2015

Divergent Summers: Measuring the Effect-Size of Summer Vacation on Reading and Mathematics Achievement Scores for Different Populations of Maine Students

Brian I. Mazjanis PhD
University of Southern Maine, Muskie School of Public Service

DIVERGENT SUMMERS:

MEASURING THE EFFECT-SIZE OF SUMMER VACATION ON READING AND MATHEMATICS ACHIEVEMENT SCORES FOR DIFFERENT POPULATIONS OF MAINE STUDENTS

By

Brian I. Mazjanis

B.S. University of Connecticut, 1991

M.S. University of Southern Maine, 1999

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Public Policy

The University of Southern Maine

August 2015

Advisory Committee:

David Silvernail, Professor of Educational Leadership, Advisor

Glenn Cummings, President

Catherine Ann Fallona, Professor of Educational Leadership

LIBRARY RIGHTS STATEMENT

DIVERGENT SUMMERS:

MEASURING THE EFFECT-SIZE OF SUMMER VACATION ON READING AND MATHEMATICS ACHIEVEMENT SCORES FOR DIFFERENT POPULATIONS OF MAINE STUDENTS

By

Brian I. Mazjanis

B.S. University of Connecticut, 1991

M.S. University of Southern Maine, 1999

Approved by:

____
David Silvernail, Chair

____
Glenn Cummings, Member

____
Catherine Fallona, Member

DIVERGENT SUMMERS:
MEASURING THE EFFECT SIZE OF SUMMER VACATION ON READING AND MATHEMATICS ACHIEVEMENT SCORES FOR DIFFERENT POPULATIONS OF MAINE STUDENTS

By Brian I. Mazjanis, M.S.

Dissertation Advisor: Dr. David Silvernail

An Abstract of the Dissertation Presented
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Public Policy
August 2015

This quantitative study of summer learning for Maine students in grades three through grades eight analyzed changes in academic achievement level in mathematics and reading that occurred during the summer recess of 2009.

For mathematics, it appeared that when school was not in session, students showed a cumulative loss of nearly 11 percent of a standard deviation. Although small, the change in performance over the summer was not uniform across all grades studied. For the youngest students in this study, the summer recess represented a time where children collectively lost nearly 40 percent of a standard deviation in mathematics. While gender did not show a statistically significant affect on a child's mathematics achievement over the summer, a child's socioeconomic status (SES) did. Taken cumulatively over the course of this study, high-SES children made a cumulative gain of just over one third of a performance level in mathematics as compared to their low-SES classmates.

For reading achievement, it appeared that when school was not in session, students showed a slight gain in reading of just about 2 percent of a standard deviation. Again the change was not uniform: children in the youngest grades of the study appeared to gain in achievement level during the summer, while the oldest children in this study lost nearly 32 percent of a standard deviation. Both gender and SES had a statistically significant impact on a child's summer learning. Over the five grade spans of this study, high-SES children gained nearly 25 percent of a performance level over their low-SES classmates while female students gained nearly 40 percent of an achievement level over their male classmates.

The patterns of learning exposed in this study for different categories of students during the summertime have meaningful implications for policymakers attempting to close the achievement gap. First, it suggests that efforts to close the achievement gap must include efforts to address out-of-school learning factors. Second, by including the summer learning in their calculations accountability measures that use an annual assessment to measure the effectiveness of teachers and schools at closing the achievement gap contain a substantial error.

DEDICATION

In the first class of our Ph.D. cohort, Dr. Mark Lapping discussed the figurative concept of having one's family, colleagues, and friends serve as "ladders" in one's life. Such people, he said, make emotional, intellectual, and personal contributions that one uses to climb towards enlightenment, or at least to a vantage point where one can see further than before. As I reflect on my life, as well as on my most recent educational journey, I realize how fortunate I have been to have had so many wonderful people act as ladders.

My parents' words of encouragement for educational achievement and scholarship were matched only by their actions in support of those words. I can remember my father's hard work and determination to finish his undergraduate degree while engaged in full-time employment. I can also remember my mother typing at the kitchen table on her black manual typewriter helping prepare my dad's work. The many hours she spent on this manuscript, making numerous cogent comments, and helping me in my scholarship as she did my father many years ago, felt a little like déjà vu.

My two wonderful children, Madeline and Peter, have also been ladders and motivators. Both of them were part of the 66,857 student sample in this study. They made this analysis real for me, and were one of the inspirations for my study of summer learning. One night in Paris, we were trying to determine which subway train would take us to our hotel. It was late; my wife and I were tired and confused. But ten-year-old Maddy and six-year-old Peter looked at a map and somehow figured out the way back. I'm still not sure how they did it, but I am sure that it was a summer-

learning experience that mattered intellectually. To see them now as they meet life's other challenges in their own unique ways may be the greatest joy a father can hope to have.

Finally, I am deeply indebted to my wife, Debbie, for her unwavering encouragement and acceptance. While her willingness to provide critical feedback on both content and style was significant, her most important contribution was her absolute belief in my ability to do meaningful work. She inspires me to be better than I am in everything that I do.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. David Silvernail, my dissertation chairperson, for his countless hours reading drafts, making suggestions, and engaging me in conversations about the data. His deep understanding of quantitative study, as well as his expertise with the intricacies of the Maine educational system, was invaluable. When I would stray, or try to exceed the reach of the data, Dr. Silvernail would challenge my thinking. He allowed me to make mistakes and learn from them. Perhaps most importantly, he encouraged me to wander into the data to understand their meaning, limitations, and potential, and only then to start writing.

I am also deeply indebted to Dr. Catherine Fallona for her critical feedback on my work. Several conversations we had about this study had profound impacts on my writing and thinking. Her questions and observations made the final product much better.

I would like to thank Dr. Glenn Cummings who was invaluable in helping me understand the public policy implications of this research. Dr. Cummings' political perspective shaped my thinking, and changed my opinion about the political process. I am fortunate to have had such an insightful educator on my dissertation committee.

Outside of my committee, James Sloan at the Center for Educational Policy, Applied Research and Evaluation, provided the raw data for this investigation. Jim contributed hours of time helping me understand how to work with the data and suggesting possible ways of investigation. I am also thankful for Dr. Anne Ruffner Edwards' assistance in editing this document. Her work was both impressive and humbling.

Finally, there was a cadre of professors, as well as the twenty-one members of our Ph.D. cohort, whose ideas, admonitions, and critical feedback contributed in a positive way to this dissertation, as well as to my thinking. I am most thankful to them all.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I: INTRODUCTION

Statement of Problem

The United States educational system is different from those of many other countries. Not only does it have fewer hours during the school year (Silva 2007), it also has an uncharacteristically long summer recess (Wiseman, Baker 2004). Changes in student achievement during the summer recess, and its effect on the classroom, have been documented in the educational literature over the past century. The earliest researchers such as Bruene (1928) tried not only to quantify summer loss, but also attempted to define a causal relationship between a child's intelligence and his/her rate of learning loss. Cook (1942) pondered, "In the experience of the writer, it has been noted that children with intelligence quotients below 90 usually do very little reading during the summer" (p. 215).

While summer learning-loss appears to be an inefficiency of the American educational calendar, the problem appears to have implications beyond mere wastefulness. What makes summer learning-loss both an educational problem and an ethical concern is the differential effect that time away from school has on children from differing backgrounds. The Downey, von Hippel and Broh (2004) investigation using the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) data suggested that, "…for students in a typical school, the non-school environment encourages advantaged children to pull ahead…" (p. 623). Cooper, Nye, Charlton, Lindsay, and Greathouse, in their 1996 meta-analysis of 39 studies concluded, "…middle class students appeared to gain on grade-equivalent reading recognition tests over summer, while lower class students lost on them" (p.265).

Alexander, Entwisle and Olson (2007) quantified the cumulative effects that differences in non-school time learning had on elementary-aged children from different economic backgrounds. Their study suggested that from grade one through grade five, two thirds of the difference in achievement scores on the California Achievement Test-Reading (CAT-R) between students of low socioeconomic status (SES) and high-SES could be attributed to differential learning during the summer. From their study they concluded that, “Since it is low SES youth specifically whose out-of-school learning lags behind, this summer shortfall relative to better-off children contributes to the perpetuation of family advantage and disadvantage across generations” (p.175).

While the educational literature suggests that students from different SES backgrounds learn differently during the summer recess, local and federal accountability measures consistently fail to take those differences into account in their calculations of school and student progress. When The No Child Left Behind Act of 2001 (NCLB) required schools to close the achievement gap for all sub-categories of students over time, the mandated measurement was an annually administered assessment. Because annual measures not only measure the efficacy of a school, but also include what a child learned (or forgot) during the summer recess, they include an error factor. Entwisle noted that problem in her 1997 book *Children Schools and Inequality*. She observed, “When seasonal differences in growth rates are ignored, the differences in children’s achievement in summers favoring those who are better off are hidden, as is the equality of their achievement in winters” (p.37). While an annual measure is an appropriate measure of a student’s growth, its use as a measure for determining the performance of a teacher, a school, a district or a state is likely flawed by the inclusion of the summer recess.

Nearly fifty years ago, the Coleman Report correctly used a measure of annual student growth to determine if American minority children were moving toward greater equality. When explaining the finding that achievement scores between “Negro” and “white” students did not converge over the time that children were enrolled in the free public educational system, Coleman suggested that, “The difference in achievement at grade 12 between the average Negro and the average white is, in effect, the degree of inequality of opportunity, and the reduction of that inequality is the responsibility of the school” (1966, p. 21). The Coleman Report’s flaw was that of assuming that a student’s annual growth is solely the responsibility of the educational system. It is not. A student’s annual growth is a function of both schooling and what learning happens outside of school.

After nearly fifty years of investigation supporting differential summer learning, school accountability measures continue to make the error made in the Coleman Report by failing to take into account differences in summer learning. While it does matter how much a child learns each year, schools provide value only during part of that year, and therein is a fundamental problem with the current school accountability movement in the United States. There are currently no federally approved school-accountability measures that isolate academic achievement exclusively attributable to the affects of a child’s schooling. While it is appropriate to hold schools accountable for the learning that they control while school is in session, it is misleading to include in any calculation of a school's performance "out-of-school" learning for which the school has little or no control.

The same is true of teacher-evaluation systems that use annual measures to determine the effectiveness of a particular teacher's instruction. The ESEA Flexibility Review Guidance for Window 3 from the United States Department of Education asks states filling out the ESEA waiver: "Does the SEA incorporate student growth into its performance-level definitions with sufficient weighting to ensure that performance levels will differentiate among teachers and principals who have made significantly different contributions to student growth or closing achievement gaps?" (2012, p.19). That inclusion of a student-growth measure in the calculation of a teacher's effectiveness may create a misleading system in many states. For example, to comply with the ESEA waiver, Maine enacted LD 1858, which amended Title 20-A: Education. Chapter 508, § 13704. The law states that, "… measurements of student learning and growth must be a significant factor in the determination of the rating of an educator." With its passage it is likely that LD 1858 will cause many of Maine's school administrative units (SAU) to use the Smarter Balance Assessment Consortium's (SBAC) annual assessment as the measurement of student learning and growth. But any system that uses an annual assessment will necessarily have an error factor that represents learning that occurred outside of school during the summer recess.

Thus for any annual measure of student achievement to be an approximate representation of a teacher's affect on that child's learning during the school year, differential summer learning must be controlled for, or the measure risks being inaccurate and misleading. This study attempted to measure any differential student learning that occurred for Maine students in grades three through eight when the state of Maine opted to change from the spring Maine Educational Assessment (MEA) to the fall administration of the New England Common Assessment Program (NECAP) in 2009. That change in test administration by the Maine Department of Education (MDOE) created a natural experiment that allowed for the isolation and measure of student learning for different categories of Maine students during the summer recess of 2009. The study quantified out-of-school learning for different categories of students in grades three through eight.

Purpose of the Study

The purpose of this study is to determine the magnitude of the error term for different categories of students at different grade levels that is included when an annual measure is used to determine learning that occurred during the school year for elementary and middle school students in the state of Maine. The error term in this study represents the learning (or learning loss) that occurred while school was not in session during the summer of 2009 for Maine students in grade 3 through grade 8.

Research Questions

This study was organized to answer the following research questions:

1. Was there a statistically significant change in mathematics achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually, as well as for all grade spans combined.

2. Was there a statistically significant change in reading achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually as well as for all grade spans combined.
3. Were there statistically significant differences in summer achievement changes in mathematics for students of differing SES? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?
4. Were there statistically significant differences in summer achievement changes in mathematics for students of differing gender? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?
5. Were there statistically significant differences in summer achievement changes in reading for students of differing SES? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?
6. Were there statistically significant differences in summer achievement changes in reading for students of differing gender? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?

Significance

There are two primary areas of significance addressed in this study. First, quantifying the learning differential for different sub-categories of students will add to the growing literature on summer learning-loss. It will also add detail to that body of knowledge by comparing summer learning patterns by gender and SES, as well as by grade level. Second, if the magnitude and direction of summer learning can be accurately measured for particular categories of students, then accountability measures for teachers, schools, and districts can be adjusted to remove summer learning in accountability calculations.

Apart from contributing to the body of knowledge, the significance of measuring which categories of students experience summer learning-loss has important policy implications. Since the Coleman Report in 1966, policymakers have been trying to enact legislation and procedures to close the achievement gap between different subcategories of children in the United States. That is the first stated goal of NCLB. In the first section titled “Achieving Equality through High Standards and Accountability” the act states that, “The federal government can, and must, help close the achievement gap between disadvantaged students and their peers” (p.7). While that goal has a long history in educational policy, NCLB’s method for measuring progress toward it is flawed. It prescribes that: “Annual reading and math assessments will provide parents with the information they need to know how well their child is doing in school, and how well the school is educating their child” (p.8). The statement “…how well their child is doing in school…” implies that learning occurs only in school. The evidence suggests that is clearly not the case. Second and more misleading is the claim that an annual measure will show parents, "…how well the school is educating their child." That claim fails to recognize the research (Heyns, 1978, 1987; Cooper et al., 1996; Downey et al., 2004, Alexander, Entwisle and Olson 2007) that indicates that a significant aspect of the achievement differences between different categories of children is due to out-of-school influences.

If the annual growth measure of NCLB is flawed because it contains an error factor that misrepresents a school's contribution to a student's learning, policymakers may need to rethink the timing of school accountability measures, or at least account for differential student learning loss in their accountability measures. Policymakers may also need to refocus their efforts in closing the achievement gap by including educational programming for students when schools are not in session.

In the wake of NCLB, and in the era of the NCLB waivers and school accountability measures, the state of Maine has committed to use annual assessments to measure student progress. Those schools and SAUs not making defined progress face sanctions; those exceeding growth expectations, or having a high rate of children meeting or exceeding the standards, receive special recognition. While annual measures are perfectly suited for measuring student progress, in their raw form they are likely to be inappropriate for measuring a teacher's, a school's or an SAU's progress. By quantifying the error factor for Maine students during the summer of 2009, this study will provide information for policymakers to consider adjustments to annual growth measures to better isolate a school's contribution to a student's annual growth.

Limitations of the Study

The change in the state of Maine from MEA to NECAP occurred in 2009. Since then, there have been changes to educational structures, as well as an emerging dialogue with respect to summer learning-loss. Yet there have not been any major overhauls to summer programming. State funding has declined as a percentage of the total educational spending, and no additional funds have been targeted towards increased summer learning programs. In fact, during the years after the recession of 2008, the Libra Foundation gradually ended a $30-million, decade-long program that provided summer activities to children in the three largest cities in Maine. Maine's reduction of summer services was mirrored throughout the U.S. according to McCombs, Augistine and Schwartz (2011) who *in Making Summer Count How Summer Programs can Boost Children's Learning* noted that "a large number of school districts have been forced to make cuts to summer funds in the midst of a recession, placing some of the largest summer learning programs at risk" (p.54).

Since 2009, there does not appear to have been a significant change in student achievement in Maine. Table 1.1 shows that the percentage of Maine students rated proficient or proficient with distinction on the NECAP annual assessment has changed little since the 2009 administration. Differences from the average in both mathematics and reading proficiency have not shown a significant trend in any direction. Therefore, while the data used in this study are five years old, it is reasonable to assume that the educational conditions that generated the data are still in place for students in the state of Maine.

Table 1.1 - NECAP Performance Maine							
Year	% Proficient Reading	% Proficient with Distinction Reading	Difference from Mean		% Proficient Math	% Proficient with Distinction Math	Difference from Mean
2009-10	56.3	13.6	-0.66		44.8	16.7	0.02
2010-11	55.5	14.8	-0.26		43.7	16.8	-0.98
2011-12	54.6	17.6	1.64		44.4	18.7	1.62
2012-13	56	15.1	0.54		44.3	17.8	0.62
2013-14	53.5	15.8	-1.26		43.1	17.1	-1.28

Another limitation of this study is the comparability of the two assessments. Although both the MEA spring 2009 administration and the NECAP fall 2009 administration were designed to measure the same learning, that of the 2009 school year, they are nonetheless different assessments. Inherent in using different assessments to measure the same learning is the accuracy of those assessments.

The final relevant limitation to the study is the time between the administrations of the two assessments. The MEA was administered during the last part of March 2009 and the NECAP was administered during the first part of October 2009. In-school learning that would occur between the administration of the MEA assessment in the spring and the NECAP assessment in the fall would presumably account for a good deal of student growth. There were between sixty and seventy school days between the MEA administration in the spring and the NECAP administration in the fall: slightly more than one third of the total number of school days in the average Maine academic year. Nonetheless, all students would have been exposed to approximately the same number of learning days between the two test administrations.

