johnny can't read.* New York, NY: Harper & Row.

Fosco, A., Schleser, R., & Andal, J. (2004). Multiage programming effects on cognitive developmental level and reading achievement in early elementary school children. *Reading Psychology*, *25*(1), 1-17.

Fox, C. L. (1998). *The Michigan multiage continuous progress model* (Rep. No. 141). Lansing, MI: Michigan Department of Education.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. Boston, MA: McGraw-Hill.

Franklin, M. P. (1967). Multigrading in elementary education. *Childhood Education*, 43(9), 513-515.

Fu, D., Hartle, L., Lamme, L. L., Copenhaver, J., Adams, D., Harmon, C., & Reneke, S. (1999). A comfortable start for everyone: The first week of school in three multi-age (K-2) classrooms. *Early Childhood Education Journal*, *27*(2), 73-80.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2009). *Educational research: Competencies for analysis and applications*. Upper Saddle River, NJ: Merrill/Pearson.

Gayan, J., & Olson, R. K. (2003). Genetic and environmental influences on individuals. *Journal of Experimental Child Psychology*, *84*, 97-123.

Giroux, H. A. (1983). Theories of reproduction and resistance in the new sociology of education: A critical analysis. *Harvard Educational Review*, *53*(3), 257-293.

Goodland, J. & Anderson, R. (1987). *The non-graded elementary school.* New York, NY: Teachers College Press.

Gramsci, A. (1971). *Selections from the prison notebooks of Antonio Gramsci*. New York, NY: International.

Graue, E. (2005). (De) centering the kindergarten prototype in the child-centered classroom. In S. Ryan & S. Grieshaber (Eds.), *Practical transformations and transformational practices: Globalization, postmodernism, and early childhood education* (pp. 39-58). Amsterdam: Elsevier JAI.

Green, T. (2008). Three ideologies of Individualism: Toward assimilating a theory of individualisms and their consequences. *Critical Social Policy, 34*(1), 117-137.

Grieshaber, S.J., & Ryan, S. (2006). Beyond certainties: Postmodern perspectives, research, and the education of young children. In B. Spodek, & O.N. Savacho (Eds.). *Handbook of research on the education of young children* (pp. 533-554). *Mahwah, NJ: Lawrence Erlbaum Associates.*

Gutek, G. L. (1986). *Education in the United States: An historical perspective*. Englewood Cliffs, NJ: Prentice-Hall.

Gutièrrez, R., & Slavin, R. (1992). Achievement effects of the nongraded school: A best evidence synthesis. *Review of Educational Research*, *62*(4), 333-376.

Hagestad, G. O., & Uhlenberg, P. (2006). Should we be concerned about age segregation? *Research on Aging, 28*(6), 638-653.

Hamilton, D. (1989). *Towards a theory of schooling.* London: The Flamer Press.

Holzman, L. (2006). Activating postmodernism. *Theory & Psychology*, *16*(1), 109-123.

Horm-Wingerd, D. M., Winter, P. C., & Plotchan, P. (2001). *Primary level assessment for IASA Title 1: A call for discussion* (Rep.). Washington, DC: Council of Chief State School Officers.

Jacobson, L. (2003). Once-popular multiage grouping loses steam. *Education Week*, *23*(2), 1,15.

Kasten, W. C. (1998). Why does multiage make sense? Compelling arguments for educational change. *Primary Voices K-6*, *6*(2), 2-9.

Katz, L. G., Evangelou, D., & Hartman, J. A. (1990). *The case for mixed-age grouping in early childhood education* (Rep.). Washington, DC: National Association for the Education of Young Children.

Kinsey, S. J. (2001). Multiage grouping and academic achievement. *ERIC Digest Clearinghouse on Elementary and Early Childhood Education*, *January*, 1-2.

Knopik, V. S., Alarcon, M., & DeFries, J. C. (1998). Common and specific gender influences on individual differences in reading performance: A twin study. *Personality and Individual Differences*, *25*(2), 269-277.

Konner, M. (1975). Relations among infants and juveniles in comparative perspective. In M. Lewis & L.A. Rosenblum (Eds). *Friendship and peer relations* (pp. 27-66). New York: Wiley.

Langford, R. (2010). Critiquing child-centered pedagogy to bring children and early childhood educators into the centre of a democratic pedagogy. *Contemporary Issues in Early Childhood*, *11*(1), 113-127.

Levenson, D. (1977, March). Where do they belong? *Teacher*, 54-56.

Lloyd, L. (1999). Multi-age classes and high ability students. *Review of Educational Research*, *69*(2), 187-212.

Logue, M. E. (2006). Teachers observe to learn: Differences in social behavior of toddlers and preschoolers in same-age and multiage groupings. *Young Children*, *61*(3), 70-76.

Mapping California's Educational Progress 2008. (2008, November 25). *U.S. Department of Education*.

Mariano, L. T., & Kirby, S. N. (2009, June). Achievement of students in multigrade classrooms: Evidence from the Los Angeles Unified School District. *Working Paper*, 1-17.

Marshall, C., & Gerstl-Pepin, C. I. (2005). *Re-framing educational politics for social justice*. Boston, MA: Pearson/Allyn and Bacon.

Marx, K., & Bottomore, T. B. (1964). *Karl Marx: Early writings*. New York, NY: McGraw-Hill.

Maslow, A. H. (1970). *Motivation and personality*. New York: Harper & Row.

Mason, D. A., & Burns, R. B. (1996). "Simply no worse and simply no better" may simply be wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research*, *66*(3), 307-322.

McClusky, F. D. (1920a). Introduction of grading into the public schools of New England, Part 1. *The Elementary School Journal*, *21*(1), 34-46.

McClusky, F. D. (1920b). Introduction of grading into the public schools of New England, Part II. *Elementary School Journal*, *21*(2), 132-145.

McIntyre, E., Petrosko, J., Jones, D., Powell, R., Powers, S., Bright, K., & Newsome, F. (2005). Supplemental instruction in early reading: Does it matter for struggling readers? *The Journal of Educational Research*, *99*, 99-107.

Meyer, J. W. (1977). The effects of education as an institution. *American Journal of Sociology*, *83*, 55-77.

Miller, B. A. (1994). *Children at the center implementing the multiage classroom*. Portland, OR: Northwest Regional Educational Laboratory.

Moll, L. C. (1990). *Vygotsky and education: Instructional implications and applications of sociohistorical psychology*. Cambridge: Cambridge University Press.

Morris, V., Proger, B. & Morrell, J. (1971). Pupil achievement in a nongraded primary plan after three and five years of instruction. *Educational Leadership*, *29*, 621-625.

Morrow, L. M. (1997). *Literacy development in the early years: Helping children read and write*. Needham Heights, MA: Allyn & Bacon.

Mulryan-Kyne, C. (2005). The grouping practices of teachers in small two-teacher primary schools in the Republic of Ireland. *Journal of Research in Rural Education*, *20*.

No Child Left Behind (NCLB). (2001). *U.S. Department of Education*. Retrieved from http://www2.ed.gov/nclb/landing.jhtml

Noddings, N. (2003). *Caring: A feminine approach to ethics & moral education*. Berkeley, CA: University of California Press.

Noddings, N. (2005). What does it mean to educate the whole child? *Educational Leadership*, *63*(1), 8-13.

O'Connor, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. *Journal of Learning Disabilities*, *38*(5), 440-455.

Olson, R. K., & Byrne, B. (2005). Genetic and environmental influences on reading and language ability and disability. In A. Kamhi (Author) & H. Catts (Ed.), *The connections between language and reading disabilities* (pp. 173-200). Mahwah, NJ: Laurence Erlbaum Associates.

Ong, W., Allison, J., & Haladyna, T. (2000). Student achievement of 3rd-graders in comparable single-age and multiage classrooms. *Journal of Research in Childhood Education*, *14*(2), 205-215.

Osin, L., & Lesgold, A. (1996). A proposal for the reengineering of the educational system. *Review of Educational Research*, *66*(4), 621-656.

Papay, J. P., Costello, R. J., Hedl, J. J., & Speilberger, C. D. (1975, December). Effects of trait and state anxiety on the performance of elementary school children in traditional and individualized multiage classrooms. *Journal of Educational Psychology*, *67*, 840-846.

Pardini, P. (2005). The slowdown of the multiage classroom. *School Administrator*, *62*(3), 22-3, 27-30.