Foundational to this study is the premise that during the time that school is in session children learn at similar rates. That premise is supported by a number of researchers including Heyns (1978, 1987), Alexander, Entwisle and Olson (2001, 2007), and Ready (2010). For example, Entwisle et al. note, “Between the fall and spring of that first year, poor children in the Baltimore sample gained fifty-seven points in reading and forty-nine points in math, and their more affluent counterparts gained almost exactly the same number of points—sixty-one points in reading and forty-five points in math” (2001, p.10). Furthermore, studies indicate that summer loss is a cumulative disadvantage over time. Entwisle et al. in the same study noted, “In the course of the first five summers in elementary school, the low-SES students gained less than one point total in reading, and they lost eight points in math. At the same time, the higher SES children gained forty-seven points in reading and twenty-five points in math” (2001, p.10). Therefore if learning while school is in session is equivalent for all children, then differences between children or groups of children are attributable to differences that occur when school is not in session, and a simplified equation can be generated as follows:

$$\text{Score}_{(\text{NECAP})} = C * \text{Score}_{(\text{MEA})} + \text{Summer Learning}$$

In that equation the fall NECAP score for each child will equal a child’s spring MEA score multiplied by a constant that allows the scores to be compared, plus a child’s summer learning. Because a child may lose or gain academic ground, summer learning may be a negative or positive value. The equation is the foundation of the theory behind this study’s focus to determine if the significant time away from school during the summer recess of 2009 impacted students uniformly, or if there were statistically significant differences for student learning over the summer recess related to a child's grade, SES, or gender.

Delimitations of the Study

In their 2007 paper Alexander et al. argued that, "The remainder of the difference is built up over the school years, and Table 1 shows that the largest component, 48.5 points, or about two-thirds of the total, traces to summer learning differences over the elementary years" (p.171). This study is by design a snapshot of one testing situation that occurred in Maine during the switch from the MEA to the NECAP in 2009. That is both a limitation of the data set, as well as a delimitation of the study. While one can hypothesize that summer-learning patterns in Maine would be cumulative, as they have been shown to be by Alexander et al. using the BSS data, further investigation into that phenomenon would be an excellent topic for a future study using a different data set.

The study deliberately chose to explore the impact of summer on reading and math, excluding the writing section of the assessment. That decision was made in an effort to keep the study focused on the aspects of the current accountability measures by NCLB and the NCLB waiver. There is currently a dearth of information on the impact that time away from school has on writing or science achievement. Those topics may be an excellent area for further investigation.

Finally, while NCLB requires that "… results must also be reported to the public disaggregated by race, gender, English language proficiency, disability, and socio-economic status" (2001, p.8), this study focused exclusively on gender and socioeconomic status.

CHAPTER II: A REVIEW OF THE LITERATURE

The educational literature is rich with researchers attempting to address the differential aspects of summer recess on student learning (Bruene, 1928; Cook, 1942; Stanovich, 1986; Heyns, 1978, 1987; Cooper et al., 1996; Downey et al., 2004; Alexander et al., 1997, 2001, 2007; Entwisle et al., 1997, 2001; Vales et al., 2013). The earliest researchers tried to quantify the qualitative effects that teachers noted after the summer recess. The question that many of those researchers attempted to answer was not so much whether students regress in their learning during the summer, but rather what was the size of the effect for different types of learners in different content areas, and for different age groups.

For example Bruene (1928) questioned the differential effect that the summer recess might have had on elementary-aged children based on their intelligence. Her research question, “Does the vacation affect differently the children on different levels of intelligence?” (p. 309) attempted to answer if “intelligence” is the key variable in student regression over the summer vacation. To quantify student progress over the summer, Bruene used the Stanford Achievement Test to measure student achievement in the spring and then again in the fall.

Bruene’s findings were that student regression was indeed dependent on a child’s intelligence as measured by the Stanford Achievement Test. She noted that while children of “high” intelligence made slight gains in reading achievement during the summer, children of “low” intelligence were much more likely to lose reading skills over the same period. For mathematics achievement, all students lost some skills, but those of lower intelligence saw greater regression than those of higher intelligence.

While Bruene did not have access to rich sources of the data about socioeconomics, family condition, education levels of parents, and other factors that would be studied later in the century and into the next, there was nonetheless no attempt to address exogenous variables in her work. Based on her writing, it appears that her research reflects the thinking of intelligence of her era: that a child's intelligence quotient (IQ) is a fixed trait. While she concludes that summer school may be beneficial for those students "below norm or just above the border line [in intelligence]" (1928, p. 314), her foundational assumption is that the difference between students performance is based on some "innate" quality of the child rather than the conditions that a child faces during the summer when he or she is away from school.

Building on the body of summer-regression research, Cook (1942) sought to determine if the activities that children did during their summer vacation affected their summer learning. As Cook wrote, "In the experience of the writer, it has been noted that children with intelligence quotients below 90 usually do very little reading during the summer." She continued, "Children of high intelligence or those who have found reading a pleasure usually read copiously when material for reading is available" (p. 215). While that again treats reading ability and intelligence as a fixed trait, her study attempts to determine if a student's experiences away from school influence his or her performance in school.

Cook's study involved summer reading and mathematics packets given to children in first and second grades in an elementary school in Minnesota. She wanted to determine if the amount of student time on summer work packets moderated summer-learning regression. In her analysis of the data Cook noted, "Children with intelligence quotients above 100 were more faithful to their [summer] work than those with intelligence quotients below 100" (p. 218). Her conclusion was that the data supported her hypothesis that, "Children with intelligence quotients above 100 were more faithful to their work than were those with intelligence quotients below 100" (1942, p.218).

Cook's findings suggest that summer regression--or a lack of regression--can be predicted by looking at the way a student spends his or her time in the summer. While her reasoning may reflect her perspective, her work does point to an idea brought forth by Stanovich (1980). He describes a situation akin to a positive feedback loop in biology in which two related factors reinforce each other. He called it "the Matthew Effect." The term originally coined in sociology by Merton (1968) refers to the passage in the New Testament: "For to everyone who has, more shall be given, and he will have an abundance; but from the one who does not have, even what he does have shall be taken away (Matthew 25:29, *New American Standard Bible*). For Cook, children who are good readers tend to read more, and therefore become even better readers, while children who struggle with reading tend to read less, and therefore do not progress in their reading development. Over time those divergent paths lead to profoundly different academic outcomes.

While Cook's findings point again to the connection between IQ and the amount of reading that students did during the summer recess, it failed to question whether those students with the higher achievement scores had those scores due to some inherent ability, or if they were due to conditions outside (and possibly inside) of school. While Cook assumed that the children read more *because* they had a higher IQ, she could have just as easily concluded that the children had a higher IQ *because* they read more. In fact much of the current research in the field of summer regression makes the assumption that differences in summer learning are due to cultural and environmental factors such as access to literature at home, parents' attitudes towards reading, amount of human interaction, and an environment conducive for reading.

For example, Gershenson (2013) found that, "The largest summer time-use gap is found in children's television viewing, as the analysis of time diaries from the Activity Pattern Survey of California Children shows that children in low-income households watched nearly 2 more hours per day during the summer vacation than their peers in wealthier households." (p. 1240). Other researchers hypothesize that differences in out-of-school learning is based on language in the home (Hart Risley, 1995), financial resources (Ladd, 2012), and parental involvement (Ramey and Ramey, 2008). While the specific differences that occur during the summer that lead to differential learning are not well understood, a theory for that pattern of differential summer learning was put forth by Entwisle in 1997.

In Entwisle's 1997 foundational book *Children Schools and Inequality,* from which she along with co-researchers Alexander and Olson wrote follow-up investigations both individually and together over the next decade, she hypothesized that resources necessary for children to learn are like water pouring out of a faucet. That is, “when school is in session, the faucet is turned on for all children, the resources children need for learning are available to everyone, so all children gain. When school is not in session, children whose families are poor stop gaining because for them the faucet is turned off” (p.37). While that pattern of resource access termed the “faucet theory” does not delve into the “black box” of what resources are disproportionately missing in SES disadvantaged homes and neighborhoods, it nonetheless simplifies the investigation by suggesting the general underpinnings of differential summer learning.

The impact of factors outside of school that influence student achievement begins even before a child enters school. When children enter school at age four or five they have had the equivalent of a four or five year vacation from school, during which the school resource “faucet” has been largely turned off. During that time there are large differences in experiences that lead to large differences in student achievement. In their 2007 book *Annual Growth for All Students, Catch-Up Growth for Those Who are Behind*, Fielding, Kerr, and Rosier assert, “On the first day of kindergarten, the range between students in the bottom and top quartile midpoints is six years in reading skills and four years in math” (p. 226). Alexander et al., echo this claim in their 2007 paper. They report that, “About a third of that SES difference, 26.5 points, traces to disparities in place when these children started 1st grade, implicating experiences and family resources that predate school entry” (p.171). In the ECLS-K study, Ready (2010) came to a similar conclusion. In the study he found that children from high-SES backgrounds start kindergarten with a sizable advantage over their low-SES classmates. That initial difference at the start of kindergarten creates a “head-start” that high-SES students add to during subsequent recesses from school.

Certainly the years before a child enters formal schooling isolate the out-of-school influence on a child’s learning, and therefore act as a control for the effects of schooling on that child. There is a growing body of evidence that suggests that there are a variety of important factors that influence a child’s academic performance and life trajectory. One of those is the effect that a parent can have as their child’s first teacher.

In their paper, “The Rug Rat Race,” Ramey and Ramey (2008) document the recent and dramatic increase in the time spent in childcare activities by college-educated parents as compared to their non-college educated peers in the United States. Using data from a variety of time diary surveys including the American Heritage Time Use Study (AHTUS) from 1965 until 2007, the researchers showed a dramatic increase in the time that college-educated parents were involved in childcare activities. While their study confirmed other researchers’ work regarding the increased involvement of college-educated parents in childcare activities, Ramey and Ramey’s research goes further by highlighting the increase in the differential between high and low-SES families starting in the late 1990s.

They found that between the years 1965 and 1995 the college-educated mothers spent between 0.06 and 2.1 hours more time invested in childcare activities then their non-college-educated neighbors with no lasting trends in the data. While both groups have experienced a steady increase in parental involvement in childcare beginning in 1998, the gap between college-educated mothers and their non-college-educated counterparts began to widen. By 2007 the increase in childcare by non-college-educated mothers was about four hours per week, while the increase by college-educated mothers was more than eight hours. From 1998 to 2007 college-educated mothers increase in investment in childcare activities was twice as great as for non-college-educated mothers. During the same period of time a survey of activities described as “free time” for both college-educated and non-college-educated mothers declined indicating that both groups were making a choice in terms of time spent on child care.

Ramey and Ramey hypothesize that the increase in parental childcare activities is a direct result of parents’ efforts to help their children gain access to elite and prestigious colleges. They theorize that the increased time spent with a child represents an attempt to give that child a slight advantage every step of the way to college: from selective preschools to the Ivy League. Regardless of the reason, the increased differential of parental involvement will likely lead to differential out-of-school learning opportunities which are likely at the foundation of the differential learning noted in the above studies. Those results are concerning, as they suggest that parental resource “faucet” differential between high and low-SES families has been growing even larger.

Starting in the 1960s, and seeming to mirror the civil rights movement in the United States, much of the literature for the next half century addressing issues of student achievement differences examined the social conditions or SES differences leading to differences in learning, and measuring the effect size of those differences.

In 1966, in accordance with the Civil Rights Act of 1964, James Coleman et al. published "The Equality of Educational Opportunity," often referred to as the Coleman Report. It set the stage for a great deal of the research on student achievement and SES that followed it. The goal of the report was to, "assess the 'inequalities of educational opportunities' among racial and other groups in the United States" (1966 p.12). In the report, Coleman concluded that, "It appears the variations in the facilities and curriculums of schools account for relatively little variation in pupil achievement insofar as this is measured by standardized tests" (p.22). He arrived at this conclusion based on the fact that schooling did not seem to mitigate academic achievement differences between students of different races in different areas of the country.

In the report and in subsequent publications, Coleman argued that the divergent educational attainment between students of different backgrounds that remained the same or widened during the twelve years of schooling indicated that schools had a relatively weak influence on students. Coleman wrote: "If the school's influences are not only alike for the two groups, but very strong, relative to the divergent influences, then the two groups will move together. If they are weak, then they will move apart." He continued that the power of schools to create opportunity "…is determined, then, not merely by the <u>equality</u> of educational inputs, but by the <u>intensity</u> of the school's influences, relative to the external divergent influences" (p.20).

While the Coleman Report was mainly concerned with racial equality of opportunity in education, much of the later educational research considers race a proxy for the underlying effect of a child's SES. Nearly a half a century after the Coleman Report, researchers continue to attempt to understand the SES connection to academic achievement. Reardon's 2011 analysis indicated that since 1970, "…family income has become more predictive of children's academic achievement" (p. 111).

Ladd (2012) argued that current public policy does not adequately address the relationship between poverty and academic achievement. She makes that argument by demonstrating the correlation between eighth-grade National Assessment of Educational Progress (NAEP) scores in reading and math to the child poverty rate in those states. Using a bivariate regression to compare state test scores and state poverty-rates Ladd (2012) demonstrates that, "a full 40 percent of the variation in reading scores and 46 percent of the variation in math scores is associated with variation across states in child poverty rates" (p. 4). Ladd goes on to show that the connection between SES and academic achievement holds true internationally. By comparing the data from the Programme for International Student Assessment (PISA), an international assessment that measures student achievement between countries, with the Organization for Economic Cooperation and Development's (OECD) measure of the economic, social cultural and status (ESCS), Ladd illustrates the international connection between student achievement and a child's economic conditions. She notes that average test scores for the children in the 5th percentile of the ESCS had an average PISA score of 350: significantly below the 660 average of students in the 95th percentile of the ESCS.

Blau (1999) also investigated the relationship between a family's income and the cognitive, social, and behavioral development of its young children using the data set from the National Longitudinal Survey of Youth (NLSY79) administered by the U.S. Department of Labor's Bureau of Labor Statistics since1979. The sampling includes assessments of cognitive, social, emotional and physical development. Starting with 1986 data, Blau used fixed effects estimation to explore if permanent income or temporary income produced a measureable difference in the exhibited cognitive and behavioral traits of the children in the study.

Blau's analysis suggested that temporary changes in income levels have a negligible effect on student achievement and behaviors, but that permanent income has a much larger effect on both of those areas. He found that the Behavioral Problems Index (BPI) showed the largest effect to changes in permanent income. Nonetheless, according to Blau, the effect size was too small to be instructive for policy decisions. Blau argued that, "The empirical results from analysis of NLSY79 data show that permanent family income has effects on child development that are too small to make income transfers a feasible approach to achieving substantial improvements in developmental outcomes of low-income children" (p. 273). Nonetheless, Blau's findings suggest that some educational characteristics seem to be malleable, albeit only moderately, to changes in household-income levels.

Dahl and Lochner (2012), using an instrumental variable (IV) strategy, also investigated the effects of income on children's math and reading achievement. They based their study on the policy changes that affected the Earned Income Tax Credit (EITC) schedule from 1988 until 2000. By using changes in the EITC, their methodology controlled for other external influences on income. Like the Blau study, Dahl and Lochner took their data from the NLSY79 sample, and focused their attention on student outcomes as measured by the Peabody Individual Achievement Tests (PIAT).

With their 4,412-children sample matched to their mothers, Dahl and Lochner conclude that during their study period in the years 1988 through 2000, an increase of $1,000 (in year 2000 dollars) in family income improved math/reading achievement by six percent of a standard deviation. While that effect is not dramatic, it is significant. When they further pared down their data, they found that the effects of an increase in income on math/reading achievement are more pronounced for disadvantaged families, younger children, and boys.

What makes those studies germane in the context of school accountability is that they demonstrate that student achievement as measured by annual standardized assessments is effected by functions likely to be exogenous to the quality of a child's school or teacher. While that is not to imply that schools do not have an effect on student learning as Coleman concluded in 1966, it does indicate that children from different backgrounds are likely to have different academic trajectories for reasons having nothing to do with the schools they attend. Those studies also support the idea that families with additional resources are able to make up for periods when schools are not in session.

In 1978, Heyns' book *Summer Learning and the Effects of Schooling* and subsequent 1987 paper "Schooling and Cognitive Development: Is There a Season for Learning?" created a framework from which much of the modern research on summer learning over the last thirty-five years has been based. The importance of Heyns' work stems from both her investigative approach and her conclusions. Citing heavily the Coleman Report (1966), Heyns attempted to determine the effect that out-of-school influences had on individual student achievement.

Heyns logically reasoned that a student's cognitive growth is a function of both in-school and out-of-school factors. She wrote, "The central premise of this study is that achievement is a continuous process, whereas schooling is intermittent" (1978, p.43). She continued, "As a quasi-experimental control for the effects of education (schooling), the summer months represent a plausible interval in which to contrast patterns of learning" (1978, p.43). She contended that the summer recess is, "a temporal control for the effects of all factors linked to cognitive growth that operate year-round such as family background" (Heyns 1987, p.1156). In effect, by measuring student growth when school was not in session the Heyns study found an elegant way to measure Coleman's "external divergent influences" (1966, p.20).

To measure academic achievement Heyns used the Metropolitan Achievement Test (MAT) as her academic measure. Her sample was 1,499 sixth graders and 1,460 seventh graders who were enrolled in the Atlanta public schools from spring of 1971 until the fall of 1972. That gave Heyns two measures of summer learning for both sixth and seventh graders: spring 1971 to fall 1971 and spring 1972 to fall 1972. By comparing student spring scores to their subsequent fall scores Heyns measured student achievement during the summer months--a time when school factors are controlled--and thus arrived at a measure of non-school achievement. She also measured student achievement during the school year, when both non-school and school factors presumably effected student growth, by measuring achievement changes from fall to spring.