Pavan, B. (1977). The nongraded elementary school: Research on academic achievement and mental health. *Texas Tech Journal of Education*, *4*, 91-107.

Pavan, B. (1992). The benefits of nongraded schools. *Educational Leadership*, *50*, 182-185.

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., Dethorne, L. S., & Vandenbergh, D. J. (2007). Longitudinal genetic analysis of early reading: The Western Reserve reading project. *Reading and Writing*, *20*, 127-146.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65-85.

PISA 2009 Results: Executive Summary. (2009). *Organisation for Economic Co-operation and Development*. Retrieved from http://www.oecd.org/

Policy Briefs. (1991). *Ungraded classrooms: Fail-safe schools?* (Vol. State Policy Program, pp. 1-3, Rep.). Charleston, WV: Appalachia Educational Laboratory.

Popkewitz, T. S. (1998). Dewey, Vygotsky, and the social administration of the individual: Constructivist pedagogy as systems of ideas in historical spaces. *American Educational Research Journal*, *35*(4), 535-570.

Pratt, D. (1986). On the merits of multiage classrooms. *Research in Rural Education*, *3*(3), 111-116.

Quas, J. A., Murowchick, E., Bensadoun, J., & Boyce, W. T. (2002). Predictors of children's cortisol activation during the transition to kindergarten. *Developmental and Behavioral Pediatrics*, *23*(5), 304-313.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chavez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology, 54*, 175-203.

Rose v. Council for Better Education, 790 S.W.2d 186 (Ky. 1989).

Schrier, D., & Mercado, B. (1994, Spring). A center moves toward multiage grouping: What have we learned? *Day Care and Early Education*, 9-12.

Schulting, A. B., Malone, P. S., & Dodge, K. A. (2005). The effect of school-based kindergarten transition policies and practices on child academic outcomes. *Developmental Psychology*, *41*(6), 860-871.

Song, R., Spradlin, T., & Plucker, J. (2009) The advantages and disadvantages of multiage classrooms in the era of NCLB accountability. *Education Policy Brief*, *7*(1), 1-8.

Shepard, L. A., & Smith, M. L. (1989). *Flunking grades: Research and policies on retention*. New York: Falmer Press.

Stone, S. (2009). Multiage in the era of NCLB [Letter written Winter, 2009 to Policy Perspective]. In Center for Evaluation & Education Policy, *The advantages and disadvantages of multiage classrooms in the era of NCLB accountability* (1st ed., Vol. 7, pp. 1-8). Bloomington, IN: Stone.

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, *69*, 587-605.

Uslin, B. (2010, April 23). Information on education code 58502 [Telephone interview].

Valencia, S. W., & Pearson, P. D. (1987). Reading assessment: Time for a change. *Reading Teacher*, *40*, 726-732.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, *65*(4), 319-381.

Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, *66*(3), 323-340.

Venezky, R. (1987). A history of the American reading textbook. *The Elementary School Journal*, *87*(3), 246-265.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Way, J. W. (1981). Achievement and self-concept in multiage classrooms. *Educational Research Quarterly*, *6*(2), 69-75.

Weil, F. (1998). Capitalism and industrialization in New England, 1815-1845. *The Journal of American History*, *84*(4), 1334-1354.

Whitehurst, G. (2003, September 10). Research: Evidence based education science and learning to read: Dr. Grover Whitehurst [Interview by D. Boulton]. Retrieved from http://www.childrenofthecode.org/interviews/whitehurst.htm

Wilkinson, I. G., & Hamilton, R. J. (2003). Learning to read in composite (multigrade) classes in New Zealand: Teachers make a difference. *Teaching and Teacher Education*, *19*, 221-235.

Williams, J. L., & Strangis, D. E. (2002). Developmentally appropriate practice making sense of multiage: Socio-emotional benefits and instructional strategies. *The Journal of Early Education and Family Review, 10*(1), 6-14.

Wong-Ratcliff, M., Powell, S., & Holland, G. (2010). Effects of the reading first program on acquisition of early literacy skills. *National Forum of Applied Educational Research Journal*, *23*(3), 1-14.

Xue, Y., & Meisels, S. (2004). Early literacy instruction and learning in kindergarten: Evidence from the early childhood longitudinal study: Kindergarten class of 1998-1999. *American Educational Research Journal*, *41*(1), 191-229.