Not surprisingly her data indicated that for both low-SES children and their high-SES counterparts the most productive learning occurred while school was in session. Heyns comments: “The data clearly support the contention that schooling makes a substantial contribution to cognitive growth” (p.187). Her finding is important in the context of Coleman who argued that a school’s impact was “weak.”

While all students learned more during the school year than during the summer break, Heyns also found that during the school year, the relative growth of students was similar regardless of SES. That finding is important; it indicates that schools appear to have a consistent effect on students regardless of SES. That too is a substantial divergence from the Coleman Report, and became fodder for a great deal of subsequent educational research. In a follow-up paper, Heyns continued her argument for the equalizing effect of schooling by stating, “Education provides heightened opportunities for cognitive growth to all children, irrespective of parental background” (1978 p.93). Heyns reasoned that any achievement gaps that exist between low-SES children and their high-SES counterparts do not appear to be a function of the effects of formal schooling, but rather of non-school factors.

Compounding the severity of the differential summer growth between low-SES and high-SES students is the fact that summer cognitive regression is underestimated in most studies (Downey et al. 2004), including the Heyns study (1978 p.187). That is largely due to the fact that none of the research regarding summer regression takes into account a true spring-to-fall measure. Assessments are not administered on the last day of school in the spring, and then again on the first day of school in the fall. There is, therefore, build-in error in the data. That error represents the amount of learning that occurs after an assessment is given in the spring, and before an assessment is administered in the fall, and it presumably dampens the effect size of summer learning.

Beginning in 1992, Entwisle and Alexander, and later Entwisle, Alexander and Olson, building on the work of Heyns, wrote several papers that contend that differences in achievement between high and low-SES students can be largely attributed to differences during the summer vacation. In their 1992 study, Entwisle and Alexander argued that: “The seasonal pattern of scores emphasizes the point that home disadvantages are compensated for in the winter because, when school is in session, poor children and better-off children perform at almost the same level.” They continue: “It is mainly when school is not in session that consistent losses occur for poorer children” (1992, p. 82). In his book *Outliers,* Malcom Gladwell takes that argument further stating that: “For its poorest students, America doesn’t have a school problem. It has a summer-vacation problem…” (2008, p. 260**).** For students who come from low-SES families, time away from school appears to be the great cognitive divider. While there are many theories regarding the mechanism for which low-SES children lose cognitive ground to their high-SES counterparts, including language in the home (Hart and Risley, 1995), financial resources (Ladd, 2012), parental involvement (Ramey and Ramey, 2008), there is not consensus in the literature regarding which factors are most important.

While Heyns attempted to determine what factors led to the differential learning during the summertime with an extensive series of parental surveys, her investigation yielded an unclear picture. Heyns expressed frustration with her inability to determine the factors that led to differential summer learning with her findings about bike ownership. In her study she found that having a bicycle to visit one's friends and family had, "consistent significant effects on summer achievement when background was controlled." She then continued, "Interpreting such effects literally obviously is illegitimate…" (1978, p.194). Regardless of the reason, time away from school appears to give high-SES children a cognitive advantage over their economically less-fortunate classmates. That learning advantage over time for many students can be the difference between being college- and career-ready, and becoming a high-school dropout.

Building on their earlier work Alexander, Entwisle and Olson (2007) quantified the cumulative effects that differences in non-school time have on children. They again used data from the Beginning School Study (BSS) that began in 1982, and tracked Baltimore elementary school children's progress through their schooling using the reading sub-test of the California Achievement Test (CAT-R) during 11 different testing periods. For the BSS cohort, student progress tracking began in the fall of first grade and continued to grade five. In their study Alexander et al., reviewed the data from 787 students: 397 children categorized as low-SES, 204 children classified as middle-SES and 186 children classified as high-SES.

The results from the study were remarkable in that they quantified differences in learning over time. Their data showed that from grade one through grade five, students of low-SES improved over the five winters an average of 191.30 points on the CAT-R. Their middle SES cohort improved about 19 points more than their low-SES classmates (210.19). The high-SES cohort improved only 186.11 points or 5.19

points *less* than the low-SES cohort. Thus looking at the affect that school had for reading achievement on those children, while middle-SES students did much better than both groups, children from low-SES families did somewhat *better* than those students from high-SES families during the first few years of elementary school.

While the low-SES student cohort did slightly better than the high-SES student cohort during the first five years of schooling *during the school year*, they had significantly less growth during the time that school was not in session. During the summer recess, students from the high-SES group gained 46.58 points in reading as measured by the CAT-R. That gain occurred over four summers when school was not in session, and represents growth greater than the one-year average growth for any group during the study. In contrast children in the low-SES group had a cumulative summer regression in reading of 1.90 points. That difference of 48.48 points on the CAT-R is substantial, and represents about two thirds of the difference between CAT-R scores for high and low-SES groups. Differences between the two groups during the school instructional time were nearly non-existent. According to the BSS data, the majority (two thirds) of achievement differences between high and low-SES groups at the end of fifth grade were attributed to differences in summer learning.

Another set of data that provides rich information about children's entry into public education was the Early Childhood Longitudinal Study-Kindergarten (ECLS-K). The data for this study came from children around the country who were educated in public and private schools, attended full- and half-day kindergarten, and were from diverse cultural, ethnic, and socioeconomic backgrounds. Taken from a sample of more than 13,000 children across the United States, the advantage of those data over the BSS data is that they represent a greater geographic and socioeconomic cross-section for study as well as a larger sample size.

Information from the ECLS-K data set has been analyzed by several researchers to measure the effect of summer on student learning during the early elementary grades. Ready (2010), used the ECLS-K data to quantify student learning both during the school year as well as during the summer. To more accurately measure the effects that school had on a student's learning, he adjusted the data to look at groups of students from different SES with comparable absenteeism during the school year. In his study, Ready found that in reading, low-SES children actually learn more during their first two years of school than their high-SES classmates. He also concluded that while students of average-SES stay at about the same cognitive level during the summer recess, children of high-SES show gains, while children of low-SES show literacy-skill decreases. That finding supports the premise that achievement differences between low-SES children and high-SES children are not a function of the school, but rather occur due to out-of-school factors.

Differences in the academic growth of children during the summer recess are noteworthy in light of the NCLB and the Global Educational Reform Movement (GERM) described by Pasi Sahlberg in his 2012 book *Finnish Lessons*. Among other things, NCLB and other global accountability initiatives focus on using annual student-assessment data as a measure of teacher and school effectiveness. In the U.S., NCLB specifies minimum achievement scores for all children disaggregated by SES, race, gender, English language learner (ELL) status, and disability. It mandated that all students in all subgroups would be proficient in math and reading by 2014. Citing flaws in the current law, President Obama asked Congress in March 2010 to reauthorize ESEA with changes. When there appeared to be no movement in the ESEA reauthorization process, President Obama in September of 2011, chose to exercise a clause in ESEA that allowed Education Secretary Arne Duncan to grant waivers to states for certain provisions of ESEA.

In the U.S. Department of Education's invitation for states to be granted a waiver from the provisions of NCLB they asked for, "…rigorous and comprehensive state-developed plans designed to improve educational outcomes for all students, close achievement gaps, increase equity, and improve the quality of instruction" (htttp://www2.ed.gov/policy/elsec/guid/esea-flesibility/index.html). In the state of Maine, according to the waiver submitted on September 6, 2012, that meant adopting the Common Core State Standards CCSS as Maine's learning standards, and transitioning from the NECAP assessment to the SMARTER Balance Assessment Consortium (SBAC) as the annual measure of student progress. The state of Maine's application also included two additional features germane to this study. First, it set a "school accountability index" (Waiver p. 59) that uses annual measures to determine a school's performance in reading and math. Second, it passed LD 1859, "An Act to Ensure Effective Teaching and School Leadership." LD 1859 is Maine's response to section 3 of the ESEA waiver, which mandates that the state educational agency (SEA**)** "…develop and adopt guidelines for local teacher and principal evaluation and support systems by the end of the 2012-2013 school year." LD 1859 mandates that a "significant" part of the teacher and principal evaluative systems include student growth. While Chapter 180 defines "significant" as consisting of at least 20 percent of any evaluation system, local educational authorities (LEA) are responsible for determining which assessment will be used to measure student growth. Although far from certain, it appears likely that the statewide annual assessment will be chosen by at least some LEAs.

While the Coleman Report was correct in measuring annual student growth to determine if American children were moving toward greater racial equality, the report's error was in assuming that a student's annual growth is solely a function of the school a child attends. A mounting body of research suggests that differences between school year growth and annual growth is a function of the differences in learning both inside and outside of school. That is the fundamental problem with school accountability: while it does matter how much a child learns, it does not matter where that learning takes place. The great civil rights issue of closing the achievement gap between different groups of students might not be about how to fix schools, but rather how society can make up for differences that children experience during times school is not in session.

The current measure of student achievement based on annual growth is accurate and appropriate for measuring a child's academic development but, because it also measures what a child learns or does not learn during the summer months when school is not in session, it is an unreliable measure of the effectiveness of a school or a teacher. A more accurate measure of the effectiveness of a teacher or a school would be to measure a child's academic performance at the beginning of the school year and then again at the end to control for non-school influences on student learning that occur during the summer. Another way to address the inaccuracies of using an annual assessment would be to mathematically account for differences in summer learning in growth models.

Summary

When compared to other countries, the United States educational system is characterized by a long summer recess (Wiseman, Baker 2004). Over the past century educational researchers such as Bruene (1928) and Cook (1947) tried to determine why the long recess affected children differently. Heyns (1978) made the argument that students learn both in and out-of-school. To isolate out-of-school factors, she measured student achievement gains from spring to fall, and compared them to student growth from fall to spring. What she found was that all children made relatively consistent gains during the school year, but had dramatically different learning patterns during the summer recess. Her research conflicted with the Coleman Report (1966), which concluded that schools do not have a dramatic effect on learning. Heyns showed that schools do indeed have a great deal of influence on learning, but so to do out-of-school factors that are especially apparent during the summer recess when school is “controlled for.”

Since Heyns’ seminal book, several researchers have attempted to quantify summer learning loss for different categories of children. The research of Stanovich, 1986, Cooper et al., 1996, Downey et al., 2004, Entwisle et al., 1992, 1997, 2001, Alexander et al, 2001, 2007, and Vales et al., 2013, indicates that summer learning-loss affects economically disadvantaged children much more than it does non-disadvantaged children. Nonetheless, accountability systems continue to use annual measures to determine the effectiveness of educators or institutions, and therefore continue to ignore the mounting evidence about summer learning-loss.

Like other states, Maine uses an annual assessment to measure the performance of teachers, schools, and districts, and therefore policymakers might want to examine the affects that summer learning has on Maine children. This study examined summer learning in mathematics and reading for students in grade 3 through grade 8 in the state of Maine by comparing student performance on the 2009 spring MEA assessment to the 2009 fall NECAP assessment.

CHAPTER III: METHODOLOGY

Purpose

There is a growing body of evidence that suggests that summer learning-loss affects economically disadvantaged children to a greater extent than non-disadvantaged children (Stanovich, 1986; Cooper et al., 1996; Downey et al., 2004; Entwisle et al., 1992, 1997, 2001; Alexander et al, 2001, 2007; Vales et al., 2013). While the educational research suggests that different students have different learning experiences during summer recess, local and federal accountability measures fail to take those variations into account in their calculations, due to their reliance on annual assessments to measure nine and a half months of instruction. In federal legislation, the use of an annual assessment to determine the effectiveness of a school or teacher started with the NCLB Act of 2001 but the root of the practice can be traced back to the 1966 Coleman Report. That report used annual assessment data to conclude that schools had very little impact on student learning.

In the state of Maine, school accountability measures are currently tied to the annual NECAP assessment. Starting in SY15, those accountability measures were to be connected to the SBAC assessment: yet another annual assessment. While LD1859 does not specify that student achievement will be tied to SBAC, it was expected that many LEAs would use the SBAC assessment for that purpose as well. Because there is evidence that annual assessments contain an error factor as a result of their inclusion of the summer recess, their data may be misleading with respect to a school's influence on achievement. To assure that the data used to make judgments about programming and instructional effectiveness are correct, this study measures the error introduced by including the summer recess for students at different grades, and from different backgrounds.

The study is quantitative to capture a value or a range of values for the error factor created by including the extended time away from school that the summer recess adds to accountability measures. It takes advantage of the fact that the spring 2009 administration of the MEA assessment and the fall 2009 NECAP assessment in the state of Maine for all children in grades three through eight were both designed to measure the same learning based on the same learning standards. That change in test administration created a natural ex post facto research opportunity that allows for the isolation and measurement of student learning for different groups of Maine students during the summer recess of 2009.

Methodology

This study was organized to answer the following research questions:

1. Was there a statistically significant change in mathematics achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually, as well as for all grade spans combined.
2. Was there a statistically significant change in reading achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually, as well as for all grade spans combined.

3. Were there statistically significant differences in summer achievement changes in mathematics for students of differing SES? If so, what was the magnitude of those differences, and did they vary according to a child's grade level?
4. Were there statistically significant differences in summer achievement changes in mathematics for students of differing gender? If so, what was the magnitude of those differences, and did they vary according to a child's grade level?
5. Were there statistically significant differences in summer achievement changes in reading for students of differing SES? If so, what was the magnitude of those differences, and did they vary according to a child's grade level?
6. Were there statistically significant differences in summer achievement changes in reading for students of differing gender? If so, what was the magnitude of those differences, and did they vary according to a child's grade level?

While there is evidence in the literature that indicates that students from economically disadvantaged families show stalled learning, or even regression, over the summer break, there has not been research conducted on a largely rural population in grades three through eight to form a directional research hypothesis. Much of the work has been on large urban populations. For example, Heyns' (1978) work addressed the summer learning loss of seventh and eighth graders in Atlanta while, Alexander et al.'s series of papers from 1997 to 2007 worked with younger students in urban Baltimore. Although several studies of summer learning using ECLS-K, such as that done by Ready (2010), include non-urban students, the data set does not include the span of grades to be addressed in this study. Consequently the null hypothesis of this study is that there will not be a significant difference in summer learning on either the math or reading assessment for children in grades three through eight based on the following independent variables in this study:

a. Gender – Male/Female
b. Economic Status – Economically Disadvantaged/Non-Economically Disadvantaged

Note that all definitions are based on MDOE's standards and are a part of each student's Maine Department of Education Data System (MEDMS) record.

The study itself compares a student's MEA score in the spring with that same child's NECAP score in the fall, and therefore is a within subjects analysis also known as a repeated measures independent variable analysis. The "treatment" in this study is the summer recess of 2009 that created a situation in which in-school learning factors were controlled for, and thus out-of-school learning factors were the primary agents acting on student academic achievement.

Operational Definitions

Low-SES – Economically disadvantaged as defined in the MEA student demographics are those students who are eligibility for Federal Free or Reduced Lunch (FRL) program. The author acknowledges that there are some well-documented concerns with respect to using FRL status as a proxy for SES. According to Harwell and LeBeau (2010) "…A significant percentage of students are incorrectly certified as eligible or not eligible" (p. 124). The authors go on to assert that the, "variety of variables have served as SES measures…"(P.120) are ignored when using FRL as a proxy for SES. Nonetheless, the preponderance of the data attributing low achievement with SES use FRL. As an operational definition, even though FRL is an imprecise proxy for SES, it serves well enough to be a predictor of student achievement. For the purposes of this study, all classifications are based on a child's MEA classification. Coding in the results section of this analysis for economically disadvantaged will be high-SES for non-economically disadvantaged children, and low-SES for economically disadvantaged.

Maine Educational Assessment (MEA) – In 2009 the MEA was a standardized assessment designed to assess all Maine students in grades three through eight on their reading and mathematics abilities. It also consisted of a science and writing assessment for students in grades five and eight. For the purposes of this study only reading and mathematics will be addressed.

New England Common Assessment Program (NECAP) – In the fall of 2009, the NECAP replaced the MEA as the assessment to measure reading and mathematics ability for all Maine students in grades three through eight. It also has a writing component for students in grades five and eight but for the purposes of this study these will not be included.

Summer Learning - This term is used to describe the influence that time away from school has on student achievement. It is not directional, and does not imply either an improvement or degradation in achievement scores.

Sample and Data Procedures

Students in this sample will include all Maine students who took the end-of-year MEA assessment in the spring of 2009, and the beginning-of-year NECAP assessment in the fall of 2009. (Because the students were tested in two different school years, the sample will "lose" SY09 eight graders who would not have taken the NECAP in the fall of their ninth grade year in SY10.)

Fundamental to this study is the assertion that both assessments were designed to measure the same learning standards. That is, the spring third-grade MEA assessment was constructed to measure third-grade learning standards from the MLR. The following year when those children were promoted to fourth grade, they were to have taken the fourth-grade NECAP assessment. The fourth-grade NECAP assessment was constructed to measure the prior year's learning standards: third grade learning. Therefore, during the year that Maine switched from the MEA to the NECAP, in the fall of 2009 the children in grades four through eight were assessed twice on the same learning standards: once in the spring of the prior year (SY09) and then again in the fall of the current year (SY10). It was the intent of this study to use the duplication of assessments in Maine as a natural experiment to measure differences in learning that occurred for different categories of students. Because the period between the spring MEA and the fall NECAP assessment was the part of the year with the greatest amount of time away from school, it represented a period when out-of-school factors had the greatest influence on student learning.

Sampling

In accordance with Maine state statute, “Each school administrative unit and each student enrolled in a school covered by this rule shall participate in the Maine Education Assessment (MEA) in grades 4 through 8 (Chapter 127 § 4.1).” Therefore, in the spring of 2009 the 70,497 students who were enrolled in grades three through seven were required to take the MEA. In the following fall 70,622 were enrolled in grades four through eight. Table 3.1 shows the break-down of the sample size of the raw data set for each grade studied, as well as the year-to-year difference in enrollment. From that data set, students who did not take either of the two assessments, students who were retained or skipped a grade, students who moved in or out of the state, and students who took the Personalized Alternative Assessment Portfolio (PAAP) are excluded from the study sample.

Table 3.1

Sample Size by Grade Level

School Year	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**
2008-09	13,782	13,822	14,146	14,272	14,475
School Year	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
2009-10	13,753	13,891	14,221	14,337	14,420
Difference	-29	69	75	65	-55

Instrumentation

Maine Educational Assessment (MEA)

The MEA assessment was administered to all students in grades three through eight. According to the 2009 MEA technical manual, "The MEA is designed to determine the extent to which students know and are able to do what is articulated in Maine's 2007 Learning Results: Parameters for Essential Instruction (MLRs)" (MeCAS Technical Report, p.3). According to the Maine Department of Education website, "When the Maine State Legislature adopted the initial *Learning Results* in 1996...The legislation also required a new system for assessing student progress resulting in the MeCAS program. The Maine Educational Assessment (MEA) fulfilled this requirement…" (http://www.maine.gov/education/lsalt/). MeCAS is the Maine comprehensive assessment system, and among other things it required that starting in 1997, all students in Maine be assessed by the MEA to measure their progress towards meeting the standards of the Maine Learning Results (MLR).

The MEA was created by Measured Progress, an assessment company based in Dover, New Hampshire. It was administered to Maine public school students from 1985 until 2009. According to the 2008- 2009 MeCAS technical manual, the assessment consists of three types of questions:

- Multiple-choice items (MC) were used to provide breadth of coverage of a content area. Because they require no more than a minute for most students to answer, these items make efficient use of limited testing time and allow coverage of a wide range of knowledge and skills.

• Short-answer items (SA) were used only in mathematics to assess students' skills and their abilities to work with brief, well-structured problems that have one or a very limited number of solutions. Short-answer items require approximately 2 to 5 minutes for most students to answer. The advantage of this item type is that it requires students to demonstrate knowledge and skills by generating, rather than merely selecting, an answer.

• Constructed-response items (CR) typically require students to use higher order thinking skills—evaluation, analysis, summarization, and so on—in constructing a satisfactory response. Constructed-response items should take most students approximately 7 to 10 minutes to complete. (p. 5)

The MEA was designed by Measured Progress then reviewed by, "…item review committees composed of Maine classroom teachers, curriculum supervisors, higher education faculty, content specialists of the MDOE, and curriculum and assessment specialists at Measured Progress" (MeCAS Technical Report, P.9). Scoring quality is maintained by embedded committee review responses, read behind procedures, double scoring, recalibration sets and scoring reports. Student raw scores are then converted to an eighty-point scale, and cut points are made for the various achievement levels. Scores are reported in a three-digit format with the hundreds place representing the student's grade, and the tens and ones places representing the child's achievement on the eighty-point scale.

New England Common Assessment Program (NECAP)

The NECAP was also created by Measured Progress. A collaboration among the New England states of New Hampshire, Rhode Island and Vermont, NECAP was intended to measure student achievement, and meet the annual student assessment requirements of NCLB. It was piloted in SY05 and was administered in all three founding states in the fall of 2005. When Maine joined the NECAP collaborative in 2009, "Teachers and other education professionals from the four states participated in the March 2009 Item Review Committees, and Bias and Sensitivity Review meetings, in order to provide recommendations for field test items" (http://www.measuredprogress.org/necap). Because "NECAP test items are directly linked to the *content standards* and *performance indicators* described in the GLEs/GSEs" (NECAP Technical Manual 2009-2010, p.4), according to the MDOE website, the NECAP replaced the MEA to, "certify achievement of Maine's learning standards as articulated in Chapter 131 legislation" (http://www.maine.gov/education/lsalt/).

Unlike the MEA, the NECAP is administered in the fall. According to the 2009-2010 NECAP Technical Manual, "It is important to note that the NECAP tests in reading, mathematics, and writing are administered in the fall at the beginning of the school year and test student achievement based on the *prior years* GLEs/GSEs" (NECAP Technical Manual 2009-2010, P.1). It is that divergence in the NECAP's test administration timeline from the MEA that makes this study possible.

Like the MEA, the NECAP consists of multiple-choice, short answer and constructed response items. Scoring quality is maintained by embedded committee review responses, read behind procedures, double scoring, recalibration sets and scoring reports. Like the MEA, student raw scores on the NECAP are scaled on an eighty-point scale, and cut points are made for various levels of achievement.

Both the MEA and NECAP use Item Response Theory (IRT) to calibrate all items and cut points for each performance level derived from raw scores using the Test Characteristic Curve (TCC) plot of Theta to Raw score. While both scales are eighty-point scales, cut points on the MEA and NECAP are similar but not the exactly the same. Tables 3.2 and 3.3 show in great detail the difference between the MEA proficiency cut points and the NECAP proficiency cut points. Most notably in that difference is that the MEA uses the formula X42 as the cut off for Proficient and the NECAP uses X40 (Where X is used to denote the grade level).

Table 3.2

MEA/NECAP Reading Proficiency Cut Points
2009

Grade	Assessment	Min.	DMS/ PMS	# of Points	PMS / MS	# of Points	MS/ ES	# of Points	Max.
3	MEA	300	332	32	342	10	362	20	380
	NECAP	300	331	31	340	9	357	17	380
	Difference	0	-1	-1	-2	-1	-5	-3	0
4	MEA	400	432	32	442	10	462	20	480
	NECAP	400	431	31	440	9	456	16	480
	Difference	0	-1	-1	-2	-1	-6	-4	0
5	MEA	500	532	32	542	10	562	20	580
	NECAP	500	530	30	540	10	556	16	580
	Difference	0	-2	-2	-2	0	-6	-4	0
6	MEA	600	630	30	642	12	662	20	680
	NECAP	600	629	29	640	11	659	19	680
	Difference	0	-1	-1	-2	-1	-3	-1	0
7	MEA	700	730	30	742	12	762	20	780
	NECAP	700	729	29	740	11	760	20	780
	Difference	0	-1	-1	-2	-1	-2	0	0
8	MEA	800	830	30	842	12	862	20	880
	NECAP	800	828	28	840	12	859	19	880
	Difference	0	-2	-2	-2	0	-3	-1	0

Table 3.3

MEA/NECAP Math Proficiency Cut Points 2009

Grade	Assessment	Min.	DMS/ PMS	# of Points	PMS / MS	# of Points	MS/ ES	# of Points	Max.
3	MEA	300	326	26	342	16	362	20	380
	NECAP	300	332	32	340	8	353	13	380
	Difference	0	6	6	-2	-8	-9	-7	0
4	MEA	400	430	30	442	12	462	20	480
	NECAP	400	431	31	440	9	455	15	480
	Difference	0	1	1	-2	-3	-7	-5	0
5	MEA	500	530	30	542	12	562	20	580
	NECAP	500	533	33	540	7	554	14	580
	Difference	0	3	3	-2	-5	-8	-6	0
6	MEA	600	628	28	642	14	662	20	680
	NECAP	600	633	33	640	7	653	13	680
	Difference	0	5	5	-2	-7	-9	-7	0
7	MEA	700	728	28	742	14	762	20	780
	NECAP	700	734	34	740	6	752	12	780
	Difference	0	6	6	-2	-8	-10	-8	0
8	MEA	800	830	30	842	12	862	20	880
	NECAP	800	834	34	840	6	852	12	880
	Difference	0	4	4	-2	-6	-10	-8	0

Data Analysis

As tables 3.2 and 3.3 indicate, while the MEA and NECAP scales are similar, they are not the same. Nonetheless, for the purpose of the first part of this study there is no need for the two scales to be the same, only that both scores represent an assessment of the same or similar content. Descriptive statistics comparing the MEA spring with the NECAP fall will be conducted to determine the “shape” of the relationship, and determine the data’s skewedness and kurtosis. If the data conform to expected results, a regression analysis for each grade level and each content area (math, reading) with MEA spring scores being the independent variable and NECAP fall scores being the dependent variable will be performed. It is expected that regression will be highly correlated as both assessments measured the same Grade Level Expectations (GLE).

Residual Values

If the formula for the line created by the entire set of students in a given grade at a given content area perfectly predicts a student's NECAP fall score using that same student's MEA spring score, than the residual for that student by definition will be zero. If on the other hand, different groups of students (i.e. SES, gender) have different patterns of residual values, that will be an indicator that those groups experienced different effects from the treatment. For the purposes of this study the treatment is the effect that the summer vacation of 2009 had on different populations of students. Should there be a pattern for the residuals that has a statistical significance (p-value) of less than 0.01 then the null hypothesis can be rejected.

Analysis of Variance

The predictor variables are all binary in nature, and thus will be dummy coded (Table 3.6). Due to the nature of the independent variables, and that there is more than one independent variable, analysis of variance (ANOVA) will be performed on the data to determine if the variance between the predictor variables is statistically significant. ANOVA will be done independently for each grade (4 through 8 Fall 2009) and content area (mathematics and reading) resulting in ten independent results. Each of those results will produce data with respect to each of the dummy variables coded in table 3.4 to determine the direction and magnitude of summer learning for different groups sorted by grade and content area. This data will then provide relevant feedback as to the direction and magnitude of the error factor included when annual assessments are used to measure the effectiveness of the instructional year for Maine students.

Table 3.4 – Gender Coding

Predictor Variables	Coding	
	0	1
Gender	Male	Female
SES	No	Yes

Setting up the Data

The data for this study come from the MDOE's MEDMS database. In accordance with FERPA and accepted ethical research standards, the data was cleaned of individually identifiable information by the data manager at the Center for Educational Policy, Applied Research and Evaluation (CEPARE). To create the test/retest scenario, each student was assigned a unique project ID number that allowed each student's spring MEA data to be paired with that same student's fall NECAP data. Data analysis was done using IBM SPSS Version 22 software.

The data set consisted of all students in the state of Maine in grade three through grade seven who took the 2009 MEA spring assessment and all students in the state of Maine in grade four through grade eight who took the fall NECAP assessment. That made for a total starting sample size of 70,477 MEA student scores and 70,796 NECAP student scores.

The data were duplicated and each of the two identical data sets for each grade was assigned to either the math or the reading analysis for content-specific participation adjustments. To make those adjustments, the data were sorted to remove students who had any missing scores for either of the tests being analyzed. Students missing a particular score represent children who did not complete a particular section of the test for a variety of reasons such as moving out of the state during the assessment, illness, or refusal to complete a particular component of the test. The cleaned data were then paired by project ID with their corresponding cleaned data set: MEA Grade X with NECAP Grade X+1 for a particular content area. The data were again sorted to remove all students with only one score from either test. Such students presumably moved into the state, left the state, refused to do a particular section in a particular year, or were otherwise excused from either the spring MEA or the fall NECAP. The final usable sample for the analysis of reading was (N = 66,828) and for mathematics (N = 66,857). Table 3.5 shows a complete listing of the sample size after the data were prepared for study for each grade level and content area in the investigation, along with each grade level's percentage of the total sample.

Table 3.5 - Sample Size by Grade

Grade	Reading	Reading Percentage	Mathematics	Mathematics Percentage
3 and 4	13,036	19.5	13,043	19.5
4 and 5	13,189	19.7	13,210	19.8
5 and 6	13,472	20.2	13,451	20.1
6 and 7	13,487	20.2	13,510	20.2
7 and 8	13,644	20.4	13,643	20.4
Total:	66,828	100	66,857	100

Scale and Equivalence of the Assessments

While the MEA and the NECAP were designed to measure the same learning standards for the children of Maine, and both have technical similarities, the assessments themselves are not the same. They do not have the same number of questions, possible points, or raw-score proficiency cut-points. What the MEA and NECAP do have in common is that both assessments have a comparable 4-point proficiency scale, as well as an underlying design requirement to measure the same learning standards. As stated above, both 4-point proficiency scales were designed to report on a student's progress toward meeting the standards of the MLR. Because the MLR did not undergo any changes over the summer of 2009, the scales are assumed to be equivalent.

To create statistical comparability between the two assessments, as well as to address the limitations of using data from different assessments, the following three approaches were used for data analysis: raw score comparison; 4-point linked scale conversion; and Z-score comparison. For the majority of the analysis, the 4-point linked scale was used because it corresponds directly to a student's achievement level. Other scales were primarily used to confirm the finding-point linked scale findings.

The 4-point Linked Scale

The MEA and NECAP assessments both rely on a 4-point proficiency level scale (1-4) and both also have the same number of scaled score values (0-80). Nonetheless, a complicating factor for comparative purposes was that the raw score and scaled-score cut points for proficiency were different for the two assessments.

That can be seen in Tables 3.2 and 3.3. To address that comparability challenge, this study took advantage of the fact that the 4-point proficiency measure for both assessments represents a parallel scale from which a comparison of the student achievement on each assessment could be based. That was possible because the proficiency levels between the tests are defined as corresponding to the same level of proficiency that a student must reach in order to meet the learning standards of the MLR. While the 4-point proficiency scale provided by Measured Progress for both assessments represents a clear link between the two assessments, in its nominal form it was rejected as being too crude a measure to pick up the potentially subtle changes in learning patterns.

To address the lack of sensitivity in the nominal 4-point proficiency scale, a conversion was made to turn each of those scales into a real number scale. The 4-point linked, or real number, scale takes advantage of the fact that there are a variety of raw scores within each performance level. Using the assumption that there is a relationship between a child's raw score in a particular performance level and that child's academic achievement, different raw scores within a performance level were weighted to create a much more-sensitive measure of a student's performance level. That assumption was founded on Measured Progress' practice of generating scaled scores from raw scores. Table 3.6 shows that conversion for grade three MEA mathematics and grade four NECAP mathematics as a basis for the explanation that follows.

Table 3.6

4-Point Linked Scale Conversion by Performance Level

Proficiency Level																													
MEA 1	Raw Score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15													
	Conversion	0.06	0.13	0.19	0.25	0.31	0.38	0.44	0.50	0.56	0.63	0.69	0.75	0.81	0.88	0.94													
MEA 2	Raw Score	16	17	18	19	20	21	22	23	24	25	26																	
	Conversion	1.00	1.09	1.18	1.27	1.36	1.45	1.55	1.64	1.73	1.82	1.91																	
MEA 3	Raw Score	27	28	29	30	31	32	33	34	35	36	37	38	39															
	Conversion	2.00	2.08	2.15	2.23	2.31	2.38	2.46	2.54	2.62	2.69	2.77	2.85	2.92															
MEA 4	Raw Score	40	41	42	43	44	45	46	47	48																			
	Conversion	3.00	3.13	3.25	3.38	3.50	3.63	3.75	3.88	4.00																			
NECAP 1	Raw Score	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
	Conversion	0.03	0.07	0.10	0.14	0.17	0.21	0.24	0.28	0.31	0.34	0.38	0.41	0.45	0.48	0.52	0.55	0.59	0.62	0.66	0.69	0.72	0.76	0.79	0.83	0.86	0.90	0.93	0.97
NECAP 2	Raw Score	29	30	31	32	33	34	35	36	37	38																		
	Conversion	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90																		
NECAP 3	Raw Score	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53													
	Conversion	2.00	2.07	2.13	2.20	2.27	2.33	2.40	2.47	2.53	2.60	2.67	2.73	2.80	2.87	2.93													
NECAP 4	Raw Score	54	55	56	57	58	59	60	61	62	63	64	65																
	Conversion	3.00	3.09	3.18	3.27	3.36	3.45	3.55	3.64	3.73	3.82	3.91	4.00																

Table 3.6 illustrates the differences in scoring between the two assessments. It shows the 4-point linked scores from both assessments for a comparison of student achievement over the summer of 2009. The 4-point linked scores were derived by using the specific cut-point for each proficiency level of the given test as the anchor, and dividing the whole number for that proficiency level into the parts of whole using the number of raw points possible for that proficiency level. For example, as seen in Table 3.6, a child with a MEA proficiency level of 1 could have had a raw score of 0 to 15 points. For a student who had a proficiency level of 1 on the MEA, by dividing that student's raw score by 16 (the raw score needed for a proficiency score of 2), the 4-point linked scale was created. A similar calculation was done for each of the four performance levels for all of the tests. The following is the formula that was used to calculate the 4-point linked scale for all achievement levels, grades:

$$S_{4p} = ((S_{Raw} - S_{PC} / S_{pp}) + (P - 1)$$

Whereas S_{4p} is the 4 point linked score, S_{Raw} is the raw score, S_{PC} is the lower-proficiency cut point, S_{pp} is the total number of points within a proficiency level and P is the proficiency level. For a child with a MEA raw score of 32 on the MEA mathematics assessment, the calculation would be as follows:

$$S_{4p} = ((32 – 27) / 13) + (3-1)$$

$$S_{4p} = 2.38$$

That formula would convert the child's initial 4-point nominal score of a 3 to a value of 2.38 in the 4-point linked scale, and would allow that child's achievement to be contrasted with other children who scored a 3 on the nominal scale, but had different raw scores.

Comparability Assessments Using Pearson Correlation

This study uses the MEA as the independent variable and NECAP as the dependent variable in a test-retest or repeated measures format. To confirm the comparability of MEA and NECAP scores, a bivariate Pearson Correlation was performed for each data set in the study for both the 4-point linked conversion and a raw score comparison (see Table 3.7 and 3.8). For all grades in both math and reading for both raw score comparison as well as for the 4-point linked scale, the MEA assessment was a statistically significant, $p < .001$ (2-tailed) predictor of that same student's performance on the fall NECAP with a mean r value for math raw scores of 0.837 and a mean r value for math linked 4-point scale of 0.829. For reading, the mean r-value for raw scores is 0.773 and 0.772 for the linked 4-point scale. The results suggest that there is a high degree of correlation between the two assessments, both in the raw score analysis and with the 4-point linked analysis.

Table 3.7 - Mathematics Pearson Correlation

	Raw Score Analysis			4-point Linked Score Analysis	
Grade	Pearson Correlation	Sig. (2-tailed)		Pearson Correlation	Sig. (2-tailed)
3 and 4	.805**	0.000		.802**	0.000
4 and 5	.814**	0.000		.806**	0.000
5 and 6	.841**	0.000		.828**	0.000
6 and 7	.857**	0.000		.850**	0.000
7 and 8	.868**	0.000		.861**	0.000

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

Table 3.8 - Reading Pearson Correlation

	Raw Score Analysis			4-point Linked Score Analysis	
Grade	Pearson Correlation	Sig. (2-tailed)		Pearson Correlation	Sig. (2-tailed)
3 and 4	.756**	0.000		.752**	0.000
4 and 5	.735**	0.000		.738**	0.000
5 and 6	.764**	0.000		.765**	0.000
6 and 7	.802**	0.000		.802**	0.000
7 and 8	.808**	0.000		.803**	0.000

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

Comparability and Equivalence Using Cronbach's Alpha

Because it was fundamental to this research that there was comparability between the MEA and the NECAP, an analysis of the equivalence of the two assessments was also conducted using Cronbach's Alpha. According to Cronbach (1951),

> A retest after an interval, using an identical test, indicates how stable scores are and therefore can be called a coefficient of *stability*. The correlation between two forms given virtually at the same time, is a coefficient of *equivalence*, showing how nearly two measures of the same general trait agree. Then the coefficient using comparable forms with an interval between testings is a coefficient of equivalence and stability (p.298).

In 2004, Cronbach described that idea of equivalence of two forms of a test with the following;

> At the other extreme were random-parallel tests, where each test was (or could reasonably be regarded as) a random sample from a specified domain of

> admissible test items. It was the latter level of parallelism that seemed best to explain the function of coefficient alpha; it measured the consistency of one random sample of items with other such samples from the same domain (p. 400).

Given that the MEA and NECAP were designed to assess the same learning standards, they can be seen in Cronbach's terms as "parallel tests" with "a random sample (of questions) from a specified domains of admissible test items."

According to the NECAP technical manual (p.57), Cronbach's α for raw scores for the NECAP for grades four through seven for the different versions of the assessment ranged from 0.87 to 0.90 for reading and from 0.92 to 0.93 for mathematics. The MEA Technical Manual (p.54) reports Cronbach's α for raw scores for grades four through seven ranging from 0.87 to 0.90 for reading, and from 0.92 to 0.93 for mathematics. In comparing the equivalence of the MEA with the NECAP using calculations of Cronbach's α, similar levels of consistency were demonstrated between the MEA and NECAP as were found in the internal measures of equivalence for different versions of the MEA and NECAP (Table 3.9). The result provides a statistical basis for treating the MEA and the NECAP as a situation "where each test was (or could reasonably be regarded as) a random sample from a specified domain of admissible test items" (p.400). That condition allows for the supposition that differences in scores for each child can be attributed to external factors rather than being a function of the assessment itself.

Table 3.9 Comparability and Equivalence - Cronbach's α

Cronbach's α Mathematics				
	Raw Score α	4-point Linked α	MEA2009 α	NECAP 2009 α
MEA 3 NECAP 4	0.892	0.888	0.93	0.92
MEA 4 NECAP 5	0.898	0.893	0.92	0.92
MEA 5 NECAP 6	0.913	0.906	0.92	0.92
MEA 6 NECAP 7	0.923	0.917	0.92	0.92
MEA 7 NECAP 8	0.929	0.925	0.92	0.93
Cronbach's α Reading				
	Raw Score α	4-point Linked α	MEA2009 α	NECAP 2009 α
MEA 3 NECAP 4	0.861	0.859	0.89	0.89
MEA 4 NECAP 5	0.847	0.849	0.89	0.87
MEA 5 NECAP 6	0.866	0.867	0.87	0.89
MEA 6 NECAP 7	0.89	0.89	0.89	0.9
MEA 7 NECAP 8	0.894	0.89	0.9	0.88

Summary

When in 2009 the state of Maine switched from the spring MEA assessment to the fall NECAP assessment to measure the same student learning on the same GLEs separated by the summer break, a natural experiment to measure summer learning was created. While the two assessments are not identical, they both were designed to measure the same learning. The comparability of the two assessments was verified by conducting a regression to determine how accurately a child's spring MEA score predicted that same child's fall NECAP score. Once comparability was confirmed, patterns of difference were analyzed to determine if there was a statistically significant relationship between a child's predicted NECAP score and that child's actual score (on both math or reading) based on that child's gender, grade level or SES.

While the MEA and NECAP assessments were spaced approximately six months apart, during which time students had just over three months of additional schooling, the students were also on a nearly three-month summer recess. Because a great deal of research, most notably from Heyns (1978), Cooper et al. (1996), Entwisle (1997), Downey et al. (2004), and Alexander et al. (2007), indicates that students learn at about the same rate during the school year, statistically significant patterns of student's expected achievement as compared to their actual achievement on the NECAP are to be attributable in large part to out-of-school influences that occurred during the summer recess. Furthermore, because those differences in learning are not attributable to the influence of schooling they should be considered in any educational accountability measure.

CHAPTER IV: FINDINGS

Overview of the Study

The purpose of this research was to examine the impact of time away from school on student learning. More importantly it attempts to determine if there are statistically significant differences in learning over the summer that are attributable to factors exogenous from school such as gender, poverty status, or grade level. The foundational assumption is that a student's academic achievement is a product of both in-school and out-of-school factors. To accurately measure one, the other must be controlled for in some way. Because many out-of-school factors are constant (e.g., gender), they are nearly impossible to isolate. In contrast, schooling is discontinuous.

While there are many breaks from schooling during the calendar year, the largest by far is the summer recess. That makes summer recess the best de facto control for the direct effects of schooling on student achievement. By analyzing student achievement data measured at the end of one school year, and then again at the beginning of the next school year, variations in learning between different groups of students during a time when the effects of school are controlled for can be exposed. That method of measuring the part of student learning that is a result of the non-school factors by examining learning patterns when school is not in session has a long history in the educational literature (Bruene, 1928; Cook, 1942; Stanovich, 1986; Heyns, 1978, 1987; Cooper et al., 1996; Downey et al., 2004; Alexander et al., 1997, 2001, 2007; Entwisle et al., 1997, 2001; Vales et al., 2013). As Heyns succinctly stated, "…achievement is a continuous process, whereas schooling is intermittent." Furthermore, she continued: "As a quasi-experimental control for the effects of education (schooling), the summer months represent a plausible interval in which to contrast patterns of learning" (1978, p.43).

The data for this study were created due to a natural experiment that arose when the state of Maine changed from assessing students' progress toward meeting the standards of the MLR from the spring-administered MEA to the fall-administered NECAP. Because "NECAP test items are directly linked to the *content standards* and *performance indicators* described in the GLEs/GSEs" (NECAP Technical Manual 2009-2010, p.4), student achievement on the NECAP is comparable to student achievement on the MEA. The MDOE reinforced that claim on its website, stating that the NECAP replaced the MEA as the measure to, "…certify achievement of Maine's learning standards as articulated in Chapter 131 legislation" (http://www.maine.gov/education/lsalt/). The veracity of the DOE's claim regarding that level of comparability was confirmed using both a Pearson Correlation and Cronbach's Alpha.

The analysis was conducted from both an absolute change in achievement between the two assessments, as well as from the relative differences in change in achievement over the summer for students with different characteristics: gender, grade, and economic status. The absolute change in achievement is dependent on the equivalence of the two assessments' scales. That is, any claim of an absolute loss or gain for any particular grade or group of students is only as accurate as the equivalence of the scales. Given that the MLR standards did not change over the summer of 2009, there is a solid foundation to make that assumption. Nonetheless, further analysis with respect to relative changes between different nominal groups over the summer of 2009 was conducted both to highlight learning differences for different groups over the summer, as well as to circumvent any error that may be inherent in absolute changes.

Achievement Changes During the Summer

The time interval between the administration of the spring MEA and fall NECAP, namely the summer of 2009, was the control in this research for academic learning that occurs as a result of schooling. Because schooling does not directly influence a student's academic growth during the summertime, patterns of achievement that emerge during the summer recess are therefore largely attributable to the effects of out-of-school factors. The theory behind that approach is well documented in the literature on summer learning. Entwisle, Alexander and Olson in their 2001 paper, "Keep the Faucet Flowing" cogently explain the underpinnings of this pattern of learning with their "faucet theory." In their paper they asserted:

> …when school was in session, the resource faucet was turned on for all children, and all gained equally; when school was not in session, the school resource faucet was turned off and all did not gain equally. In summers, poor families could not make up for the resources the school had been providing, and so their children's achievement reached a plateau or even fell back. Middle-class families could make up for the school's resources to a considerable extent so their children's growth continued, though at a slower pace (p.2).

This research seeks to determine what happened to Maine students' reading and mathematics achievement during the summer of 2009 when the school resource “faucet” was turned off, and only out-of-school influences affected student achievement.

A multi-layered analysis of means was performed on each of the data sets. The variable list included the MEA 4-point joined scale and NECAP 4-point joined scale. The first set of analysis included paired sample t-tests with a child’s score on the MEA as the independent variable, and that same child’s NECAP score as the dependent variable. This analysis was conducted independently for both mathematics and reading. Each grade was analyzed independently as well as in aggregate. The analysis quantified the cumulative affect that time away from school had for all children, as measured by changes in mean performances between the MEA and NECAP assessment.

With a baseline of summer achievement changes established for all students in all grades in both mathematics and reading, a more in-depth analysis was then conducted to determine if there were differences in achievement-level changes over the summer for different categories of students. In this analysis the dependent variable was the change in mean scores between the MEA and NECAP assessments while the independent list included the binary categorical variables: MEA Economically Disadvantaged and MEA Gender. The choice to use the MEA distinction for economically disadvantaged rather than the NECAP distinction was intentional: a way to minimize students recently economically disadvantaged as a result of the recession of 2008. The assumption was that using the MEA would capture more long-term economically disadvantaged families, and therefore be more representative of the effects of persistent poverty.

The first analysis was conducted to determine the affect that time away from school had on mathematics and reading academic achievement for students in Maine during the summer of 2009. Because the area of study consisted of five different grade-spans each of the grade spans was analyzed individually, as well as in aggregate. The analysis was done separately for both mathematics and reading. The analysis that follows is organized by content area. Mathematics achievement changes are addressed first, followed by a parallel analysis for changes in reading achievement. SPSS Version 22 was used to conduct the paired sample t-tests that follow.

Mean Mathematics Changes Over the Summer of 2009

Research Questions Addressed

- Was there a statistically significant change in mathematics achievement scores for children in Maine during the summer of 2009? If so what was the magnitude of that change?" This question was analyzed for each grade span individually as well as for all grade spans combined.

The paired sample t-test analysis of the mean mathematics change over the summer of 2009 indicated a significant finding $p < .001$ for all grades combined, as well as for all grades individually except grades 7 to 8, which had a p value of 0.130 (see Table 4.1). That result supports the conclusion that there was a statistical basis for treating differences in mean performance between the MEA and NECAP as significant. A descriptive analysis indicated that students, over the summer between grades 3 and 4, had the greatest amount of summer learning loss in mathematics. For those students, the summer between grades 3 and 4 represented a loss of just over 2 tenths of a performance level or 40 percent of a standard deviation. For students over the summer between grades 4 and 5 the loss is more modest, at just under 1 tenth of a performance level (0.0838 or 16 percent of a SD) and for students in grades 5 and 6 the loss is just over 0.04 of a performance level (8 percent of a SD). The data shifted for children during the summer of grades 6 and 7, who showed a slight increase in mathematics performance (0.0478 or 10 percent of a SD). Finally, for children during the grades 7 to 8 summer there is a very slight loss of 0.0061 of a performance level in mathematics (1 percent of a SD).

Taken in total, the average loss during the summer for all children was a modest 0.0560 of a performance level or 11 percent of a SD. That finding is consistent with other researchers' conclusions regarding the effects of summer on learning such as Cooper et al., who suggest that, "the overall effect of summer vacation on standardized test scores is at issue, students appear at best to demonstrate no academic growth over the summer p.259."

Table 4.1 - Mathematics t-Test Summer 2009

	Mean Difference	SD	t	Sig.
Grades 3 to 4	-0.2037	0.5097	45.653	0.000
Grades 4 to 5	-0.0838	0.5310	18.126	0.000
Grades 5 to 6	-0.0402	0.5051	9.246	0.000
Grades 6 to 7	0.0478	0.4899	-11.337	0.000
Grades 7 to 8	-0.0061	0.4693	1.514	0.130
Total	-0.056	0.5082	28.480	0.000

Result: For mathematics the null hypothesis that there will not be a statistically significant change in summer learning was rejected for all grades studied except for children over the summer of grades 7 to 8 at the $p < .001$ level. The descriptive data suggested that during the summer of 2009, the youngest children in this study had the greatest declines in mathematics achievement, but that decline diminished for each subsequent grade of the study, and reversed for children in the upper grades of the study.

Mean Reading Changes Over the Summer of 2009

Research Questions Addressed

- Was there a statistically significant change in reading achievement scores for children in Maine during the summer of 2009? If so what was the magnitude of that change? This question was also analyzed for each grade span individually, as well as for all grade spans combined.

The paired sample t-test analysis of the mean reading change over the summer of 2009 showed a significant finding $p < .001$ for all grades combined, as well as with each grade when analyzed individually (see Table 4.2). The result supports the conclusion that there is a statistical basis for treating differences in mean reading performance between the MEA and NECAP as significant. A descriptive analysis indicated that students experienced achievement gains in the earliest grades of the study. Reading gains were at their largest level during the grades 3 to 4 summer at just over a tenth of a performance level (0.1120) or 21 percent of a standard deviation, but were less pronounced during the grades 4 to 5 summer at 0.0754 of a performance level (15 percent of a SD) and were even smaller for students over the grades 5 to 6 summer at 0.0579 of a performance level (12 percent of a SD).

In the upper grades of this study, the data indicated that student performance in reading declined over the summer. During the grades 6 to 7 summer, children showed a slight decline in reading performance of 0.0470 of a performance level (11 percent of a SD). Finally during the grades 7 to 8 summer, students showed the greatest loss in achievement, regressing nearly one-sixth of a performance level (-0.1510 or 32 percent of a SD). The total change over the five summers studied was a negligible gain of 0.0081 of a performance level or just under 2 percent of a SD.

Table 4.2 - Reading t-Test Summer 2009

	Mean Difference	SD	t	Sig.
Grades 3 to 4	0.112	0.5183	-24.672	0.000
Grades 4 to 5	0.0754	0.4786	-18.109	0.000
Grades 5 to 6	0.0579	0.4641	-14.491	0.000
Grades 6 to 7	-0.047	0.4217	12.945	0.000
Grades 7 to 8	-0.151	0.4424	39.863	0.000
Total	0.0081	0.4755	-4.408	0.000

Result: For reading the null hypothesis that there will not be a statistically significant change in summer learning was rejected for all grades studied at the $p < .001$ level. The data suggest that the summer of 2009 had an effect on student achievement for every grade studied. The descriptive analysis indicated that while children improved in their reading-achievement levels during the earliest grades of the study, the trend reversed in the upper grades, where children in the highest grades of this study showed summertime regression.

Summer Achievement Changes for Different Groups of Students

The above analysis addressed the question of student learning over the summer for all students in each grade level studied both individually and in aggregate for reading and mathematics. The analysis used absolute values of achievement as measured by the 4-point linked scale. Differences over the summer were reported as changes in mean performance, based on the fraction of a performance level, or the percentage of a standard deviation changed. The following analysis shows the findings disaggregated by SES and gender. It is an analysis of differences in learning that occurred over the summer of 2009 between different groups of students, and seeks to answer the questions:

1. Did students of different SES have different patterns of academic growth in reading or mathematics during the summer recess when school was not a contributing factor to learning?
2. Did students who differ in gender have different patterns of academic growth in reading or mathematics during the summer recess when school was not a contributing factor to learning?

A split-plot ANOVA using SPSS version 22 was performed on a combined all-grades data set, as well as with each of the five individual grade-span data sets for both reading and mathematics to determine if there were statistically significant differences in summer achievement for different categories of students. The initial measure for the split-plot ANOVA was a child's performance on the spring MEA, while the second measurement was that same child's performance on the fall NECAP in the same content area. All inferential analyses were conducted using the 4-point linked scale and results were verified using an analysis of both Z-scores and raw scores. The between-subjects factors or independent variables explored in this analysis were MEA Economically Disadvantaged and MEA Gender.

The analysis that follows is organized in two sections. The first section is an analysis of mathematics, followed by a parallel investigation of reading. The first part of each of the analyses is an inferential investigation to establish the statistically significant independent variables. All significant findings are explored in a descriptive analysis.

Mathematics

Analysis of Achievement Changes in Mathematics Over the Summer of 2009 with Between Subjects Factors.

Research Questions Addressed

1. Were there statistically significant differences in summer achievement changes in mathematics for students of differing SES? If so what was the magnitude of those differences, and did they vary according to a child's grade level?
2. Were there statistically significant differences in summer achievement changes in mathematics for students of differing gender? If so what was the magnitude of those differences, and did they vary according to a child's grade level?

The split-plot ANOVA of mathematical achievement scores for each of the grade spans in the study indicated that in all cases students classified as MEA economically disadvantaged had a statistically significant difference in mathematics achievement during the summer than did their non-economically disadvantaged classmates. For determining statistical significance all cases had a $p < .001$. The partial η^2 in the analysis varied from 0.085 to 0.093, which according to Cohen indicates that there was a moderate effect size, based on a child's classification of economically disadvantaged. Gender had a statistically significant effect on student mathematics achievement only for students during the summer of grades 5 and 6 and grades 6 and 7 with $p < .001$. However the partial η^2 indicated a negligible effect size of between 0.000 and 0.001 respectively. The data did not show a strong interaction between SES and gender. While there was a statistically significant interaction between economic status and gender with a $p = 0.031$ and partial $\eta^2 = .0000$ for grades 3 and 4 and a $p = 0.032$ and a partial $\eta^2 = .0000$ for grades 4 and 5, both of those interactions had a negligible effect size and were not included in the means analysis for mathematics (Table 4.3).

Table 4.3 - Mathematics Split-Plot ANOVA

Grade	Source of Variation	F	Significance Level	Partial Eta Squared
All Grades	SES	6421.486	0.000	0.088
	Gender	19.336	0.000	0.000
	Interaction	3.784	0.052	0.000
3 and 4	SES	1203.660	0.000	0.085
	Gender	5.386	0.200	0.000
	Interaction	4.967	0.031	0.000
4 and 5	SES	1304.986	0.000	0.090
	Gender	2.553	0.112	0.000
	Interaction	4.564	0.032	0.000
5 and 6	SES	1383.921	0.000	0.093
	Gender	14.570	0.000	0.001
	Interaction	0.254	0.219	0.000
6 and 7	SES	1348.371	0.000	0.091
	Gender	19.663	0.000	0.001
	Interaction	0.179	0.672	0.000
7 and 8	SES	1370.550	0.000	0.091
	Gender	3.188	0.074	0.000
	Interaction	0.261	0.609	0.000

Result: For mathematics the null hypothesis that there will not be a statistically significant change in summer learning based on SES is rejected for all grades studied at the $p < .001$ level. The data suggest that changes in student achievement in mathematics over the summer of 2009 were affected by a child's SES for every grade studied. Also for mathematics, the null hypothesis that there will not be a statistically significant change in summer learning based on gender is rejected for grades 5 and 6 and 6 and 7 at the $p < .001$ level, but confirmed for grades 3 and 4, grades 4 and 5 and for grades 7 and 8. The data suggest that changes in student achievement in mathematics over the summer of 2009 were affected by a child's gender for some grades, but not for others. Further analysis showed that while gender had significance for some grade spans, the effects size was negligible. The following analyses quantify the magnitude of the effects of SES on student achievement in mathematics over the summer of 2009.

Research Question Addressed

- What was the magnitude of the affect that SES had on summer learning in mathematics?

A descriptive analysis was performed on the mathematics data looking at each grade individually, and then in aggregate. The goal of that individual grade analysis was to determine if any grade-specific trends in summer-learning changes emerged from the data. The investigation was done exclusively between low-SES and high-SES students, due to the inferential statistical analysis that indicated that both gender and the gender-SES interaction were not statistically significant for most grade levels, and that for the grade levels where it showed significance, it had a negligible η^2 value.

A descriptive analysis for mathematics over the summer of 2009 (Table 4.4) indicated that the mean summer learning-loss was greatest for children in the lower grades of this study. Both economically disadvantaged children (low-SES) and their non-economically disadvantaged (high-SES) classmates showed a fairly substantial decline in mathematics achievement over the grades 3 to 4 summer. While on average all students lost about two tenths of a performance level, the loss was not evenly distributed between high-SES and low-SES students. For students over the summer of grades 3 to 4, high-SES students lost nearly a fifth of a performance level ($\overline{X}$ = -0.1799) while their low-SES counterparts lost nearly a quarter of a performance level ($\overline{X}$ = -0.2368). For both high-SES and low-SES children, summer had the greatest impact on mathematics learning-loss during the early grades of this study.

Table 4.4 - Change in Mathematics Scores (Fall-Spring)

Grade	High-SES	Low-SES	Summer Loss for Low-SES as Compared to High-SES
3 and 4	-0.1799	-0.2368	-0.0569
4 and 5	-0.0407	-0.1143	-0.0736
5 and 6	0.0078	-0.1106	-0.1184
6 and 7	0.0718	0.0112	-0.0606
7 and 8	0.0055	-0.0259	-0.0314
Total:	-0.1355	-0.4764	-0.3409

While low-SES students showed the greatest decline in achievement in mathematics during the summer of grades 3 to 4 ($\overline{X}$ = -0.2368), the data indicated that low-SES children experienced learning loss during the grades 4 to 5 summer, as well as the grades 5 to 6 summer, with economically disadvantaged children losing just over a tenth of an achievement level in mathematics during each of those summers ($\overline{X}$ = -0.1143 and $\overline{X}$ = -0.1106 respectively). That trend appeared to moderate in the later grades for low-SES children who experienced a slight gain in mathematics achievement during the summer of grades 6 to 7, and a slight loss for children during the summer of grades 7 to 8 ($\overline{X}$ = 0.0112 and $\overline{X}$ = -0.0259 respectively). When viewed collectively, low-SES children in the five grades of this study experienced a total summertime learning-loss in mathematics of nearly a half of a performance level ($\overline{X}$ = - 0.4764).

High-SES children, like their low-SES classmates, show a similar pattern of learning loss in mathematics during the younger grades. High-SES students over the summer of grades 3 to 4 lost about a fifth of a performance level in mathematics ($\overline{X}$ = -0.1799) while high-SES students over the grades 4 to 5 summer declined 0.0407 of a performance level. In the upper three grades (grades 5 to 6, grades 6 to 7 and grades 7 to 8), high-SES children showed modest gains in mathematics achievement during the summer ($\overline{X}$ = 0.0078, $\overline{X}$ = 0.0718 and $\overline{X}$ = 0.0055 respectively). When viewed collectively, high-SES children in all grades of this study had a total summertime learning-loss in mathematics of just over a tenth of a performance level ($\overline{X}$ = - 0.1355).

The total summer learning-loss measured in all grades of this study for low-SES students was nearly one half of an achievement level ($\overline{X}$ = -0.4764). Their non-economically disadvantaged classmates who experienced some loss of achievement in the earlier grades, but made academic progress during the three summers between grades 5 through 8, showed only a cumulative summer loss of just over a tenth of an achievement level ($\overline{X}$ = -0.1335). Differences in achievement over the summer break between high-SES and low-SES cumulatively account for slightly more than one third of a performance level ($\overline{X}$ = 0.3409) of difference between these two groups. That sizable disadvantage during summer recess for low-SES children is consistent with the findings of Cooper et al., 1996 and Downey et al.

Moreover, table 4.4 shows that at every grade high-SES students show a summer learning-advantage in mathematics over their low-SES classmates. While the quantity of those achievement-differences varies based on a student's grade level, high-SES students nonetheless gained an academic advantage as a result of the summer recess when compared to their low-SES classmates in every case studied. That finding implies that out-of-school factors contributed to a widening of the learning gap between economically different groups of students.

Mean Analysis of Mathematics Achievement Organized by MEA Performance Level.

The prior analysis suggests that high-SES students gained an advantage in mathematics achievement over their low-SES counterparts during the summer of 2009 for all grades studied. One concern with that finding was that high-SES students had a higher starting (MEA) average mathematics achievement level than did low-SES students. As Table 4.5 shows, in all cases high-SES students have a mean mathematics score above the grand mean, while in all cases low-SES students have a mean score below the grand mean. To determine if starting achievement dissimilarities between low-SES and high-SES students was a contributing factor to the findings above, an additional analysis was conducted. It compared high-SES and low-SES children who performed at the same mathematics MEA achievement-level.

Table 4.5 - Math Mean Scores by SES

Grade	MEA			NECAP		
	Mean High-SES	Mean Low-SES	Mean Total	Mean High-SES	Mean Low-SES	Mean Total
3 and 4	2.5402	2.1205	2.3644	2.3603	1.8837	2.1607
4 and 5	2.4253	1.9998	2.2483	2.3846	1.8555	2.1645
5 and 6	2.4091	1.9717	2.2314	2.4169	1.9717	2.1912
6 and 7	2.253	1.7602	2.0574	2.3248	1.7714	2.1052
7 and 8	2.3054	1.7846	2.1131	2.3109	1.7587	2.107

To compare academically similar groups of students with respect to SES, an analysis of mean performance was divided based on each student's mathematics achievement level on the spring MEA assessment. The MEA was used as the dividing point because it was the independent variable in this analysis, and represents the pre-treatment assessment or the starting condition for student achievement in this study.

A descriptive analysis of mean achievement level changes comparing high-SES to low-SES students over the summer of 2009 disaggregated by MEA achievement level indicated that for every achievement level, high-SES students gained an achievement advantage during the summer of 2009 over their low-SES classmates (See Table 4.6). For students at the lowest achievement level, both high-SES and low-SES children show a fairly sizable gain in performance level. High-SES students gained about a seventh of an achievement level ($\bar{X}$ = 0.1434), while low-SES students gained nearly a tenth of an achievement level ($\bar{X}$ = 0.0996). For students who scored at an achievement level of 2, high-SES students showed a modest average gain of 0.0162 of an achievement level while their low-SES classmates showed a decline of 0.0926 of an achievement level. Both high-SES and low-SES students who scored an MEA achievement level of 3 showed a decline over the 2009 summer recess, with high-SES students losing 0.0235 of an achievement level, while low-SES students lost a more substantial 0.1734 of an achievement level. Finally all students who performed at the highest achievement level on the MEA showed the greatest losses. High-SES students lost 0.1346 of an achievement level while their low-SES classmates lost on average just over a quarter of an achievement level ($\overline{X}$ = -0.2672).

Table 4.6 - Change in Achievement Level by Starting Achievement Level

Performance Level Change	Mean High-SES	Mean Low-SES	Summer Loss for Low-SES as Compared to High-SES
MEA Level 1	0.1434	0.0996	-0.0438
MEA Level 2	0.0162	-0.0926	-0.1088
MEA Level 3	-0.0235	-0.1734	-0.1499
MEA Level 4	-0.1346	-0.2672	-0.1326

In general, students who performed poorly on the MEA in mathematics in the prior year showed a mean improvement when they took the fall NECAP assessment, while students who performed well on the MEA in the prior year showed a mean decline in achievement level when they took the fall NECAP assessment. While that trend was fairly consistent for both high-SES and low-SES students, the losses for low-SES students were greater than they were for high-SES students, and the gains for low-SES students were smaller than they were for high-SES students (see Table 4.6).

In comparing high-SES with low-SES students over the summer of 2009 at every performance level, high-SES students made relative achievement gains over their low-SES classmates. For students scoring at MEA level 1 in the spring, the gain for high-SES students over the summer of 2009 was relatively small, at 0.0438 of a performance level. For students scoring at MEA level 2 in the spring, the relative gain for high-SES students over the summer of 2009 was more robust, at just over a tenth of a performance level ($\overline{X}$ = 0.1088). For students scoring at MEA level 3 or 4 in the spring, the relative gain for high-SES students over the summer of 2009 was even greater: 0.1499 and 0.1326 of a performance level respectively. Those results suggest that regardless of starting performance-level on the MEA, when children were away from the influences of school, high-SES students experienced a relative achievement gain in mathematics over their low-SES classmates.

A deeper analysis (not included) comparing the mathematics achievement score changes between high-SES and low-SES students for each grade individually showed that regardless of a child's starting performance level in the spring, the average achievement-change during the summer break showed that high-SES students made relative academic gains over their low-SES classmates. That was true in *every* grade studied and at *every* achievement level. Every way that the data were sorted, high-SES students made measurable academic gains in mathematics during the summer when compared to their low-SES classmates.

The analysis is important because it controls for a child's initial ability to perform on a standardized assessment. By grouping students who achieved equally on the MEA, this analysis compared a cohort of academically similar students who differed only by their economic status.The analysis suggests that even when controlling for a child's achievement level in mathematics, a high-SES child on average will have a higher mathematics achievement level after the summer recess than a low-SES child, due to factors having nothing to do with the value that school is able to add to that child's learning.

Reading

Analysis of Achievement Changes in Reading Over the Summer of 2009 with Between Subjects Factors.

Research Questions Addressed:

1. Were their statistically significant differences in summer achievement changes in reading for students of differing SES? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?
2. Were their statistically significant differences in summer achievement changes in reading for students of differing gender? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?

The earlier analysis (t-test) of changes in reading achievement over the summer for all students suggested that in the earliest grades students continue to improve their reading levels when school is not in session. The data also suggested that the difference was greatest for children over the grades 3 to 4 summer who showed a robust improvement of just over a tenth of a performance level ($\bar{X}$ = 0.1120), but was dampened for children over the summer of grades 4 to 5, who showed a mean improvement of 0.0754 of a performance level, and was even smaller for children over the summer of grades 5 to 6 who showed a mean improvement of 0.0579 of a performance level. Children over the summer of grades 6 to 7 showed a slight decline in reading achievement of 0.0470 of a performance level, and children over the summer of grades 7 to 8 showed a more substantial decline of 0.1510 of a performance level.

While this descriptive analysis suggests the general learning trends that occurred in reading for students when school was not in session, it does not contain information about how different categories of students fared over the summer. In the analysis that follows, those general trends for all students have been disaggregated to compare the statistically significant changes in learning that occurred over the summer in reading for different categories of students. The analyses report absolute changes for each group, as well as relative differences between contrasting groups of students.

The split-plot ANOVA of reading achievement score changes for all of the grade-spans in the study indicated that in all cases, the categories of MEA economically disadvantaged (high-SES/low-SES), as well as gender, had a statistically significant impact in reading performance over the summer. A child's SES designation had a statistically significant ($p < .001$) affect on student learning, with a partial η^2 varying from 0.082 to 0.101 for all grades studied (Table 4.7). A child's gender also had a statistically significant effect on student performance in reading for all grade levels studied with $p < .001$ and partial η^2 varying from between 0.020 to 0.041 for all grades studied. Finally, the inferential analysis indicated that there was not a statistically significant relationship between the interaction of gender and SES on a child's reading score over the summer.

Table 4.7 - Reading Split-Plot ANOVA

Grade	Source of Variation	F	Significance Level	Partial Eta Squared
All Grades	SES	6782.892	0.000	0.092
	Gender	1972.636	0.000	0.029
	Interaction	0.709	0.400	0.000
3 and 4	SES	1159.143	0.000	0.082
	Gender	276.276	0.000	0.021
	Interaction	2.715	0.099	0.000
4 and 5	SES	1254.999	0.000	0.087
	Gender	379.117	0.000	0.028
	Interaction	3.86	0.049	0.000
5 and 6	SES	1508.933	0.000	0.101
	Gender	281.599	0.000	0.020
	Interaction	0.142	0.706	0.000
6 and 7	SES	1383.647	0.000	0.093
	Gender	511.624	0.000	0.037
	Interaction	0.199	0.656	0.000
7 and 8	SES	1421.938	0.000	0.094
	Gender	577.909	0.000	0.041
	Interaction	2.533	0.112	0.000

Result: For reading, the null hypothesis that there will not be a statistically significant change in summer learning based on SES was rejected for all grades studied at the $p < .001$ level. The data suggest that changes in student achievement in reading over the summer of 2009 were affected by a child's SES for every grade studied. For reading, the null hypothesis that there will not be a statistically significant change in summer learning based on gender was rejected for all grades studied at the $p < .001$ level. The data suggest that changes in student achievement in reading over the summer of 2009 were affected by a child's gender for all grades. The following analyses quantify the magnitude of the effects of those two binary variables on student achievement in reading over the summer of 2009.

Research Question Addressed:

- What was the magnitude of the affect that SES had on summer learning in reading?

Because the inferential statistical analysis showed that SES had a statistically significant impact on summertime achievement in reading, a descriptive analysis was performed for each grade individually, and then in aggregate. The goal of that individual grade-level analysis was to determine if any grade-specific trends in summer learning changes emerged from the data.

A descriptive analysis for reading over the summer of 2009 suggested that both low-SES and high-SES children showed a mean improvement in their reading scores over the grades 3 to 4, grades 4 to 5 and grades 5 to 6 summers. During the summer between grades 3 to 4, grades 4 to 5 and grades 5 to 6, high-SES students had a mean improvement in reading achievement of $\overline{X}$ = 0.1608, $\overline{X}$ = 0.0970 and $\overline{X}$ = 0.0686 of a performance level respectively. For low-SES students during the summer between grades 3 to 4, grades 4 to 5 and grades 5 to 6, the mean improvement in reading achievement scores was somewhat more modest at $\overline{X}$ = 0.0444, $\overline{X}$ = 0.0450 and $\overline{X}$ = 0.0424 of an achievement level respectively. During the summer of grades 6 to 7 and grades 7 to 8, both high-SES and low-SES children had a mean loss in reading achievement scores. For high-SES children those losses over the summer of grades 6 to 7 were $\overline{X}$ = -0.0399 of an achievement level, while over the grades 7 to 8 summer the average loss was about one-seventh of an achievement level ($\overline{X}$ = -0.1451). For low-SES children, losses over the summer of grades 6 to 7 were 0.0578 of an achievement level, while over the grades 7 to 8 summer the average loss was just over one-seventh of an achievement level ($\overline{X}$ = -0.1610).

Table 4.8 - Change in Reading Scores (Fall-Spring)

Grade	High-SES	Low-SES	Summer Loss for Low-SES as compared to High-SES
3 and 4	0.1608	0.0444	-0.1164
4 and 5	0.0970	0.0450	-0.0520
5 and 6	0.0686	0.0424	-0.0262
6 and 7	-0.0399	-0.0578	-0.0179
7 and 8	-0.1451	-0.1610	-0.0159
Total:	0.1414	-0.0870	-0.2284

While changes in student achievement over the summer of 2009 indicated that children tended to improve in their reading achievement levels over the summer in the earlier grades, but decline in their reading achievement levels over the summer in the upper grades, for every grade in this study high-SES students had a relative advantage during the summer when compared to their low-SES classmates (Table 4.8). While both low-SES and high-SES students improved their achievement scores in reading during the summers between grades 3 to 4, grades 4 to 5 and grades 5 to 6, the achievement gap between high-SES and low-SES students grew during those summers. Low-SES children during the summer of grades 3 to 4 had a relative loss to their high-SES classmates of just over a tenth of a performance level ($\overline{X}$ = -0.1164). That trend continued, although somewhat moderated, for children in grades 4 to 5 and children in grades 5 to 6 where low-SES students had a mean relative loss of $\overline{X}$ = -0.0520 and $\overline{X}$ = -0.0262 of a performance level respectively. Finally, while both high-SES and low-SES appeared to show decline in reading achievement over the summer, the decline was more pronounced for low-SES children. During the summer of grades 6 to 7, low-SES students lost 0.0179 of an achievement level, and over the summer of grades 7 to 8 they lost 0.0159 of an achievement level to their high-SES classmates. The total summer differential in reading measured for all grades of this study indicate that low-SES students lost nearly one-quarter of an achievement level ($\overline{X}$ = -0.2284) to their high-SES classmates over the five grades analyzed. If cumulative, that difference of nearly one-quarter of a performance level represents a widening of the achievement gap in reading between different SES children, having nothing to do with the effects of schooling.

The mean differences in learning between high-SES and low-SES students in this study indicate that the biggest differences in learning between the groups occurred during the earliest grades. Nearly three-quarters (74 percent) of the differential between high-SES and low-SES came in the first two summers of the study. That finding suggests that time away from school creates the biggest reading achievement differences between high-SES and low-SES children, and academically divides those children in the earliest grades of this study.

Analysis of Reading Achievement Changes for SES Organized by MEA Reading Performance Level.

The above analysis suggested that high-SES students gained an advantage in reading achievement over their low-SES counterparts during the summer of 2009. One concern with this finding was that high-SES students had a higher starting (MEA) average reading achievement level than did low-SES students. As Table 4.9 shows, in all cases high-SES students had a mean reading score above the grand mean, while in all cases low-SES students had a mean score below the grand mean. To determine if initial achievement dissimilarities between low-SES and high-SES students was a contributing factor to the findings above, an additional analysis was conducted that compared high-SES and low-SES children who performed at the same reading achievement level on the MEA.

Table 4.9 - Reading Mean Scores by SES

	MEA			NECAP		
Grade	Mean High-SES	Mean Low-SES	Mean Total	Mean High-SES	Mean Low-SES	Mean Total
3 and 4	2.2907	1.9854	2.1628	2.4515	2.0298	2.2748
4 and 5	2.4037	2.0729	2.2664	2.5007	2.1179	2.3418
5 and 6	2.3804	1.991	2.2218	2.449	2.0334	2.2797
6 and 7	2.4567	2.0739	2.3048	2.4168	2.0161	2.2578
7 and 8	2.6216	2.2114	2.4701	2.4765	2.0504	2.3191

To compare academically similar groups of students with respect to SES, an analysis of mean achievement was divided based on each student's achievement level on the spring MEA assessment. The MEA was used as the dividing point because it was the independent variable in this analysis, and represents the pre-treatment assessment or the starting condition for student achievement in this study.

A descriptive analysis of mean reading achievement level changes comparing high-SES to low-SES students over the summer of 2009 disaggregated by MEA reading achievement level, indicated that for every achievement level, high-SES students gained an achievement advantage in reading over their low-SES classmates during the summer recess. For students at the lowest achievement level, both high-SES and low-SES children showed gains in reading achievement level with high-SES students gaining about a third of an achievement level ($\overline{X}$ = 0.3124) while low-SES students gained just over a quarter of an achievement level ($\overline{X}$ = 0.2524). For students who scored at an MEA reading achievement level of a 2, high-SES students showed a somewhat more modest average gain of just over a tenth of an achievement level ($\overline{X}$ = 0.1189), while their low-SES classmates showed an even smaller gain of 0.0445 of an achievement level. For students who had a spring MEA reading achievement level of a 3, high-SES students showed a slight gain of 0.0145 of an achievement level, while low-SES students showed a decline of 0.0830 of an achievement level. Finally, all students who performed at the highest reading achievement level on the MEA reading assessment showed losses over the summer of 2009. High-SES students lost 0.1460 of an achievement level while their low-SES classmates lost on average nearly a quarter of an achievement level ($\overline{X}$ = -0.2358).

In general, students who performed poorly on the MEA in the prior year showed a mean improvement when they took the fall NECAP assessment, while students who performed well on the MEA in the prior year showed a mean decline in achievement level when they took the fall NECAP assessment. While that trend was fairly consistent for both high-SES and low-SES students, the losses for low-SES students were greater than they were for high-SES students, and the gains for low-SES students were smaller than they were for high-SES students.

An analysis of the differential between high-SES and low-SES students over the summer of 2009 in reading achievement scores showed that at every achievement level high-SES students made relative achievement gains over their low-SES classmates (Table 4.10). For students scoring at an MEA reading achievement level of a 1, high-SES students gained an average of 0.0600 of an achievement level over low-SES students. For MEA reading achievement level 2 students, the difference was a slightly larger 0.0744, and for MEA reading achievement level 3 students the difference was nearly one-tenth of an achievement level ($\overline{X}$ = 0.0975). Finally, for students with the highest MEA reading achievement level the difference between high-SES and low-SES students was 0.0898 of an achievement level.

Table 4.10 - Reading Change in Achievement Level by Starting Performance Level (SES)

Performance Level Change	Mean High-SES	Mean Low-SES	Summer Loss for Low-SES as Compared to High-SES
MEA Level 1	0.3124	0.2524	-0.0600
MEA Level 2	0.1189	0.0445	-0.0744
MEA Level 3	0.0145	-0.0830	-0.0975
MEA Level 4	-0.1460	-0.2358	-0.0898

The results suggest that regardless of a student's initial achievement level in reading, when children are away from the influences of school, high-SES students experience a relative reading achievement gain over low-SES students. While that difference in reading achievement over the summer between low-SES and high-SES students is smaller than that found for mathematics, it nonetheless suggests that out-of-school factors influence a child's academic development differently for economically dissimilar students.

Descriptive Analysis of the effect of Gender on Reading Achievement Over the of 2009 Summer by Grade.

Research Question Addressed:

- What was the magnitude of the affect that gender had on summer learning in reading?

Because the inferential analysis indicated that there was a statistically significant difference in summer learning patterns in reading based on gender, a descriptive analysis of the data was conducted to determine the magnitude of the affect. The following analysis was first divided by gender, then by grade individually. Finally, the analysis compared changes in achievement by gender to quantify learning differences between genders over the years in this study.

Female students in all grades other than grades 7 to 8 had achievement gains in reading over the summer of 2009. That improvement in achievement level was greatest in the earliest grades, with girls in the grades 3 to 4, grades 4 to 5, and grades 5 to 6 showing an average summer gain of just over one-tenth of an achievement level each year ($\overline{X}$ = 0.1313, $\overline{X}$ = 0.1241 and $\overline{X}$ = 0.1214 respectively). The gain dampened for girls during the grade 5 to 6 summer, with a mean growth of $\overline{X}$ = 0.0045, and reversed for girls in grades 7 to 8; who showed regression of just over one-tenth of an achievement level ($\overline{X}$ = -0.1330) over the summer of 2009.

Male students also made modest increases in reading achievement levels during the earliest grades of the study, but showed losses over the summer in the later grades. Over the summer of grades 3 to 4, boys gained nearly one-tenth of an achievement level ($\overline{X}$ = 0.0936), but had a much more modest gain during the grades 4 to 5 summer ($\overline{X}$ = 0.0299). Starting during the grades 5 to 6 summer, and increasing over the grades 6 to 7 summer and grades 7 to 8 summer, boys demonstrated a pattern of increasing summer learning loss with mean losses of, $\overline{X}$ = -0.0043, $\overline{X}$ = -0.0952 and $\overline{X}$ = -0.1687 of an achievement level respectively.

Table 4.11 - Reading Academic Change over the Summer of 2009 by Gender

Grade	Mean Female Change	Mean Male Change	Summer Gain for Females as Compared to Males
3 and 4	0.1313	0.0936	0.0377
4 and 5	0.1241	0.0299	0.0942
5 and 6	0.1214	-0.0043	0.1257
6 and 7	0.0045	-0.0952	0.0997
7 and 8	-0.1330	-0.1687	0.0357
Total:	0.2483	-0.1447	0.3930

Over every summer in the study, females made relative academic progress over their male classmates (Table 4.11). In the earliest grades that learning differential was a modest 0.0377 of a performance level, but the differential increased over each subsequent summer grade spans until it reached a maximum of 0.1257 of an achievement level during the grades 5 to 6 summer. That learning differential then became less pronounced over the grades 6 to 7 and grades 7 to 8 summers, with a mean change of 0.0997 and 0.0357 of an achievement level respectively. That pattern creates a nearly symmetrical graph with an apex at the grades 5 to 6 differential. Taken cumulatively over the grades of this study, the learning differential in reading between male and female students was well over one-third of a performance level ($\overline{X}$ = 0.3930).

Analysis of Reading Achievement Changes for Gender Organized by MEA Performance Level.

In a separate analysis of the data, mean scores were sorted by a child's MEA reading achievement level to compare summer achievement changes for children of different genders who performed at the same reading achievement level. That was done to control for any differences in mean performance on the MEA reading assessment between female and male students.

Table 4.12 - Reading Change in Achievement Level by Starting Performance Level (Gender)

Performance Level Change	Mean Change Female	Mean Change Male	Summer Loss for Males as Compared to Females
MEA Level 1	0.2959	0.2582	0.0377
MEA Level 2	0.1315	0.0383	0.0932
MEA Level 3	0.0353	-0.0755	0.1108
MEA Level 4	-0.1318	-0.2133	0.0815

A descriptive analysis of changes in mean reading-performance level comparing female to male students over the summer of 2009 disaggregated by MEA reading performance level shows that for every performance level, female students have an achievement advantage in reading over their male classmates (see Table 4.12). For students at the lowest achievement level, both female and male children showed gains in reading achievement level, with female students gaining just under one-third of an achievement level ($\overline{X}$ = 0.2959), while male students gained just over one-quarter of an achievement level ($\overline{X}$ = 0.2582). For students who scored at an MEA reading achievement level of a 2, female students showed a somewhat more modest average gain of just over one-tenth of an achievement level ($\overline{X}$ = 0.1315), while their male classmates showed a somewhat smaller gain of 0.0383 of an achievement level. For students who had a MEA reading achievement level of a 3, female students had a slight gain of 0.0353 of an achievement level, while male students showed a decline of 0.0755 of an achievement level. Finally all students who performed at the highest reading achievement level on the MEA reading assessment had academic losses over the summer of 2009. Female students lost 0.1318 of an achievement level while their male classmates lost on average just over one-fifth of an achievement level ($\overline{X}$ = -0.2133).

Looking at the differential between female and male students over the summer of 2009 in reading achievement scores, at every performance level, shows that female students made relative achievement gains over their male classmates. For students scoring at an MEA reading performance level of a 1, female students gained an average of 0.0377 of a performance level over male students. For MEA reading performance level 2 students, that difference is a larger 0.0932 of an achievement level, and for MEA reading achievement level 3 students the difference is just over one-tenth of an achievement level ($\overline{X}$ = 0.1108). Finally, for students with the highest MEA reading achievement level the difference between female and male students is 0.0815 of an achievement level. Those results suggest that regardless of a student’s initial achievement level, when children are away from the influences of school, female students experience a relative reading achievement gain over male students.

CHAPTER V: CONCLUSIONS

Introduction

The purpose of this study was to determine the affect that out-of-school factors had on student achievement in reading and mathematics. This ex post facto research was conducted using a test-retest or repeated measures method. The initial measurement of student achievement was the spring 2009 MEA. The retest was the fall 2009 NECAP. By measuring student achievement in reading and math before the summer recess of 2009 and then again afterwards, a measure of student growth when school was not a contributing factor for learning was obtained. The “treatment” in this experimental design was the summer recess of 2009.

Using a repeated measures format as a way to isolate the affect that out-of-school factors had on student learning has a rich history in the literature of summer learning. Most notably in recent studies, the idea was proposed in 1978 by Heyns to refute some of the findings of the Coleman Report. Because the summer recess represents a time when the affects of school on student learning are negligible, it is reasonable to attribute changes in achievement levels on a repeated measures assessment to out-of-school factors.

The sample for this study comprised all children in all districts from grade 3 to grade 8. From that large data set, students who did not take either of the two assessments, students who were retained or skipped a grade, students who moved in or out of the state, and students who took the Personalized Alternative Assessment Portfolio (PAAP) were excluded from the study.

To confirm the similarity of the two assessments, a regression analysis was performed comparing the consistency of a child’s scores on the MEA to that same child’s score on the NECAP. To further confirm the similarity of the assessments, as well as to verify the validity of the 4-point linked scale, an analysis using Cronbach’s Alpha was performed. Both analyses provided a statistically sound basis for treating the MEA and the NECAP as a situation “where each test was (or could reasonably be regarded as) a random sample from a specified domain of admissible test items (p.400).”

This study was organized to answer the following research questions:

1. Was there a statistically significant change in mathematics achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually as well as for all grade spans combined.
2. Was there a statistically significant change in reading achievement scores for children in Maine during the summer of 2009? If so, what was the magnitude of that change? This question was analyzed for each grade span individually as well as for all grade spans combined.
3. Were there statistically significant differences in summer achievement changes in mathematics for students of differing SES? If so, what was the magnitude of those differences, and did those differences vary according to a child’s grade level?
4. Were there statistically significant differences in summer achievement changes in mathematics for students of differing

gender? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?

5. Were there statistically significant differences in summer achievement changes in reading for students of differing SES? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?

6. Were there statistically significant differences in summer achievement changes in reading for students of differing gender? If so, what was the magnitude of those differences, and did those differences vary according to a child's grade level?

Summary of Study

This quantitative ex-post-facto study first attempted to determine if there were statistically significant changes in reading and mathematics over the summer for all students in all grades. Because of the very large sample size and the lack of supporting research to support a directional hypothesis, only a significance level of p <.001 was used to determine statistical significance. Paired sample t-tests were used to show the size and significance of summer learning in mathematics and reading for each grade and for all students. With general patterns of achievement change over the summer in mathematics and reading established, a split-plot ANOVA was performed on a combined all-grades data set, as well as with each of the five individual grade-span data sets for both reading and mathematics to determine if there were statistically significant differences in summer achievement for different categories of students. The initial measure for the split-plot ANOVA was a child's performance on the spring MEA, while the second measurement was that same child's performance on the fall NECAP in the same content area. All inferential analyses were conducted using the 4-point linked scale, and results were verified using both Z-scores and raw scores. The between-subjects factors or independent variables explored in this analysis were MEA Economically Disadvantaged and MEA Gender. Finally, a comparison of mean changes was performed on all statistically significant findings to quantify the size of academic changes over the summer.

Discussion of Results

There is a dearth of information on the effects of summer learning in areas of the country that are as rural as the state of Maine. The majority of studies cited in the literature on summer learning use samples from urban areas like Baltimore (Alexender et. al.) and Atlanta (Heyns). Furthermore, this study employs a very large sample size (reading N = 66,828 and mathematics N = 66,857). The analysis was conducted on each individual student's achievement change over the summer, irrespective of that child's town, school, or location in the state. That makes this study a unique investigation into the effects of out-of-school factors on student achievement on standardized assessments.

The discussion of results that follows was divided by content area. Mathematics preceded reading. The discussion addresses general trends in summer achievement then focuses in on differences in summer learning between different groups of students.

Mathematics

When student achievement for all grades of this study was analyzed collectively, it appeared that when school was not in session, students showed a cumulative loss of just about 11 percent of a standard deviation or an average loss of just over 2 percent of a standard deviation per year. From that finding one might conclude that when Maine children in grades 4 through 8 returned in the fall of 2009, they began school at approximately the same math achievement level that they held the prior spring. But a closer examination showed that the change in performance over the summer was not uniform across all grades studied. For the youngest children in this study, summer learning-loss in mathematics was greater. Nearly 70 percent of the total learning-loss experienced in all grades occurred during the summer between grades 3 to 4. For those children, the summer recess represented a time where they collectively lost nearly 40 percent of a standard deviation in mathematics from the previous spring.

While there was not a strong relationship between a student's summertime mathematics achievement change and that child's gender, there was a significant affect for a child's SES on summertime mathematics achievement. In each of the five summers studied, high-SES children gained a relative advantage in mathematics over their low-SES classmates. High-SES children made small gains over their low-SES classmates in the earliest years of the study, but the gains became greater each summer until the grades 6 to 7 summer when the difference between high and low-SES children began to decline. The largest of those occurred over the summer of grades 5 to 6 as many Maine children are transitioning to middle school. Taken cumulatively over the course of this study, high-SES children made a cumulative gain of just over one-third of a performance level in mathematics over their low-SES classmates during the five years studied.

Reading

When student achievement in reading for all grades of this study was analyzed collectively, it appeared that when school was not in session, students showed a slight gain in reading achievement of just about 2 percent of a standard deviation. While that cumulative difference was negligible, the change was by no means representative of what happened over the summer in reading achievement at different grade levels. For the youngest children in this study, summer was a time where achievement in reading improved. That improvement became less significant in the later years of this study until the summer of grades 6 to 7 when children experienced academic loss. The loss then became much greater during the summer between grades 7 to 8, when children lost 32 percent of a standard deviation in reading.

While the general pattern of reading achievement for all students has some implications for schools, a more in-depth analysis of reading achievement for different groups of students gave a more nuanced look at the influence that out-of-school factors had on student achievement in reading.

The split-plot ANOVA indicated a significant relationship between a child's SES and that child's change in reading achievement level over the summer. High-SES children gained in reading achievement over their low-SES classmates for every grade studied. What was more interesting was that nearly three quarters (74 percent) of the differential between high-SES and low-SES came in the first two summers of the study. That finding implies that at the earliest grades of this study when children were transitioning from learning to read to reading to learn, low-SES children had a large disadvantage over the summer when compared to their high-SES classmates.

The ANOVA for reading also indicated that there was a significant relationship between a child's gender and that child's change in achievement for all grades over the summer of 2009. This study indicated that female students made gains over their male classmates in every grade studied. While the difference started out small during the summer of grades 3 to 4, it increased to nearly one-tenth of a performance level the next year, and peaked at 0.1275 of a performance level during the summer between grades 5 to 6. The difference between male and female reading achievement level over the summer then declined during the summer of grades 6 to 7, as well as for students during the grades 7 to 8 summer. While that result suggests that out-of-school influences favored girls' reading achievement, there are no data from this study to suggest a cause or mechanism to create such a difference in out-of-school achievement between the genders.

Controlling for Starting Achievement Level

Because low and high-SES students varied so largely in MEA and NECAP performance in both mathematics and reading, a further analysis was conducted to determine if the difference in achievement level between high and low-SES affected the findings. To control for a student's starting MEA performance, a t-test was conducted comparing changes in achievement over the summer of students who started at the same performance level. In this analysis, low-SES students showed a relative achievement loss to their high-SES classmates at every starting achievement level for both reading and mathematics. That finding supported the conclusion that differential losses that occurred for low-SES children were not a product of their initial test scores, but rather were a result of the out-of-school factors that impacted children's learning during the summer break when school was not an influence.

Implications

The long summer recess has been a characteristic of the American school system since the beginning of the nineteenth century. While that break represents an inefficiency in the American educational system with respect to the amount of time a child has to learn, the findings from this study also suggest that the summer break can actually be harmful to certain children. Because this study found that not all children reacted to the affects of time away from school in the same way, the summer recess acted as a mechanism to academically sort children.

While this study did not probe the specific mechanisms that affected Maine students during the summer of 2009, it did indicate that there was something different that happened for high-SES children over the summer that led to relative academic growth when compared to their low-SES classmates. For low-SES students each summer represented an opportunity to slip just a little further academically behind their high-SES classmates. The difference in achievement over the summer was robust, and persisted even when the initial achievement level was controlled for in the analysis. That slow widening of the achievement gap based on a child's SES suggests that policymakers hoping to close the achievement gap should focus at least some of their efforts on summertime differences in learning, rather than exclusively attributing the achievement gap to deficiencies in teaching.

For accountability measures that mandate an annual assessment to measure the effectiveness of teachers, the finding that summer affects children differently based on their SES suggests that inclusion of the summer recess in teacher or school accountability measures represents an error factor that needs to be controlled for in some way. That finding is significant in an era of high-stakes assessments that rate schools or teachers based on annual measures. Any annual assessment, by virtue of being annual, has the error factor created by the addition of out-of-school factors in its calculations. The results of this study suggest that schools with a high percentage of low-SES students will have deflated ratings, while those schools with a high percentage of high-SES students will have inflated ratings that have nothing to do with the quality of their instructional programming but rather are a result of factors exogenous to the efforts of the school.

The ESEA Flexibility Review Guidance for Window 3 from the US Department of Education asks states filling out the ESEA waiver: "Does the SEA incorporate student growth into its performance-level definitions with sufficient weighting to ensure that performance levels will differentiate among teachers and principals who have made significantly different contributions to student growth or closing achievement gaps?" (2012, p.19). But the findings of this study suggest that any annual measure that includes summer learning will have the effect of magnifying any achievement gap irrespective of the impacts a school or teacher may have had.

Just as making a mark on a wall or door jam each year is an accurate measure of a child's annual growth in height, an annual academic assessment (apart from errors in the instrumentation) is an appropriate measure of a child's total annual academic growth. From the lens of a child, an annual academic measure is the most important measure because it reports how much growth occurred during the year. Just as the annual mark on the wall answers the question of how much taller a child grew in a given year, so too does the annual assessment (at least in theory) indicate how much academic growth a child made in a given year. This study indicates, however, that it is important to distinguish a child's annual academic growth from the academic growth that a child made as a result of the affects of schooling. This study clearly shows that different categories of children learned (or regressed) at different rates during the summer when in-school factors were controlled for.

While it is beyond the scope of this research to suggest reasons for the differential in summer learning in math between high-SES and low-SES children, or the differential in reading between high-SES and low-SES students, as well as between male and female students, if policymakers are serious about the premise of No Child Left Behind with respect to closing the achievement gap for all sub-categories of students over time, then funding for programming that addresses the out-of-school learning differential between different groups of students is warranted. If schools are to be the great equalizer with respect to SES differences, then their roles must grow proportionally to address out-of-school learning differences between high-SES and low-SES students. An obvious place to start is with funding for summer educational programming for low-SES students.

It is important to note that the analysis was intentionally conducted at the student level so that aspects such as school size, school funding or any other school or instructional influences on student learning did not interfere with the analysis. Put another way, this study indicated that when school was not in session, high-SES children on average gained a relative advantage over their low-SES classmates during every summer in both reading and mathematics for every achievement level and grade studied, independent of that child's community or type of school attended.

Limitations of Results

The foundation of this investigation was a comparison of students' achievement on the MEA in the spring with those same students' achievement on the NECAP in the fall. By virtue of the fact that the two assessments were different, some caution with respect to the magnitude of student differences should be exercised. That is more of a concern for absolute changes in student achievement that compare spring to fall achievement levels than it is for the relative or comparative student achievement levels that compare differences in achievement for different categories of students.

Another limitation of the study was the time lapse between the administrations of the assessments. The MEA was administered near the end of March of 2009, while the NECAP took place during the first weeks of October of 2009. The gap of just about half a year represents a good deal of learning time in school. While the interval also included the two and a half months of the summer recess, the data were likely affected by the learning that occurred in school between the two tests. Cooper et.al (1996) suggest that an extended interval between test administrations should lead to an undervaluation of summer loss.

The study was also limited to the single year of 2009 when the state of Maine changed from the MEA to the NECAP assessment. Because the analyzed data provided a one-year snapshot in time, conclusions about cumulative nature have to be viewed with some caution. While the large sample size ($\overline{X}$=13,374.4/grade) reduces the possibility of a non-representative grade cohort, that nonetheless is a limitation of the findings.

Suggestions for Further Research

Because the study took advantage of a single summer, conclusions about the cumulative nature of differential summer-learning, as well as grade-level changes, must be viewed with some caution. It would be preferable to follow several cohorts of students over a period of many years. Such effort might yield a more reliable indicator of summer-learning patterns for different students.

The patterns of learning in the findings suggest further study in the following areas:

1. In mathematics, nearly 70 percent of the total learning loss experienced in all grades occurred during the summer between grades 3 to 4. The finding suggests further study in the primary grades to determine if that is isolated to the summer of grades 3 to 4, or if the trend continues throughout the primary grades.
2. In mathematics, low-SES children over the summer between grades 5 to 6 lost the most ground to their high-SES classmates. Does the parabola with a peak at the grades 5 to 6 summer extend into the primary grades, as well as into high school? Further study is needed to make any conclusions about differential learning outside of the grades studied.

3. Students showed gains in reading achievement level in the early grades, but the improvement declined each year until it reversed during the summer of grades 6 to grade 7. The pattern suggests:
    a. Further study to determine if the trend of increased reading achievement-loss over the summer would extend into the high school, or if it peaks for children over the summer of grades 7 to 8.
    b. Further study to determine if there is a mechanism for student gains in the early grades, which might lead to ideas to moderate losses in the latter grades. While one might hypothesize that an emphasis on early grade literacy, such as schools challenging young children to read over the summer and parents being encouraged through schools and public service announcements to read to their children, it would be instructive to confirm the mechanism for that growth in the early grades.
4. More concerning about reading is the finding that 74 percent of the reading advantage that high-SES children had over their low-SES classmates occurred in the first two grades of the study. Further study is warranted to determine if that robust learning differential is confined to the first two grades of the study, or if it extends into the primary grades. Should the pattern extend to the primary grades, then a good deal of the achievement gap and academic divergence between low and high-SES students is attributable to early literacy outside of school. That would indicate a particularly important area for policymakers to focus their efforts to close the achievement gap.

5. Finally, the fact that this study indicated that girls have the advantage over boys in summer reading achievement is unexplained, but is fodder for additional research.

Conclusion

This study indicated that when the effects of school are controlled for during the summer recess, students from different backgrounds learn at different rates. For dissimilar SES students that finding is intuitive. Economically disadvantaged children lead different lives outside school than their non-economically disadvantaged classmates. While the specifics of those differences are well beyond the scope of this investigation, Entwisle, Alexander and Olson present a likely model to explain the possible underpinnings of differences in achievement. Their "faucet theory" hypothesized that when school was in session, all children benefitted from the resources that the school had to offer, but when school was not in session, those resources were turned off. For high-SES children time away from school does not create problems: many of those children have access to learning resources by virtue of their economic status. In contrast, when school learning-resources are turned off, low-SES children's families do not have the resources to make up for what had been provided by the school. Those children thus fall academically behind their more economically fortunate peers.

The findings of this study correspond with the finding of similar investigations by Heyns (1978, 1987); Cooper, et al. (1996); Downey, et al. (2004); and Alexander, Entwisle and Olson (2007) which indicate that when school is not in session, high-SES children make relative academic gains over their low-SES classmates. That difference in achievement has a variety of implications with respect to public policy. Most obviously it suggests that efforts that focus exclusively on schools and teachers to close the achievement gap between high and low-SES children are at least in part misdirected. The findings also suggest that accountability measures that use annual assessments to measure the effectiveness of a school or teacher are inherently flawed.

While Heyns' groundbreaking research in 1978 demonstrated the error of using an annual assessment to measure the effect of a nine and a half month process, that flawed thinking has nonetheless persisted in public policy. For example, NCLB and the NCLB waiver make that error with the mandate that an annual assessment be used to measure a school's effectiveness at closing the achievement gap. That flawed national thinking is reflected in the accountability measures of a great many states. In the state of Maine, the MeCAS requires that all students in Maine be assessed annually by the MEA to measure their progress towards meeting the standards of the Maine Learning Results. While an annual assessment is an appropriate measure of student growth, in recent years it has also been used to rate school performance. This study has found that including differences in learning that occurred when children were not in school introduces an error factor that makes that rating process imprecise, and somewhat irresponsible.

If it is a public policy goal to reduce the achievement gap between high and low-SES students, then the most productive approach would be to focus efforts on the time when the differences in learning between those groups are greatest. It is for that reason efforts to close the achievement gap need to focus on out-of-school factors. If the "faucet theory" is the actor, then learning resources need to be made available to all children throughout the calendar year, not just during the school year.

It is tempting to direct efforts to close the academic achievement gap by providing more academic learning experiences for low-SES children. While "turning-on the faucet" might include extended school year services for children, such a solution may be too simplistic. Although Borman (2006) hypothesized that, "Perhaps by turning on the summer school faucet, educators can narrow achievement gaps…" (p.147) he went on to conclude of the summer school program that he was studying that his sample of treatment students served by the summer program did not exhibit large and statistically reliable achievement advance over the control. While the 2011 Rand Corporation report by McCombs, Augistine, and Schwartz (2011) attempted to explore the factors that make summer school effective, they, like Borman, concluded that ensuring regular student attendance can be difficult.

It is possible that high-SES children benefit not from access to academic materials and adult instruction but rather from enriching summer activities, or from some other resource available to high-SES children. Heyns in her 1978 study struggled with that problem when she found that having a bicycle to visit friends and family had, "consistent significant effects on summer achievement when background was controlled." She then continued, "Interpreting such effects literally obviously is illegitimate…" (p.194). While it is unclear what factors allow high-SES children to make academic growth over their low-SES classmates when school is not in session, what is clear from this research, and that done by others, is that they do. Until further study is conducted to better understand the underpinnings of differential summer learning and public policy is enacted that effectively mitigates this social injustice, low-SES children will continue to fall behind their high-SES peers, while annual accountability measures continue to irresponsibly penalize schools serving low-SES students for relative losses that occur outside of the schoolhouse.

REFERENCES

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87-107.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171-191.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180.

Alexander, K., Entwisle, D., & Olson, L. (2014). *The long shadow: Family background, disadvantaged urban youth, and the transition to adulthood*. Russell Sage Foundation.

Bible, H. (1995). New American standard bible. *Grand Rapids: World,*

Blau, D. M. (1999). The effect of income on child development. *The Review of Economics and Statistics, 81*(2), 261-276.

Borman, G. D., & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis, 28*(1), 25-48.

Bruene, E. (1928). Effect of the summer vacation on the achievement of pupils in the fourth, fifth and sixth grades. *The Journal of Educational Research, 18*(4), 309-314.

Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004). Social-class differences in summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education*, *77*(1), 1-31.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). Equality of educational opportunity. *Washington, Dc,* 1066-5684.

Cook, R. C. (1942). Vacation retention of fundamentals by primary-grade pupils. The Elementary School Journal, 214-219.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*(3), 227-268.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391-418.

Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review, 102*(5), 1927-1956.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*(5), 613-635.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. American Sociological Review, 72-84.

Entwisle, D. R. (1997). *Children, Schools, & Inequality. Social Inequality Series*. Westview Press, 5500 Central Avenue, Boulder, CO 80301-2877.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2001). Keep the faucet flowing. *American Educator, 25*(3), 10-15.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*(1), 72-84.

ESEA Flexibility. (n.d.). Retrieved August 1, 2014 from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html.

Fielding, L., Kerr, N., & Rosier, P. (2007). *Annual growth for all students: Catch-up growth for those who are behind* L. Fielding.

Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status?. American Educational Research Journal, 0002831213502516.

Gladwell, M. (2008). *Outliers: The story of success* Little, Brown.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher, 39*(2), 120-131.

Heyns, B. (1978). *Summer learning and the effects of schooling.* Academic Press New York.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning?. Child Development, 1151-1160.

Ladd, H. F. (2012). Education and poverty: Confronting the evidence. *Journal of Policy Analysis and Management*, *31*(2), 203-227.

Maine's Comprehensive Assessment System (MeCAS). (n.d.). Retrieved July 8, 2015, from http://www.maine.gov/education/lsalt.

McCombs, J. S., Augustine, C. H., & Schwartz, H. L. (2011). *Making summer count: How summer programs can boost children's learning* Rand Corporation.

Measured Progress, *Partnership with Maine*, (n.d.). Retrieved May 3, 2015, from http://www.measuredprogress.org/pt/NECAP.

Measured Progress, (2009). *Maine Comprehensive Assessments System MeCAS; 2008–2009 Technical Report Part One Maine Educational Assessment Grades 3–8.* Dover, NH.

Measured Progress, (2010). *New England Common Assessment Program 2009-2010 Technical Report.* Dover, NH.

Ramey, G., & Ramey, V. A. (2009). *The rug rat race* (No. w15284). National Bureau of Economic Research.

Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development the differential effects of school exposure. *Sociology of Education*, *83*(4), 271-286.

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, 91-116.

Sahlberg, P. (2011). Finnish lessons: What the world can learn from educational change in Finland. New York, NY: Teacher’s College Press.

Silva, E. (2007). *On the clock: Rethinking the way schools use time*. Education Sector.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Stover, D. (1989). Should Schools Plow Under the Old Agrarian Calendar?. American School Board Journal, 176(10), 37.

Vale, C., Weaven, M., Davies, A., Hooley, N., Davidson, K., & Loton, D. (2013). Growth in literacy and numeracy achievement: Evidence and explanations of a summer slowdown in low socio-economic schools. *The Australian Educational Researcher, 40*(1), 1-25.

Wiseman, A. W., & Baker, D. P. (2004). The American Summer-Learning Gap from an International Perspective. *Summer Learning: Research, Policies, and Programs, Mahwah, NJ: Lawrence Erlbaum*, 53-69.

BIOGRAPHY OF THE AUTHOR

Brian Mazjanis was born in Connecticut, and graduated from Newington High School in 1985. Upon completion of his undergraduate degree in secondary-science education at the University of Connecticut in 1991, he moved to Maine. In 1999, he completed a master of science degree in educational leadership at the University of Southern Maine. He earned a doctorate in public policy from the Muskie School at the University of Southern Maine in 2015.

Brian began his career in education in 1991 at Marshwood High School in Eliot, Maine, teaching both physical and life sciences, and founding the applied biology and chemistry program. In 1996 he began his career in educational administration at Oxford Hills Comprehensive High School as an assistant principal. In 2000, he became assistant principal at Falmouth High School, then interim principal, before moving to the Westbrook School Department in 2006. In Westbrook, he led the adoption of freshmen teaming as part of the school-within-a-school initiative.

As principal at Westbrook Middle School, Brian implemented a comprehensive school reform and improvement grant, and led the design, construction, and transition to a multi-million dollar facility. He is currently the principal of Saccarappa Elementary School, a K-4 school that serves a linguistically, culturally, and economically diverse population of students. Brian has the unusual opportunity to serve as a principal in schools with grades kindergarten through twelve.